# A system wide study on the hypertrophic response to salbutamol in skeletal muscle cells

Thesis submitted in accordance with the requirements of the

University of Liverpool for the degree of Doctor in Philosophy

by Jenna Kenyani

May 2012

**IMAGING SERVICES NORTH**
Boston Spa, Wetherby
West Yorkshire, LS23 7BQ
www.bl.uk

# THESIS CONTAINS

# CD

**Table of contents**

## List of Tables

## List of Figures

## List of Abbreviations

| | |
|---|---|
| 1D | One dimensional |
| 2DE | Two dimensional gel electrophoresis |
| ACN | Acetonitrile |
| APS | Ammonium persulfate |
| Asb | Ankyrin repeat and SOCS-box |
| ATP | Adenosine triphosphate |
| Bp | Base pairs |
| BVA | Biological Variation Analysis |
| BSA | Bovine Serum Albumin |
| cAMP | Cyclic adenosine monophosphate |
| CID | Collision induced dissociation |
| CHAPS | 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate |
| CREB | cAMP response element binding protein |
| Da | Daltons |
| DAVID | Database for Annotation, Visualization and Integrated Discovery |
| DEPC | Diethylpyrocarbonate |
| DHS | Donor Horse Serum |
| DIA | Differential In-gel Analysis |

| | |
|---|---|
| DIGE | Differential gel electrophoresis |
| DMEM | Dulbecco's Modified Eagle Medium |
| Dmnt1 | DNA methyltransferase 1-associated protein 1 |
| DMSO | Dimethyl sulfoxide |
| dNTP | Deoxyribonucleotide triphosphate |
| DTT | Dithiothreitol |
| EDL | Extensor digitorum longus |
| EDTA | Ethylenediaminetetraacetic acid |
| ePCR | Emulsion polymerase chain reaction |
| ESI | Electrospray ionization |
| ET-1 | carndothelin-1 |
| FA | Formic acid |
| FCS | Foetal calf serum |
| FDR | False discovery rate |
| Gb | Giga bites |
| GO | Gene Ontology |
| GTP | Guanosine triphosphate |
| GDP | Guanosine diphosphate |
| GPCR | G protein-coupled receptor |

| | |
|---|---|
| HCL | Hydrogen chloride |
| HEPES | Hydroxyethyl piperazine-1-ethanesulfonic acid sodium salt |
| HPLC | High performance liquid chromatography |
| HPP | Hyperkalemic periodic paralysis |
| Hsp | Heat shock protein |
| IAA | Iodoacetamide |
| ICAT | Isotope-Coded Affinity Tags |
| IEF | Isoelectric focusing |
| IGF | Insulin-like growth factors |
| INF1 | Inverted formin-1 |
| IPG | Immobilized pH gradient |
| ISO | Isoproterenol |
| iTRAQ | Isobaric tags for relative and absolute quantification |
| KEGG | Kyoto encyclopedia of genes and genomes |
| LC | Liquid chromatography |
| LC3-II | Type II microtubule-associated protein 1 light chain 3 |
| LC_SRM | Liquid-chromatography selected reaction monitoring |
| MALDI | Matrix assisted laser desorption ionisation |
| MAPK | Mitogen-activated protein kinase |

| | |
|---|---|
| MHC | Major histocompatibility complexes |
| MIAPE | Minimum information about a proteomic experiment |
| MRM | Multiple reaction monitoring |
| MS | Mass spectrometry |
| MS/MS | Tandem mass spectrometry |
| NFAT | Nuclear factor of activated T cells |
| NFκB | Nuclear factor kappa B |
| NL | Non linear |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate buffer saline |
| PCR | Polymerase chain reaction |
| PEP | Posterior error probability |
| pI | Isoelectric point |
| PKA | Protein kinase A |
| PLB | Phospholamban |
| ppm | Parts per million |
| RF | Radio frequency |
| RNAi | RNA interference |
| RNA-Seq | RNA sequencing |

| | |
|---|---|
| SAGE | Serial analysis of gene expression |
| SCX | Strong cation-exchange chromatography |
| SDS | Sodium dodecyl sulphate |
| SILAC | Stable Isotope Labelling of Amino acids in cell Culture |
| SNK | Serum-inducible kinase |
| SR | Sarcoplasmic reticulum |
| STRAP | Serine-threonine kinasereceptor-associated protein |
| TCA | Trichloroacetic acid |
| TEMED | Tetramethylethylenediamine |
| Th1 | T helper 1 cells |
| Th2 | T helper 2 cells |
| TOF | Time-of-flight |
| UCSC | University of California Santa Cruz |

**Acknowledgements**

Firstly, I would like to express my gratitude to both Professor Jonathan Wastling and Dr Andy Jones for their encouragement, patience and help throughout my Ph.D. studies. I would like to extend this thanks to the members of both groups past and present. I have been extremely lucky to have worked with such great people. They have made coming to work each day a joy and without Sanya, Stu, Dong, Andy, Marnie, Hos, Gian-luca, Kat, Achchuthan, Faviel, Da, Ritesh, Sarah, Corrado and Poom my last four years would not have been so amazing. Not only have they made the experience very enjoyable, they have also been a great asset to the constant growth of my knowledge and practical skills. I have to make a special thank you to Nadine Randle, who has been sensational and without her I would not have had a thesis. I would also like to mention Emma Newsham and Sarah Vermont who have become very good friends and are just fantastic.

I have been immensely lucky throughout my studies to have met a range of different people and had the pleasure of working with several of them. I would like to thank each and every one of these people for being part of an incredible journey that has made me the person I am today. I would like to thank PFG for the use of their labs and equipment as well as their knowledge on all things mass spectrometry related. I would like to especially thank Yvonne Woolerton and Gemma Molyneux for always being there when I needed someone to talk to. On this note, I would also like to thank Maz Yon and Bryony Parsons for being amazing friends both in and out of work.

I am also extremely grateful to all my friends outside of work who have kept me going through the ups and downs of it all. A special thanks goes to all those I have lived with; Suz Martin, Jess Seddon, James Rolls, Paul Robinson, Jen Cooper and Indy Karpha for making home a great place to escape work. I would like to especially thank Natasha George and Simon West who helped me get through one of the hardest years and kept me sane. To all "the boys" for making sure I didn't become a complete geek, especially Rolf Groenewold for always being ready with the vodka and Rob Labinski for his unwavering belief in me and his constant

encouragement. To the Tripod girls: Ashley McKeown and Ruth Cotson for always being there with a glass of wine or a cuppa when I needed it. To the engineers who made my first year the best year of my life, but notably Lucy Hipkins and Jack Longbottom for always being there no matter the distance. To Becca Pearse, Jess Martin and Amber Appleton for their patience and kindness. To Holly Mcloughlin for just being a truly great and loyal friend. To Jo Tallis for making me feel like the most special person in the world. Last but not least to Melanie Couch, Alex Colby and Sophie Lee for being the best friends a girl could ask for and supporting me through thick and thin. Although I cannot name all my friends individually, I would like to express an unbelievable amount of thanks to every one of these wonderful people.

My last but most important thanks must go to my incredible family. I have been blessed with the most amazing amount of support and belief from all my family members. I would like to thank my Grandma for all her prayers, encouragement and belief but especially for giving me the opportunity to celebrate in style. I must also mention my Aunty Else who inspired me to believe I could achieve anything and I wish she was still here to see me complete the biggest achievement of my life. I have been lucky enough to be blessed with not only a fantastic brother who is always pleased to see me even when I turn up in tears but two phenomenal parents. I cannot put into words how much their love and support has meant to me. Without this I would never have even achieved my A Levels let alone write this thesis. To my amazing Baba I would like to thank you for always being there and for your stories involving the one eyed girl and lampposts. Lastly I will thank my best friend, my confidant, my mother. I would like to thank her from the bottom of my heart for always being there no matter what, for not caring how many times a day I called, for sitting up till the early hours of the morning to encourage me, for proof reading things she didn't understand but mostly for her love and kindness which never ends.

## Author's Declaration

The work presented in this thesis was performed solely by the author except where the assistance of other has been acknowledged.

Jenna Kenyani, May 2012

## Abstract

Salbutamol was synthesised in 1967 and is widely known now as a β-adrenergic agonist. It is known that β-agonists cause hypertrophy in various mammalian muscles cells, however the mechanisms by which this occurs are not fully understood. Recent studies have suggested that there is great potential for using β-agonists, specifically $\beta_2$-agonists, in restoring muscle mass and strength in humans. Although, salbutamol is commonly used as a bronchodilator, it has been identified as having an effect on muscle growth at a dose which is safe for humans. However, like all steroids there are many side effects associated with β-agonists. By understanding the mechanisms affected within muscle cells by treatment with salbutamol it will enable not only the understanding of muscle hypertrophy but lead to more specific dugs being designed.

The technical advances in mass spectrometry, sequencing and bioinformatics enable greater understanding of the interactions within many different cell and tissue types. However, combining these technologies to enable an understanding of the system as a whole is not yet possible. Modern techniques are able to acquire large amounts of information on the different components within a cell or organism but correlating, for example, gene expression and protein expression has not been achieved sucessfully. In this study, both proteomic and transcriptomic technologies have been utilized to model changes in the development of skeletal muscle cells *in vitro* in the presence of salbutamol in an attempt to gain a system wide understanding. The first aim was to identify novel proteins and pathways that interact with salbutamol, and could therefore be potential targets for β-agonists. The second aim was to investigate the correlation between different research areas and methodologies highlighting the importance of a system wide approach. Three proteomic methods have been applied, Differential In-Gel Electrophoresis (DIGE) and SILAC (stable isotope labelling), and label free peak intensity measurements. The data gained from these studies were compared to transcriptomic data produced via RNA-seq.

By investigating the proteomic changes within rat skeletal muscle using the DIGE technique a total of 31 proteins at 24 hr and 110 proteins at 96 hr were identified as being differentially modulated. This initial study allowed a general insight into the types of proteins and different pathways being affected within muscle cells. To continue to increase the knowledge known on β-agonists a study to sequence the RNA gathered from treated and control mouse muscle cells was undertaken. This reveled that between 161 and 2756 genes over four time points (2 hr, 6 hr, 24 hr and 96 hr) were statistically differently modulated. To investigate the proteomic changes further, two mass spectrometry-based techniques were employed. SILAC determined differences between control and treated protein abundance using stable isotopic labeling, whereas the label free technique compared peak intensities of corresponding peptides. The SILAC method identified 58 and 56 non-redundant proteins at 24 hr and 96 hr respectively. Whereas, the label free techniques identified 14 and 19 proteins at 24 hr and 96 hr respectively.

Overall, the techniques identified a range of different proteins and genes but with similar functions or participating within the same pathway. However, a much larger amount of genomic modulation was observed compared to the proteomic investigations, suggesting the process between gene transcription and protein expression cannot be explored fully using these techniques. This study also determined the need for more than one technique to be used in any study and that consideration over the bioinformatical analysis post biological experimentation is essential.

# 1 Chapter 1

# Introduction

## *1.1 System wide approach*

The need to understand the complexity, versatility, and robustness of each living system is the fundamental issue in current biological research. Within any cell there is constant state of organization and interactions between cellular networks that enable complex processes to occur. There are number of components, such as genes, proteins, and metabolites, in play which connect together allowing a variety of different processes and functions to be carried out. The high calibre of current technologies such as mass spectrometry and RNA sequencing has enabled a number findings relating to each of these components. For example it is now possible to sequence entire genomes, identify abundances in mRNA levels, determine relative and absolute quantities of proteins and observe the rate of protein turn-over as well as model possible interactions. However, without knowledge of the interplay between genes, proteins and metabolites understanding the system as a whole cannot be achieved.

For the last 50 years biological research has had a reductionist view, where the understanding of complex processes, such as mechanisms within a living cell, are determined by splitting them in to the individual components. Although, successfully identifying many of the components and interactions within a number of living systems, this approach offers little comprehension of how the system as a whole functions. This type of system wide perspective, often termed "systems biology" enables "cause and effect" in biological networks to be addressed by observing and quantifying multiple components simultaneously, as well as applying mathematical models to the data (Kitano, 2002).

However, this approach is difficult and complex and comes with a number of challenges. The ability to successfully identify and quantify all components at the level of mRNA, proteins and small molecular weight metabolites as well as correlate this data between all three areas; genome, proteome and metabolome, has proven difficult. Despite the difficulties, Ishii *et al.*, (2007) were able to perform a system wide analysis using techniques spanning the three main components in a cell; transcriptomics (mRNA), proteomics (protein) and metabolomics (metabolites). This study measured the global response of *Escherichia coli* to genetic and environmental changes and was the first to examine the relationship between these components and the functional state of a metabolic network. Although, Ishii *et al.*, (2007) were able to successfully use different techniques to investigate the global mechanisms within an organism other studies have not been as fortunate. For example, studies analysing the abundances of protein and mRNA have observed that the correlation is weak (Fu *et al.*, 2009; Gygi *et al.*, 1999b) suggesting that there are still limitations to overcome before a systems biology procedure can be obtained. Both biological factors and methodological constrains, as well as the computational ability for analysis have a role in the ability to correlate this data. The capability to integrate data from "omic" data sets obtained from different sources via computational methods to allow for seamless mapping of identified proteins to corresponding mRNA data is essential with the systems biology philosophy.

Overall, the major challenge that faces systems biology is the acquisition and interpretation of large quantitative data sets across multiple samples. Ultimately, systems biology will allow not only the functional and biological aspects of individual systems to be understood but allow biological systems to be reproduced in

terms of mathematical models and simulations, which enables further hypothesis and medical possibilities to be explored. The ability to acquire accurate quantitative data for all the components of a complex biological sample is essential. Current technologies allow a range of data sets to be collected from a number of species on a large scale. However, the capacity of each technology on their own to obtain all the data required for a systems biological approach is not sufficient at present. By adapting a systems biology approach to investigating a simple and well explored model organism, such as the mouse or rat, under treatment with stimulus of known function by using a number of techniques as well as exploring the bioinformatic analyses methods available to link the corresponding data it is hoped the biological advantages of system biology can be observed. It is also essential that any limitations of the current technologies are explored to encourage any problems encountered to be solved allowing the progression of both biological knowledge and technical advances to continue.

## *1.2 Adrenergic agonists*

In 1906, Dale, was the first to identify two distinctly different types of receptors for epinephrine, a sympathomimetic amine (Barger & Dale, 1910). Today these receptors are known as adrenergic receptors or adrenoceptors. In 1948, Ahlquist, classified these adrenergic receptors as either α-receptors or β-receptors based on the receptor mediating specific responses in different organs using five sympathomimetic amines (epinephrine, norepinephrine, isoproterenol, α-methylepinephrine and α-methylnorepinephrine).The effects of both naturally occurring and manmade sympathomimetic amines on the body have been studied since.

Sympathomimetric amines have been investigated since the 1920's when epinephrine was used in the relief of bronchospasms, however, it often had disagreeable side effects such as cardiovascular and central nervous stimulation (Bresnick *et al.*, 1949). Konzett, H. (1940 a, b) found that *N*-isopropylnoradrenaline (isoproterenol) dilated the bronchi and increased the heart rate and cardiac output in dogs. In the late 1940's studies were completed on the possibility of using isoproterenol for treating asthma (Segal & Beakey, 1947). Since then research has been ongoing until this day to try and find better compounds to act as bronchodilators without any harmful side effects.

Ahlquist, (1948) was the first to suggest the two types of receptors produce different responses. He suggested that constriction results from stimulation of α-adrenergic receptors in vascular smooth muscle, whereas dilation results from stimulation of β-receptors. Most work investigating brochodilation focuses on β-receptor agonists,

however, from these studies other effects of β-agonists were identified. One of the most interesting is their hypertrophic effect on muscle cells.

β-agonists produce a dramatic increase in skeletal muscle mass (Beermann *et al.*, 1987; Emery *et al.*, 1984; McElligott *et al.*, 1987; Reeds *et al.*, 1986) and a significant reduction in body fat content (Emery *et al.*, 1984). It is thought that this muscle hypertrophy is caused by the anabolic effect of these drugs by either stimulating protein synthesis or inhibiting proteolysis. They also have an influence on the contractile activity of muscle via the stimulation of the sodium/potassium pump (Buur *et al.*, 1982; Clausen & Everts, 1989; Tashiro, 1973). This hypertrophic effect seems to be selective for certain fibre types. Several studies have shown β-agonists to increase the cross-sectional area of fast muscle fibres (Criswell *et al.*, 1996; Ricart-Firinga *et al.*, 2000; Zeman *et al.*, 1988). This effect has a potential use in the prevention of muscle atrophy (Hinkle *et al.*, 2002; Pellegrino *et al.*, 2004) and the improvement of the efficiency of muscle growth in production livestock (Anderson D.B., 1991; Anderson, 1991; Mersmann, 1998; Moody D.E., 2000).

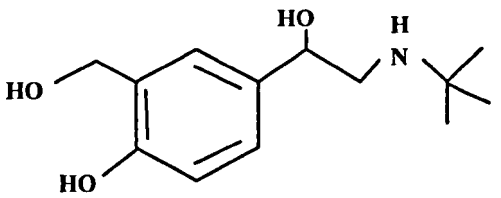
There are three classes of beta agonists; $\beta_1$, $\beta_2$ and $\beta_3$. The sub-division of the β-adrenoceptors was proposed by Lands and Brown (1967) was based on the varying responses of fat, heart and lung tissue as well as blood vessels to a number of agonists. The lipolytic response correlated well with cardiac stimulation and was classed as $\beta_1$, whereas bronchodilatation and vasodepression correlated with neither of these and were classed together as $\beta_2$. However, there were some β-agonists synthesised which stimulated the metabolic rate, adipose tissue thermogenesis, ileum relaxation, and soleus muscle glycogen synthesis (Arch *et al.*, 1984; Challiss *et al.*,

1988; Wilson *et al.*, 1984), but seemed to have a minimal effect at $\beta_1$ and $\beta_2$ sites. In 1989, a human gene was identified that encoded for a third adrenoceptor, classed as $\beta_3$ (Emorine *et al.*, 1989). The $\beta_2$ receptor agonists are of interest for this investigation, specifically a compound called salbutamol.

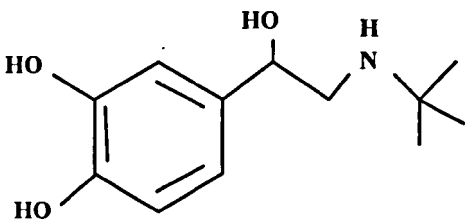
### 1.2.1 Salbutamol

In 1967 Larsen, *et al.*, synthesized a compound; 2-t-butylamino-1-(4-hydroxy-3-hydroxymethyl) phenylethanol (salbutamol, Figure 1.1), which had very selective action on bronchial muscle. This compound was made as an alternative to isoprenaline, which was used as a bronchodilator previously (Konzett, 1940, Stolzenberger-Seidel 1940), in an attempt to lessen side effects which arose from isoprenaline treatment. The addition of a $CH_2$ group to the compound, although small, would make a huge difference to the effect of salbutamol within the body. From the 1940's isoprenaline, was being used as a bronchodilator (Konzett, 1940, Stolzenberger-Seidel 1940), in an attempt to lesson side effects which arose from isoprenaline treatment. However it was only effective when given by aerosol or sublingual routes due to its rapid inactivation by catechol-O-methyl transferase (Ross, 1963). It was also found to cause distinct side effects because it stimulated the β-adrenoceptive receptors in the cardiovascular system. As previously discussed, Lands and Brown (1967) were the first to suggest that β-receptors can be divided into two different groups; $\beta_1$ and $\beta_2$. These two types of receptors reside in different organs within the body. It was proposed that a chemical that did not interact with the receptors in the cardiovascular system would be the best option for a bronchodilator, as the side effects would be reduced. Cullum *et al.*, (1969) went on to show that

salbutamol was more active on bronchial smooth muscle than on cardiac muscle as well as showing that it has greater selectivity for β-receptors in bronchi than any other β-receptors stimulant.




Salbutamol




Isoprenaline

**Figure 1.1** ***Chemical structures of two sympathomimetric amines used as broncodilators.*** The difference between the two compounds is the addition of a methylene ($CH_2$) group in salbutamol.

The $\beta_2$-agonists being used as treatment for asthma are thought to act by stimulating an increase in the intracellular level of cyclic adenosine monophosphate (cAMP), the second messenger that plays a large part in the inhibitory effect of cellular responses such as bronchial smooth muscle spasm and inflammatory mediator release (Lulich *et al.*, 1988). However, the mechanisms are still not fully understood.

Salbutamol is now widely used as a bronchodilator but it has recently been suggested as a possible way to improve the efficiency of muscle growth in humans and production livestock. Warriss, *et al.,* (1990) showed that salbutamol increases the quality of the meat in pigs, but did cause a slight toughening of certain muscles. Van Baak, *et al.,* (2000) identified that acute administration of salbutamol (4 mg) increased maximal isokinetic strength (4-5%) of the knee flexors and knee extensors. Collomp, *et al.,* (2005) also found that the same oral acute salbutamol intake improved peak power and mean power during supramaximal excerise in healthy men. Šoić-Vranić, *et al.*, (2005), identified that after administrating denervated and control rat soleus muscle with salbutamol for 14 days an increase in the cross-section for both type I and type II fibres can seen. This ergogenic effect has been related to the increased muscle tension production caused by enhanced sarcoplasmic reticulum (SR) $Ca^{2+}$ release after β-agonist stimulation (Cairns *et al.*, 1993). Slack, *et al.,* (1997) identified that phosphorylation of the SR pump regulatory protein phospholamban (PLB) in response to $\beta_2$-agonist stimulation enhanced relaxation rates within slow-twitch skeletal muscle.

Acute $\beta_2$-agonists can also cause increased stimulation in both twitch and titanic peak tension in intact slow- and fast-twitch skeletal muscle fibres in non-fatigue states (Cairns & Dulhunty, 1993; Cairns *et al.*, 1993). In intact skeletal muscle fibres, $\beta_2$-agonists act on some processes in the excitation-contraction coupling modulating force and contractility characteristics (Crivelli *et al.*, 2011). In this experiment the effects were seen after only three days of administration. Zhang, *et al.,* (1996) showed that salbutamol can up regulate genes in slow-twitch fibres that are normally expressed by fast-twitch fibres. Polla, *et al.,* (1994) also identified that β-agonists

stimulate a transition from slow to fast fibres. Pellegrino, *et al.,* (2003) noticed a change in mouse muscle fibre composition when treated with $\beta_2$-agonists. The agonist seemed to stimulate expression of fast *major histocompatibility complex* (MHC) -2B and -2X and repress the expression of MHC and MHC-2A.

There are several widely used $\beta_2$-agonists which all have similar effects to each other on the body. By comparing the research completed on all the $\beta_2$-agonists it helps to build a better picture of the mechanisms of action of these compounds.

### 1.2.2 Clenbuterol and other $\beta_2$-adrenergic agonists

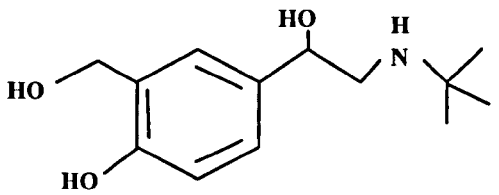
As previously discussed $\beta_2$-agonists are used as treatment for asthma and other conditions affecting the respiratory system. By analysing the methods of action for all the different $\beta_2$-agonists it is hoped a better picture will be gained of the mechanisms involved. It is thought that due to the presence of $\beta_2$-adrenoceptors on inflammatory cells $\beta_2$-agonists may act through these receptors to contribute to the beneficial effects on asthma. $\beta_2$-adrenoceptors are present on postcapillary venular endothelial cells. It has been shown that certain $\beta_2$-agonists can inhibit plasma exudation by preventing separation of endothelial cells in postcapillary venules (Baluk & McDonald, 1994; Bowden *et al.*, 1994). Formoterol, a long acting $\beta_2$-agonist, has been identified as reducing the adhesion of neutrophils and eosinophils to venular endothelial cells and so inhibiting the trafficking of granulocytes into the airway wall (Bowden *et al.*, 1994). There is also evidence to suggest that β-agonists may have an effect on mast cells *in vivo* because nebulized β-agonists have been seen to have a greater effect on AMP-induced bronchoconstriction than on histamine- or

methacholine-induced broncoconstriction (O'Connor *et al.*, 1992; O'Connor *et al.*, 1994). In addition, β-agonists can have an effect on peripheral blood monocytes causing an inhibitory effect on secretion of the cytokines TNF-α and GM-CSF (Seldon *et al.*, 1995; Seldon *et al.*, 1998). Peripheral blood lymphocytes express β-adrenoceptors as well and an increase in cAMP concentrations can occur due to treatment with β-agonists (Conolly & Greenacre, 1976; Kariman, 1980). Intracellular cAMP concentrations are also increased due to $\beta_2$-receptors being present on neutrophils (Galant *et al.*, 1980). This increase in intracellular cAMP results in reduced expression of the adhesion molecule Mac-1, which allows neutrophils to adhere to human airway epithelial cells (Bloemen *et al.*, 1997). It has also been shown that β-agonists interfere with leukocyte adhesion in the rat airway (Bowden *et al.*, 1994) and $\beta_2$-receptors have been found at high densities on human airway epithelial cells (Carstairs *et al.*, 1985; Davis *et al.*, 1990). Airway epithelial cells are able to secrete inflammatory mediators in asthma including nitric oxide, cytokines, chemokines and growth factors. Although it is known that $\beta_2$-agonists increase ciliary beating (Devalia *et al.*, 1992) it is not certain whether they also affect the secretion of any of these mediators. Salbutamol and salmeterol have also been seen to help reduce any increase in plasma histamines after exercise in mild asthma patients (Sichletidis *et al.*, 1993). Salbutamol can also inhibit the release of eosinophil peroxidase and leukotriene $C_4$ in human peripheral blood eosinophils *in vitro* (Munoz *et al.*, 1995). Eosinophils along with mast cells help control certain mechanisms associated with asthma.

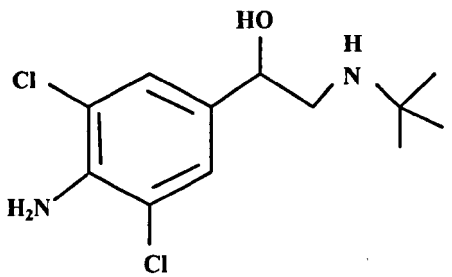
As discussed previously studies have shown that as well as being effective broncodilators, $\beta_2$-agonists also cause a very noticeable effect on muscle cells. There

has been a lot more research carried out on clenbuterol compared to salbutamol. Clenbuterol is well known to cause muscle hypertrophy. This effect is associated with an increased rate of synthesis of certain proteins, for example actin and myosin (Hesketh *et al.*, 1992; Navegantes *et al.*, 2004) and decreased rates of proteolysis of other proteins e.g. adenosine triphosphate (ATP)-dependant proteolysis (Busquets *et al.*, 2004; Yimlamai *et al.*, 2005) and calcium-dependant proteolysis (Navegantes *et al.*, 2001). In 1984, Emery, *et al.*, (1984) first reported significant increases in muscle protein synthesis in rats treated with clenbuterol by subcutaneous injections, which was verified by MacLennan and Edwards (1989). Clenbuterol treatment increases the proportion and size of fast fibres, as well as shifting the myosin heavy chain profile towards faster isoforms and elevates the levels of glycolytic enzymes resulting in larger, faster muscles (Criswell *et al.*, 1996; Polla *et al.*, 2001; Ricart-Firinga *et al.*, 2000; Zeman *et al.*, 1988). Clenbuterol is used to complement ventricular assist devices (Hon & Yacoub, 2003) and has also been proposed as a possible anticachectic agent (Baracos, 2001). Anticachectic agents help combat weight loss, wasting of muscle and general debility that can occur during a chronic disease. Clenbuterol has also been shown to reduce or restore skeletal muscle loss in animal models with muscle wasting such as denervation (Cockman *et al.*, 2001; Fitton *et al.*, 2001; Maltin *et al.*, 1986; Zeman *et al.*, 1987), genetic muscular dystrophy (Rothwell & Stock, 1985), bacterial endoxin injection (Choo *et al.*, 1989) and burn injury of 30% of the body surface area (Martineau *et al.*, 1993). Clenbuterol also enhanced the recovery rate of skeletal muscle mass and carcass protein lost because of surgical stress in senescent rats (Carter *et al.*, 1991). There are also studies showing salbutamol can also increase muscle mass and protein content in young rats (Choo *et al.*, 1992).

Both compounds work in similar ways though there is a worry that clenbuterol may have more harmful side effects. Choo, J.J. et al., (1992) compared equimolar doses of the two $\beta_2$-agonists. They found that salbutamol increased weight gain in the rats after four days of treatment but the increase was less than that produced by the clenbuterol treatment. An elevated heart mass and a reduction in liver and epididymal fat pad weights where observed in the clenbuterol treated animals but not in the animals treated with salbutamol. However, when chronic infusions of the two drugs were compared the increase in weight of gastrocnemius muscle and heart muscle, the increase protein and RNA content and the increase in the ratio of RNA to protein of gastrocnemius muscle were all almost identical. This suggests that salbutamol has a much shorter half life than clenbuterol *in vivo*. It was suggested that clenbuterol may be used to help combat muscle wastage. However, the dosage of clenbuterol needed to increase muscle mass in animal models (Carter *et al.*, 1991; Maltin *et al.*, 1986), when scaled up for estimated human dosage, exceeds the estimated safe dose for humans (Maltin *et al.*, 1992). Studies using salbutamol used a much lower dose, one which falls within the safe limit for humans (Price & Clissold, 1989). As can be seen from Figure 1.2, two hydroxyl groups and one hydrogen (H) atom on salbutamol have been substituted with two chlorine groups and an additional amine on clenbuterol which may explain the difference in side effects due to different binding capacities.




Salbutamol




Clenbuterol

**Figure 1.2** ***Chemical structures of two similar $\beta_2$ agonists.*** Clenbuterol has two chlorine groups and an extra amine group which results in a different chemical activity from salbutamol.

Many $\beta_2$-agonists have also been shown to increase muscle strength as well as muscle growth. To study the strength promoting effects of these chemicals, $\beta_2$-agonists were tested in young men by Martineau *et al.*, (1993). They were able to identify an increase in strength in both quadriceps and hamstring muscles after treatment. The study consisted of twelve healthy men who were given an oral sustained-release preparation of 8 mg salbutamol, twice daily for three weeks. The isometric strength of the quadriceps muscles increased 12% bilaterally after the second and third week of treatment. The dominant hamstring muscles increased in strength by 22% after three weeks. These drugs are known to be taken by body builders to increase strength and muscle mass. However, even though these compounds can increase muscle growth and strength it has been observed that there

is a decrease in exercise capacity after administration of large doses of clenbuterol (Duncan *et al.*, 2000; Ingalls *et al.*, 1996; Lynch *et al.*, 1996).

Studies have also shown that $\beta_2$-agonists can cause a reduction of the body fat content (Yang & McElligott, 1989) and have often been reportedly taken by celebrities as a weight loss drug. Both clenbuterol and salbutamol have been seen to reduce carcass fat content in young and old rats (Carter & Lynch, 1994). Clenbuterol has also been seen to have a significant repartitioning effect in most livestock species by increasing the growth of skeletal muscle at the expense of fat tissues (Anderson D.B., 1991; Moody D.E., 2000). In adipose tissue, the binding of β-agonists to adrenoceptors activates adenylate cyclise, cyclic AMP levels, the protein kinase cascade and eventually leads to the activation of hormone sensitive lipase and triacylglycerol hydrolysis (Yang & McElligott, 1989).

Clenbuterol can also induce cardiac hypertrophy in not only young rats but in foetuses of treated dams (Maltin *et al.*, 1990) which corresponds with evidence of placental transfer during human pregnancy (Pelkonen *et al.*, 1982). It seems that the result of drug treatment *in utero* is a reduction in myoblast proliferation and/or fusion of myoblasts which reduces the degree of hyperplasia in the neonatal muscles (Downie *et al.*, 2008; McMillan *et al.*, 1992).

Administration of clenbuterol also prevents the conversion of type II fibres to type I, which occurs during the process of maturation of rat muscle, due to the effect of the drugs on the increase in synthesis and accumulation of fast myosin light chains in

soleus muscle of young rats (Zeman *et al.*, 1988). The authors suggested that the drug may have the potential to override neurogenic effects during development. Šoić-Vranić, *et al.*, (2005) showed that under the influence of salbutamol there was a decrease in type I fibres in denervated soleus muscle compared to denervated untreated muscle. However, after the sixth day the number of type I fibres were higher in the treated muscle. It has been suggested that salbutamol re-established the process of growth and development in denervated muscles using the process of conversion of type II to type I fibres (Kugelberg, 1976).

As noted earlier the exact mechanisms of action of β-agonists in muscle cells are not fully understood. One of the most noticeable responses to acute β-agonist treatment is peripheral vasodilation, which increases blood flow to the skeletal muscles; this may provide extra substrates for muscle growth. However, this response is only temporary suggesting other processes are involved (Rothwell *et al.*, 1987). Insulin-like growth factors (IGFs) have been reported to play an important role in the hypertrophy of skeletal muscles (Awede *et al.*, 2002; Sneddon *et al.*, 2001). It is thought that β-agonist treatment may modulate IGF's to help increase muscle growth (Bhavsar *et al.*, 2010; Suzuki *et al.*, 1999). The role of IGF-I in muscle hypertrophy is well known (Adams & Haddad, 1996; Chen *et al.*, 1994; Turner *et al.*, 1988) but it is also thought that IGF-II plays a similar role. Sneddon, *et al.*, (2001) identified an increase in mRNA levels for IGF-II after 2 days of treatment with clenbuterol. There is also evidence that clenbuterol may act through specific targeting of protein kinase C isoforms (Downie *et al.*, 2008). Costelli, *et al.*, (1995) reported that clenbuterol administration to rats carrying a cachectic tumour was able to reverse muscle wastage through inactivation of the ATP-dependent proteolytic pathway. *In vitro*

studies in isolated skeletal muscle have shown that $\beta_2$-agonists stimulate the $Na^+$-$K^+$ pump enhancing cellular uptake of $K^+$ (Clausen & Flatman, 1977) and loss of $Na^+$ (Everts *et al.*, 1988). Stimulation via a β-agonist also alters the process of calcium transport across the plasma membrane and the intracellular calcium movements (Schmid *et al.*, 1985). It has also been identified that $\beta_2$-agonists can successfully counteract muscle paralysis in patients suffering from hyperkalemic periodic paralysis (HPP) (Hanna *et al.*, 1998; Wang & Clausen, 1976). HPP is caused by mutations in the gene encoding for the sodium channel expressed exclusively in the skeletal muscle (Cannon, 2002) which again suggests the β-agonists have an effect on other ion channels. Clenbuterol was also observed in blocking sodium currents in native skeletal muscle fibres (Desaphy *et al.*, 2003) further supporting the role of β-agonists on ion channels.

Smith, *et al.*, (1990) showed that $\beta_2$-agonists raised blood glucose levels by stimulating hepatic glucose output. However, β-agonists have also been seen to have an effect on homoeostasis, through direct effects on both brown adipocytes and skeletal muscle (Chernogubova *et al.*, 2005). Any adrenergic effects on glucose metabolism are of interest in skeletal muscle due to the adrenergic system having a large influence on whole body metabolism.

Although $\beta_2$-agonists have the potential to be of benefit for muscle wastage there is also evidence to suggest that use of these agents could lead to apoptosis of heart and soleus muscle. *In vivo*, clenbuterol leads to the activation of caspase 3, the externalization of phosphatidylserine and an increase in the sensitivity of myocyte DNA to heat denaturation and strand breaks which all indicate apoptosis (Dumont *et*

*al.*, 2000; Frankfurt & Krishan, 2001; Saunders *et al.*, 2000). It is thought that some β-agonists can accumulate in the heart (Soma *et al.*, 2004) which can cause effects such as apoptosis. However, studies on adult rat ventricular myocytes *in vitro* suggest that $\beta_1$-adrenoceptor stimulation is proapoptotic but $\beta_2$-adrenoceptor stimulation is antiapoptotic (Communal *et al.*, 1999; Zaugg *et al.*, 2000) which indicates not all β-adrenoceptors have an apoptotic effect.

It is thought that β-agonists may have an effect on the central nervous system as well as on muscles and it has been found that salbutamol stimulation in the rat brain had an antidepressant effect and increased serotonergic neurotransmission (Bouthillier 1991). Clenbuterol has also been shown to cross the blood-brain barrier to enhance the regeneration of motor neuron axons in motor degeneration mice (Zeman *et al.*, 2004), as well as promoting regeneration of peripheral nerves (Frerichs *et al.*, 2001) and to ameliorate denervation-induced atrophy in rats (Sneddon *et al.*, 2000).

These studies suggest that there is great potential for using $\beta_2$-agonists in restoring muscle mass and strength in humans affected by muscle wastage, especially elderly patients. However, to use these drugs effectively an increase in the knowledge of the mechanisms involved is needed. Although there is evidence suggesting that the agonists can act directly on the cells most studies show that they act via the beta-adrenoceptors.

### 1.2.3 Beta-adrenoceptors

The β-adrenoceptor (beta-adrenergic receptor) is a member of the 7-transmembrane family of receptors and is composed of 413 amino acid residues (Kobilka *et al.*, 1987). There are three types of receptors that respond to the three types of agonists; $\beta_1$, $\beta_2$ and $\beta_3$, usually identified in cardiac, airway smooth muscle and adipose tissue. Although they are predominantly found on these cells they can be found on almost every cell type. The $\beta_2$-adrenoceptor subtype is widely distributed in the lung, not only in the smooth muscle but also on other resident and infiltrating cells implicated in airway inflammation in respiratory disease. The $\beta_2$ subtype is also the predominant β-adrenoreceptor present in the skeletal muscle cells.

These receptors are G protein-coupled receptors and their physiological function works through phosphorylation of specific effects, such as ion channels, by cyclic AMP-dependent protein kinase (Desaphy *et al.*, 2003). The activation of this receptor can also be mediated via the increase in intracellular cyclic adenosine monophosphate (cAMP) levels via the $G_s$-protein. It is thought that these receptors can exist in both an active and inactive form. The $\beta_2$-receptor is in the activated form when it becomes associated with the α-subunit of the $G_s$ protein with a molecule of guanosine triphosphate (GTP) (Johnson, 2001). When GTP is replaced with guanosine diphosphate (GDP) it catalyses a conversion of ATP to cAMP, which reduces the affinity of the α-subunit for the receptor. Dissociation then occurs which returns the receptor to its low energy inactive form. One hypothesis on how $\beta_2$-agonists work is by binding to and temporarily stabilising the receptor in its active state (Onaran *et al.*, 1993).

The active receptor site, where the β-agonists must interact to exert their biological effects, lies about one-third of the way into the receptor core. Residues that have been identified as being critical for agonist binding to the active site are an aspartate (Asp)-residue 113, two serine (Ser) residues; 204 and 207 and two phenylalanines (Phe) 259 and 290 (Strader *et al.*, 1989). The β-agonist is anchored to the receptor by nitrogen atoms which interact with the aspartate, and by the hydroxyl groups on the phenyl ring interacting with the two serine residues.

All β-agonists have a β-OH group which causes the molecule to have an asymmetric centre. This centre results in a pair of optical isomers, R and S enantiomers, in a racemic mixture. The β-agonist/β-receptor interactions are stereospecific and the activity lays predominantly in the R-enantiomer. In salbutamol the R-enantiomer is a 100-fold more potent as a $\beta_2$-agonist that the S-enantiomer (Johnson *et al.*, 1993).

In smooth airway muscle, salbutamol, which is a hydrophilic molecule, acts on the $\beta_2$-adrenoceptor by entering the active site directly from the extracellular aqueous compartment (Johnson, 1992). Once the chemical binds to the active site there is a rapid response causing relaxation in the tissue allowing bronchodilation to occur. If receptor density is too low or coupling is inadequate, the $\beta_2$-agonist may behave in a partial manner meaning it may not be able to achieve the same maximal effect as an agonist of higher efficacy and may even behave like an antagonist. This is due to a β-agonist potency relying on its efficacy as well as its receptor affinity. Agonists with high efficacy are regarded as *full* agonists whereas a *pure* antagonist would have a low or zero efficacy. Most $\beta_2$-agonists have an intermediate efficacy. If receptor

numbers are high this will allow the agonists to behave as a full agonist (Johnson, 2001).

The receptors have a feedback mechanism whereby desensitisation of the receptor occurs, to prevent overstimulation of receptors in the face of excessive $\beta_2$-agonist exposure (van Koppen & Jakobs, 2004). The most common method for $\beta_2$-agonist-promoted desensitisation is phosphorylation of the receptor by the β-adrenoceptor kinase or other related G protein-coupled receptor kinases. This results in the partial uncoupling of the agonist-occupied receptor from the $G_s$ protein, limiting the receptor function. This process is transient and is reversed within minutes of the agonist being removed. Desensitization can also occur via cAMP-dependent PKA following phosphorylation of serine and threonine residues present within the third intracellular loop of the receptor in response to an increase in intracellular cAMP (Lohse, 1993).

More prolonged exposure to the agonist could cause internalisation of the receptor resulting in loss from the cell surface. It is usually possible for a full reversal to occur, although the reversal process will take longer. The cellular pathways and compartments that mediate internalisation are not fully understood. It has been suggested that this process results from rapid translocation of receptors from the plasma membrane to ensure the receptors are not degraded (Hertel *et al.*, 1983) and therefore, are recycled through this process (von Zastrow & Kobilka, 1992). It is known that β-arrestin plays an important role in receptor desensitization and internalisation (Figure 1.3). In internalisation, β-arrestin links the receptor to several proteins of the endocytotic machinery, including clathrin, producing clathrin-coated

vesicles (van Koppen & Jakobs, 2004). In the endosomes most of the receptors are dephosphorylated and returned to the plasma membrane as resensitised receptors. However, it has been seen that several G coupled-protein receptors do not need β-arrestin for internalisation (van Koppen & Jakobs, 2004).

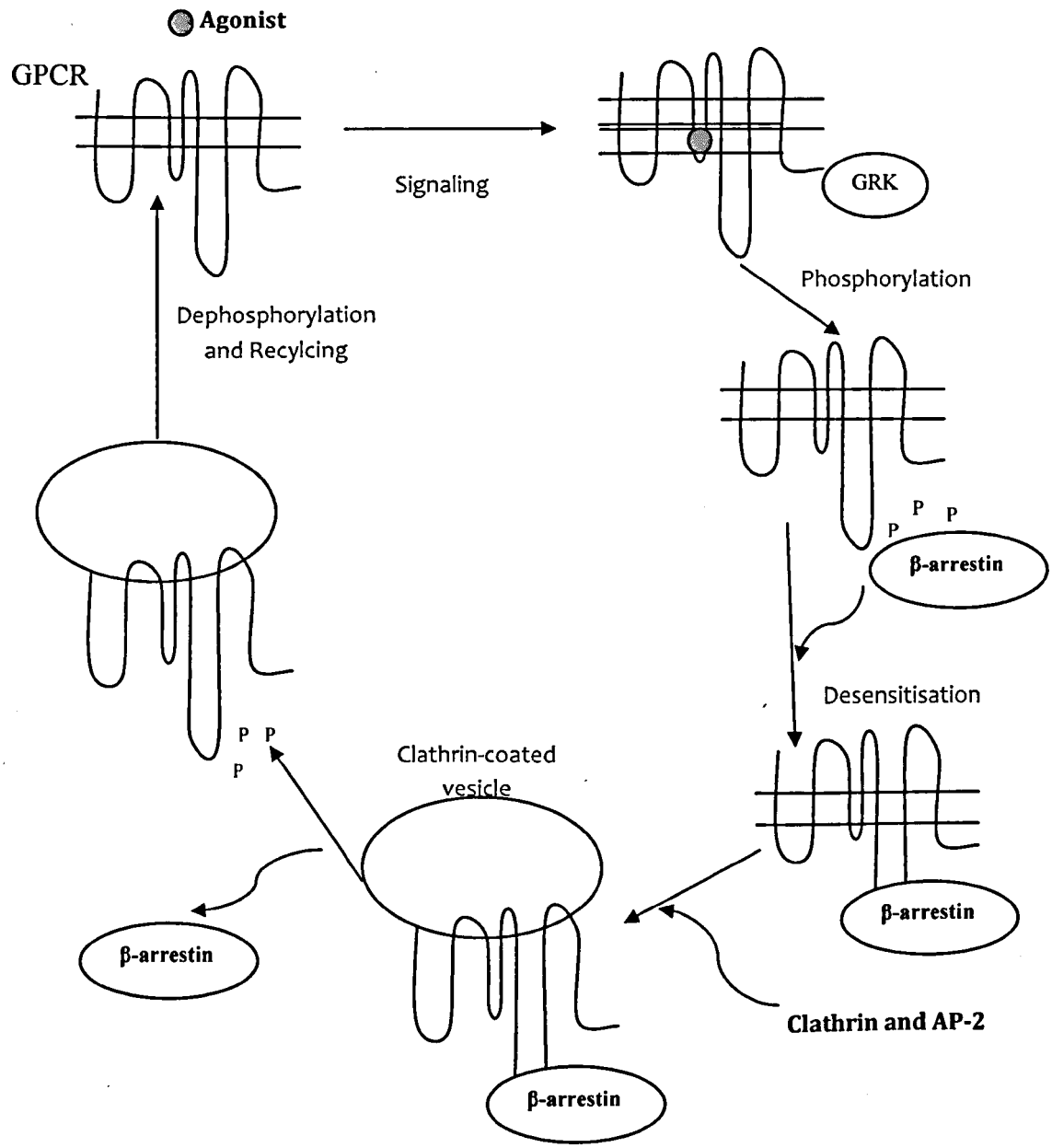



**Figure 1.3** ***Model of β-arrestin-regulated internalisation of a G protein-coupled receptor.*** If the agonist has bound to the G protein-coupled receptor (GPCR) for a prolonged period of time desensitisation will occur. This process starts with the phosphorylation of the GPCR by G protein-coupled kinases and other kinases (not shown). The β-arrestin is able to interact with the phosphorylated receptor to uncouple the GPCR from the G proteins. The β-arrestin then helps the process further by interacting with clathrin, chathrin adaptor complex AP-2 and several other endocytic components (not shown). This results in the immobilisation of the receptor within the clathrin-coated pit. Either before intracellular trafficking or after the clathrin-coated vesicle is formed the β-arrestin will dissociate from the receptor. After the release of this molecule the receptor is dephosphorylated by phosphatase and recycled back into the plasma membrane (van Koppen & Jakobs, 2004).

If the receptors are exposed to the agonist for hours then a net loss of cellular receptors occurs and transcription at the $\beta_2$-receptor gene level is required to restore the receptors. This process of desensitisation differs from tissue to tissue. It can be a big problem in airway tissue due to asthma patients often having to repeatedly use $\beta_2$-agonists for treatment over long periods of time. Airway tissue is especially affected as it is thought that inflammatory cells are more susceptible to desensitization. This may be due to a low rate in transcription within these cells, meaning that resynthesis of $\beta_2$-receptors is slow. They may also have less receptor reserve or an increased expression of β-adrenoceptor kinase, causing desensitization to be greater. Glucocorticords increase the transcription of the $\beta_2$-receptor gene in the rat and human lung and may prevent desensitization (Mak *et al.*, 1995a; Mak *et al.*, 1995b).

It is hypothesed that β-adrenoceptors play an important part in relieving asthma. A study by Kubota, *et al.,* (2000) identified that a reduced number of $\beta_2$-receptors may have an effect in the pathogenesis of exercise-induced asthma in children. Increased airway smooth muscle growth is an important pathogenic mechanism of asthma (Hirst *et al.*, 2004). It has been reported that β-agonists can inhibit cultured airway smooth muscle growth induced by various mitogens (Kassel *et al.*, 2008; Stewart *et al.*, 1997). Stimulation of the $\beta_2$-adrenoceptor by these agonists causes activation of the cAMP-dependent protein kinase (Penn, 2008). PKA activity has been shown to antagonize mitogenic signalling and growth in multiple cell types. It has also been demonstrated in several different cell types that G protein-independant signalling via the $\beta_2$-adrenoceptor can occur. Stimulation of the $\beta_2$-adrenoceptor is often mediated

by arrestins and causes either receptor tyrosine kinase or MAPK superfamily activation (Penn, 2008).

It is thought that there is a relationship between the autonomic and sympathetic nervous system and the immune system. The sympathetic nervous system helps regulate and modulate the differentiation and function of cells involved in the host immune response, often involving the $\beta_2$-receptor. All T lymphocytes, except the T helper 2 cells (Th2), possess $\beta_2$-adrenoceptors (Sanders *et al.*, 1997). The stimulation of these $\beta_2$-adrenoceptors appears to play a role in the proliferation of $CD4^+$, a sub type of T lymphocytes. These $CD4^+$ cells are able to differentiate under the influence of specific cytokines into either Th1 or Th2. The Th1 and Th2 cells are thought to predominantly, but not exclusively, govern humoral and cell-mediated host responses respectively (Wong, 2007). There is also evidence to suggest that the stimulation of the $\beta_2$-adrenoceptor of the B lymphocyte results in an amplification of the humoral host response. *In vitro* studies have also shown that stimulation of the $\beta_2$-adrenoceptors can reduce the production of the inflammatory cytokines IL-1, TNF-α, IL-6 and IL-8 from the whole blood and monocytes (Gu & Seidel, 1996; Zetterlund *et al.*, 1998).

As stated earlier these receptors are not just found in the lung but also on cardiac muscle. Stimulation of the β-adrenoceptor by β-agonists provides the most important regulatory mechanism for cardiovascular performance (Post *et al.*, 1999). In cardiac tissue the most primarily expressed β-adrenoceptor is the $\beta_1$ subtype (Bristow *et al.*, 1986). However, in cardiac fibroblasts the $\beta_2$ subtype is the most expressed receptor (Meszaros *et al.*, 2000). The simulation of these receptors on the fibroblasts has been

seen to promote DNA synthesis (Colombo *et al.*, 2003; Kim *et al.*, 2002) and modulate collagen secretion (Ostrom *et al.*, 2003).

However, Aránguiz-Urroz, *et al.*, (2010) identified that activation of the $\beta_2$-adrenoceptor lead to a strong autophagic response, which increases autophagosome formation and degradation. They identified the type II microtubule-associated protein 1 light chain 3 (LC3-II) as a critical component for autophagy induction when levels increased. Both isoproterenol and salbutamol increased the levels of LC3-II. However, it is unknown whether the role of autotrophy in the heart is detrimental or protective, although this may depend on the severity and duration of autophagy. Stimulation of β-adrenoceptors can also induce apoptosis in cardiac myocytes *in vitro* and *in vivo* (Singh *et al.*, 2001; Zaugg *et al.*, 2000). Menon, *et al.*, (2007) showed that stimulation of the β-adrenoceptor can increase the activity of glycogen synthase kinase-3β. The activation of this kinase plays a pro-apoptotic role in β-adrenoceptor-stimulated apoptosis via the involvement of the mitochondrial death pathway. There is evidence to suggest that there may be a fundamental difference between the $\beta_2$-adrenoceptor in skeletal muscle and those found in the heart. The skeletal muscle $\beta_2$-adrenoceptor appears to function more like a $\beta_1$-receptor than the $\beta_2$-receptor because it appears to stimulate cell death (Burniston *et al.*, 2005). The $\beta_2$-adrenoceptors of isolated cardiomyocytes are known to signal through both G protein $\alpha_s$ and the G protein $\alpha_i$ pathways, with protein $\alpha_i$ signalling being predominant (Xiao *et al.*, 1999). Whereas it is likely that the $\beta_2$-adrenoceptor in skeletal muscle signals predominately through the stimulation of the G protein $\alpha_s$, in a similar way to the cardiac $\beta_1$-adrenoceptor.

Activation of $\beta_2$-adrenoceptors has also been shown to increase glucose uptake in rodent skeletal muscle cells and skeletal muscle cell lines (Nevzorova *et al.*, 2002). These receptors seem to affect glucose uptake via cyclic AMP and P13K (Nevzorova *et al.*, 2006). Yamamoto, *et al.*, (2007) identified that $\beta_2$-adrenoceptors can induce glycogen synthesis and phosphorylation of glycogen synthase kinase 3 as well as effecting glucose uptake. These β-agonists can also affect glucose homeostasis through direct effects on both brown adipocytes (Chernogubova *et al.*, 2005) and skeletal muscle (Ngala *et al.*, 2008).

However, it is thought that not all effects of β-agonists are mediated through the β-adrenoceptors. There has also been pharmacological evidence suggesting there may be an atypical β-adrenoceptor that mediates the effects of clenbuterol (Challiss *et al.*, 1988; Maltin *et al.*, 1989; Reeds *et al.*, 1988; Sillence & Matthews, 1994). However, Hinkle, (2002) showed that clenbuterol lost its effect on skeletal muscle when $\beta_2$-adrenoceptors were removed.

### 1.2.4 Applications of beta agonists

The main application currently of β-agonists is for the treatment of asthma. However, there seems to be great potential for using these compounds in increasing the strength and growth of muscles. Presently there are few therapeutic approaches for the reversal of muscle wasting, a condition that affects a significant number of people worldwide. It would also be advantageous to patients suffering from muscle injuries. Skeletal muscles can be injured as a consequence of extrinsic or intrinsic events. Skeletal muscles have an ability to regenerate after injury, however this repair is

often slow and incomplete and, once complete, the muscle often has reduced functional capacity and can be prone to re-injury (Ryall *et al.*, 2008). Using β-agonists would speed up this process. There have also been trials where β-agonists have successfully counteracted muscle paralysis in patients suffering from hyperkalemic periodic paralysis (Hanna *et al.*, 1998; Wang & Clausen, 1976).

However, some data suggests that even though these agonists may be beneficial for some therapeutic therapies they also have a detrimental effect on cardiac and skeletal muscles even at low doses (Burniston *et al.*, 2005). Although salbutamol can help reverse muscle wastage its effects on the heart are problematic when treating elderly patients with underlying heart conditions. The mechanism of $\beta_2$-agonists induced anabolism is unclear, though the evidence suggests there is direct involvement from the $\beta_2$-receptors (Hinkle *et al.*, 2002). By identifying the precise cellular activity involved in muscle hypertrophy it is hoped that enhancement of drug development for this condition will occur. A greater understanding is needed into the effects of β-agonists within the cells before they can be used to their full potential within medicine. By using a number of techniques, such as proteomics and transcriptomics this study hopes to gain further knowledge of the mechanisms of salbutamol within skeletal muscle cells. An introduction to these techniques and others that could be used is given in the following sections.

## *1.3 Proteomics*

In 1996, Wilkins *et al.,* first proposed the concept of the proteome which was defined as the complete protein complement expressed by a genome. As proteins are extremely important functional components of a cell, the recent growth within the proteomic area has allowed further knowledge in protein pathways and interactions that would not have been possible using genomic data alone. Proteomics now not only focuses on which proteins are present in a cell but on protein turn-over, protein interactions and pathways as well as any post-translational modifications. Using this proteomic data it is becoming possible to gain a greater understanding of the functioning of a cell. A variety of proteomic techniques will be essential to this project in order to understand how muscle cells react when treated with anabolic agents. It is necessary to investigate key aspects such as: the relative differences in protein expression; changes in absolute protein expression and turn-over; and alterations in post-translational states of proteins. This data along with transcriptional data will be used to provide a firm basis for identifying key cellular components and pathways involved in the response to anabolic agents.

To undertake a comprehensive evaluation of protein function requires a high calibre of techniques that span the different areas of interest. One of the most important aspects of proteomics is accurate quantification of measured changes in protein expression. At present there is no one technique that fills all the needs for high resolution proteomics. It is often difficult to decide which proteomic techniques to use for an experiment, as such a wide variety are available today. There is also a

worry that different techniques can create bias in the biological results. In this investigation more than one technique will be used to study this point.

### 1.3.1 Mass spectrometry

Mass spectrometry is vital to most proteomic techniques as it enables the identification of peptides, and therefore the proteins, along with any modulation occurring within a sample. This technology works by measuring the mass-to-charge ratio (m/z) of charged molecules allowing determination of a mass of a particle, the elemental composition of a sample and for identifying the chemical structure of molecules. Mass spectrometers are comprised of three different components: an ion source that converts molecules into gas-phase ions, a mass analyzer that separates charged molecules according to their m/z and a detector that records the number of ions at each m/z value. There are several different types of ionization sources, for example, electrospray ionisation (ESI) and matrix assisted laser desorption ionisation (MALDI) which can be utilised for current proteomic techniques, each one having its own advantage and disadvantage (Ashcroft, 1997). The sample being analysed can either be introduced directly into the ionisation source or can undergo chromatography before entering the ionisation source to separate complex mixtures. There are a number of mass analysers currently available, including quadrupoles, time-of-flight (TOF) analysers and quadrupole ion traps. Some mass spectrometer instruments have more than one analyser and are termed tandem (MS/MS) mass spectrometers. Briefly, tandem MS involves the ionisation of the sample followed by separation by m/z allowing selection of precursor ions. These undergo a form of

fractionation and run through the second analyzer to determine the m/z values for the product ions.

## 1.3.2 Gel based Methods

### 1.3.2.1 Difference in gel electrophoresis

Two-dimensional gel electrophoresis (2-DE) is a well established technique that separates proteins: firstly, according to their isoelectric point and secondly, by their molecular weight. However, there are limitations to this method; often a lack of reproducibility is seen and comparing results between gels is often unreliable. This is because no two gels are identical due to inhomogeneities in the polyacrylamide gels, electric and pH fields and thermal fluctuation, which affects the ability to accurately match spots.

A modified version of this method developed in 1997 by Unlu, *et al.*, known as Difference In Gel Electrophoresis (DIGE), is more commonly used for observing differences in protein expression between two or more samples of interest. DIGE pre-labels different samples, such as treated and control samples, with two cyanine dyes (Cy2, Cy5) to enable both samples to be run on the same gel. This diminishes some of the problems seen with 2-DE as both samples will experience the same conditions and there is no variability between gels. The dyes have been manufactured so that they match the charge of the protein residue that they have modified and all three dyes have similar molecular weights and charge. They also possess distinctive fluorescence spectra allowing spots to be identified through fluorescence imaging. An internal standard is also used and is created using pooled aliquots of all the

samples in the experiment and labelled with Cy3. This is run on the same gel, allowing the abundance of each spot in the two different samples, to be expressed as a ratio compared to the internal standard. For accurate spot matching, spot quantification and statistical analysis, software such as DeCyder can be used. Once spots of interest have been identified they are excised and digested before being analysed by a mass spectrometer to obtain identifications.

The advantage of this method is its ability to compare two different samples on the same gel, which diminishes any variability. Also the DeCyder software used in conjunction with the DIGE experiments allows univariate statistical testing, meaning the analysis of the data is more reliable than that of gel-to-gel comparisons (Nelson *et al.*, 2008). However, this procedure also has its disadvantages due to the fluorescent dyes being expensive and the experiment being very labour intensive.

## 1.3.3 Non-gel based Methods

### 1.3.3.1 OFFGEL fractionation

Ros, *et al.,* (2002) developed a method which allows separation of either proteins or peptides by their isoelectric point without the use of a one dimensional polyacrylamide gel electrophoresis (1D-PAGE) called OFFGEL fractionation. Due its ability to fractionate at high resolution but also to recover the sample in-solution it is beneficial for mass spectrometry and chromatography as the sodium dodecyl sulphate (SDS) from gels can interfere with the down-stream analysis.

The OFFGEL device contains either 12 or 24 chambers which are open at the top and bottom extremities. These chambers are placed on top of an immobilized pH gradient (IPG) strip where isoelectric focusing takes place. As there is no fluidic connection between the chambers, proteins and peptides are forced to migrate through the IPG strip where separation takes place when an electric current is applied to the strip. Once the protein or peptide reaches the chamber in which the pH matches its pI it loses its charge and can be recovered from the solution by precipitation (Horth *et al.*, 2006; Lam *et al.*, 2007). Poor recovery of the fractionated samples post precipitation is the disadvantage of this technique.

However, an advantage of this technique is the ability to combine it with either isobaric tags for relative and absolute quantification (iTRAQ) or stable isotope labelling of amino acids in cell culture (SILAC) to allow for quantitative proteomics. Pre-fractionation is important due to the complexity of most human and animal models. By fractionating the samples before mass spectrometry analysis it has been demonstrated that more information about the proteome content can be gained. This fractionation technique combined with a labelling method can be used instead of DIGE to allow analysis without the bias of gel fractionation.

### 1.3.3.2 Isotope-coded affinity tags

Another non gel based method that is used to gain quantitative protein profiles is the Isotope-Coded Affinity Tag (ICAT) reagent method (Gygi *et al.*, 1999a). The reagents used consist of: a reactive group with specificity towards thiol groups; an isotopically coded linker; and an affinity tag which is used to isolate ICAT-labelled

peptides (Shiio & Aebersold, 2006; Tao & Aebersold, 2003). The reagent exists in two forms: the original heavy form that contains eight deuteriums and the light form that contains none. This gives the two different forms an 8 Dalton (Da) mass difference that can be detected using a mass spectrometer. Another version of the ICAT reagent uses carbon-13 instead of deuterium to create a heavy reagent, which gives this cleavable reagent a mass difference of 9 Da. Also in this type of reagent the affinity tag is connected to the rest of the molecule by an acid-cleavable linker. The use of carbon-13 promotes co-elution of heavy and light ICAT-labelled peptides that had partial separation on reverse-phase chromatography as a result of isobaric shift in the original ICAT reagents (Shiio & Aebersold, 2006).

This method labels the samples from two different cell states with either the heavy or light reagent. Any cysteine residues in the proteins are labelled, thus greatly reducing the complexity of the peptide mixture prior to analysis because not all peptides will contain a cysteine (Kirkpatrick *et al.*, 2005). The two samples are mixed and then digested to generate peptides. This, however, is a limiting factor, as any protein that does not contain cysteine will not be monitored. The ICAT-labelled peptides are then isolated by avidin affinity chromatography and finally separated and analysed using mass spectrometry. The relative quantification is deduced from the ratio of signal intensities of pairs of peptide ions of identical sequences that are tagged with the heavy or light forms of the reagent, which differ in mass. By combining the data collected by MS and MS/MS analysis relative quantities as well as sequence identities can be determined (Gygi *et al.*, 1999a; Shiio & Aebersold, 2006).

### 1.3.3.3 Isobaric tag for relative and absolute quantification

Another alternative to using gel based methods for quantitative proteomics utilises MS as well as stable isotope labelling. Isobaric tag for relative and absolute quantification (iTRAQ) is a quantification technique that uses four isobaric amine specific tags. These isobaric tags are placed at the N termini and lysine side chains of peptides in a digest mixture (Ross *et al.*, 2004). Once labelling has occurred, the samples are fractionated using cation exchange. This is followed by analysis of samples using LC-MS/MS. Each tag generates a unique reporter ion, 114.1, 115.1, 116.1 and 117.1 mass to charge ratio (m/z): by comparing the intensities of the four reporter ions in the MS/MS spectra, protein quantification can be achieved (Shadforth *et al.*, 2005). The total mass for each tag is identical, which increases the sensitivity. This allows the combination of two or more samples without there being an increase in spectral complexity. This extra sensitivity is also seen at the MS/MS level due to the peptide backbone fragment ions being isobaric. This technique can also be used for absolute quantification of targeted proteins, using labelled synthetic peptides of a known value and measuring the intensity of the reporter ion for this peptide giving an absolute value which corresponds to the known amount added. However, one potential drawback to this technique is that MS/MS spectra must be acquired, which will then require more analysis time than performing result-dependent analysis only on differentially expressed peptide pairs in MS (Ross *et al.*, 2004). However, another advantage of using this system is that the multiplex nature of the experiment allows the removal of any quantitative variability from chromatography that may be seen in sequential two-dimensional HPLC-MS analyses of individual peptide mixtures (Washburn *et al.*, 2003). Also, due to the relative non-specific nature of the labelling, more proteins are identified by multiple peptides.

This allows an increase in confidence in the ratio reports because quantification of multiple peptides occurs (DeSouza *et al.*, 2005).

#### 1.3.3.4 mTRAQ

The mTRAQ methodology is a nonisobaric variant of iTRAQ, which was developed to address some of the problems experienced using iTRAQ, for example, it's ineffectiveness in detecting specific proteins in a complex sample as well only providing relative quantification. The mTRAQ technique applies a method called multiple reaction monitoring (MRM) to target tryptic peptides from proteins of interest. This method performs absolute quantification based on the isotope-dilution mass spectrometry principle, which has been designed to take advantage of the MRM mode. The mTRAQ labels are nonisobaric, which allows the increase in possible differences in the MRM transition (Ross *et al.*, 2004). There are two chemically identical versions of the mTRAQ label. The lighter version is 141 Da, 4 Da lighter than the lightest iTRAQ label and does not include $^{13}C$ or $^{15}N$. The heavier version is identical to iTRAQ 117 label and has a mass of 145 Da. A known quantity of a synthetic peptide, which has an identical sequence to that of a selected tryptic peptide from the protein of interest, is labelled with one version of the mTRAQ tags. The second version of the tag is used to label the tryptically digested sample homogenate. These two differently labelled samples are then mixed together and analysed using either one-dimensional or two-dimensional liquid chromatography. Due to the differences in mass of the two tags, the MRM transitions that are chosen for a peptide labelled with one of the two versions are specific for that version. The quantification of the peptide in the digested sample, as well as the quantification of

the amount of protein in the original sample, is determined by calculating the areas under the MRM traces for the respective tags (DeSouza *et al.*, 2008).

### 1.3.3.5 The absolute quantification strategy

The absolute quantification (AQUA) strategy is a method used for quantitative analysis of proteins and post-translational modifications. This technique is designed to specifically look for phoshorylated proteins giving it an added advantage to other techniques. The difference between this method and other quantification methods is that it relies on measurements made at the peptide level. The first stage of this method involves introducing stable isotope labelled internal standard peptides into a protein sample during proteolytic digestion (Kirkpatrick *et al.*, 2005). The only difference between the internal standard and the target peptide are their masses: the labelling produces a shift in mass, which is detected by mass spectrometer (Gerber *et al.*, 2003). Liquid-chromatography selected reaction monitoring (LC_SRM) is used to measure the AQUA peptide and the unlabelled peptide. SRM is a method used in tandem MS/MS in which an ion of a particular mass is selected in the first stage of a tandem MS and in the second mass spectrometry stage an ion product from the fragmentation reaction of the precursor ion is selected for detection.

The second stage of the AQUA strategy is to quantify the native peptide from cell lysates. A 1D-PAGE is commonly used to fractionate whole cell lysates; the region of the gel where the protein migrates to is excised to perform further analysis (Kirkpatrick *et al.*, 2005). In-gel digestion is performed in the presence of the AQUA peptide and then LC-SRM analysis (Gerber *et al.*, 2003). Due to an exact amount of

the internal sample being added, the ratio between the internal standard and the levels of peptide or the phosphorylated form can be determined by the mass spectra (Gerber *et al.*, 2003; Kirkpatrick *et al.*, 2005).

For the quantification of phosphoylated proteins, the method is the same, except that a phosphorylated internal standard peptide is generated as well as a non-phosphylated internal standard peptide (Gerber *et al.*, 2003). The phosphorylated peptide is synthesized with a phosphorylated serine, threonine or tyrosine in addition to the stable isotope labelled residue. The total amount of protein and the extent of phosphorylation at that modification site can then be determined (Kirkpatrick *et al.*, 2005). However, a disadvantage to this method is the need to know the position the protein will migrate to on the gel.

#### 1.3.3.6 Stable Isotope Labelling by Amino Acids in Cell Culture

Stable Isotope Labelling by Amino acids in cell Culture (SILAC) is a labelling method that uses isotope labelling of amino acids in growth medium to allow quantification of proteins used in cellular processes of the target organism. Labelling occurs during the growth of the cell cultures by using medium deficient in certain essential amino acids that cells cannot synthesis themselves. Isotopically labelled analogs are synthesized separately and added to the medium and as these labelled anologs are supplied instead of the natural amino acids they are incorporated into each newly synthesised protein chain. There is no chemical difference between the labelled amino acids and the natural ones enabling the cell to behave like the control cell population (Ong *et al.*, 2002).

Due to there being no difference between populations the cells can be treated in specific ways, allowing a comparison between treated and control populations. These populations can then be harvested, and because the label is encoded directly into the amino acid sequence of every protein, the extracts can be mixed. Quantification is possible in exactly the same way as any other isotope labelling method. The advantage of this method compared to ICAT is that the labelling is not restricted to proteins containing cysteine residues, which is not as abundant as other amino acids. Another advantage is that as the proteins are uniformly labelled several peptides from the same protein can be compared to identify whether the change in modulation is accurate (Ong *et al.*, 2002). Although SILAC was originally designed for use in mammalian cells it has now been adapted for use in a number of species (Jiang & English, 2002; Kerner *et al.*, 2005; Nirmalan *et al.*, 2004).

This method can also be applied in detecting changes in phosphorylation states (Blagoev *et al.*, 2004). Park *et al.*, (2006), combined this method with immunopreciption to identify phosphorylation sites for the ion channel Kv2.1. Another use of this method is to study protein turnover as well as differential expression. The most common experimental design using SILAC compares cellular states in a binary comparison, however, it is often beneficial to undertake multiplex experiments by using three isotopically distinct forms of arginine and lysine.

A possible drawback to this technique is the unintended conversion of some amino acids used in the labelling processes, creating amino acids that could affect the quantification of samples. Several studies have shown that conversion of isotopically coded arginine to labelled proline can occur (Blagoev & Mann, 2006;

Hwang *et al.*, 2006; Ong *et al.*, 2002; Ong & Mann, 2006; Schmidt *et al.*, 2007; Van Hoof *et al.*, 2007). This can result in the relative abundance observed being affected as well as complicating the mass spectrometry data by increasing the number of peptide ion peaks. There are several possible ways to minimise this problem; reducing the concentration of arginine in the SILAC media which means it is metabolically unfavourable as a precursor for proline synthesis (Blagoev & Mann, 2006), avoiding arginine altogether (Hwang *et al.*, 2006), making mathematical corrections post-analysis and supplementing SILAC media with standard L-proline (Bendall *et al.*, 2008). The latter of these suggestions seems to prevent the conversion of arginine without affecting the results.

#### 1.3.3.7 Label-free techniques

By removing the labelling step errors in quantification arising from the shortcomings of the different labelling techniques are eliminated. There are currently two very different protein quantification strategies used for label-free quantification: a) spectral counting where the number of fragment spectra identifying a given peptide are counted and compared (Wang *et al.*, 2008) and b) peptide chromatographic peak intensity measurements in which the mass spectrometric signal intensity of peptide precursor ions from a particular protein are measured and compared (Bantscheff *et al.*, 2007; Wang *et al.*, 2008).

Spectral counting is based on the observation that if there is more of a particular protein present, there will be more MS/MS spectra for the peptides of that protein. Relative quantification can therefore be calculated by comparing the number of

spectra between experiments. An advantage to this technique is that both protein identification and protein quantification can be achieved simultaneously with extensive MS/MS data collected across a chromatographic time scale. This technique is controversial due to it not measuring any direct physical property of a peptide. It also relies on the spectra quality and any error would lead to inaccurate quantification. Another disadvantage to this technique is that it only provides relative quantification and assumes that all peptides will respond identically during analysis as well as being detected equally well during MS. Old *et al.*, (2005) proved that often peptides will not behave as anticipated by showing that in one experiment 90% of the total peptides eluted in a single strong cation-exchange (SCX) chromatography fraction.

Peptide ion intensity measurements integrate mass spectrometric peak areas for every peptide over the chromatographic time scale. The intensity values for each peptide can be compared between one or more experiments to gain relative quantification. An advantage to this strategy is the possibility of distinguishing between distinct peptide sequences with interfering signals of similar masses due to the zoom scan, which allows a narrow mass window to examine ions, and MS/MS data. However, a special algorithm is required to align LC-runs prior to identifying corresponding peptides. The chromatographic profile needs be optimised to allow for reproducibility and for corresponding peptides between experiments to be found without too much difficulty (Bantscheff *et al.*, 2007). An alternative method has been proposed which involves the mass spectrometer no longer cycling between MS and MS/MS mode, but instead to detect and fragment all peptides simultaneously by rapidly alternating between high- and low-energy conditions in the mass

spectrometer (Bantscheff *et al.*, 2007). Label-free methods are probably the least accurate within the mass spectrometer quantification strategies but cheaper than labelling peptides. Also, it does not restrict the amount of experiments that can be compared.

### 1.3.4 Bioinformatics analysis of proteomic data

Bioinformatics has become an integral part of proteomics, which has become more dependent on bioinformatics for storing proteomic data in databases, as well as for protein identification and interpretation (de Hoog & Mann, 2004). Databases are at the core of bioinformatics, they are required to build a system that is able to capture, store, retrieve and analyse all data concerned (Dhingra *et al.*, 2005). There are many databases available which have different advantages and disadvantages. These databases use mass spectrometry spectra to identify peptide fragments. For each proteomic technique there are often one or two specific pieces of software which allow analysis of the spectra and identification and quantification of the proteins present. These software packages not only use the different databases but also have their own algorithms and settings for identification and quantification. Without this technology proteomics would not be a powerful tool, however, as each package is slightly different the choice of package to use could result in bias. Once identifications have been made and any proteins of interest identified there are again a number of pieces of software which have the ability to use the list of proteins to determine biological processes, cluster proteins and map to pathways.

### 1.3.5 Biological implications of proteomics in investigating mammalian muscle cells

Proteomic techniques currently available are enabling the understanding of cellular processes on a protein level and are utilised for a number of different studies, including those investigating the muscle proteome. Burniston *et al.,* (2007) employed a 2-DE method to investigate the anabolic effect of clenbuterol on rat plantaris muscle. They were able to successfully identify five proteins modulated after treatment; heat shock protein 72, β-enolase, aldolase A, adenylate kinase and phosphoglycerate mutase. Stella *et al.* (2011) were able to identify 19 non-redundant differentially expressed proteins including actin, myosin and beta-enolase, when analysing the effects of clenbuterol on cattle muscle using a DIGE method. Cui *et al.,* (2009) were able to successfully use a SILAC method to identify differentially expressed proteins between rat myoblasts and myotubes. In this study a total of 400 proteins were identified as differentially expressed. Using a different labelling technique, ICAT, Toigo *et al.,* (2005) investigated the subproteome of atrophying mouse skeletal muscle and were able to identify 62 modulated proteins. The iTRAQ technique has also been used to examine the muscle proteome. For example, Hakimov *et al.,* (2009) used this form of labelling to confidently catalogue 542 proteins in porcine muscle. Warren *et al.,* (2010) combined iTRAQ with OFFGEL fractionation to study the cardiac proteome. Label free techniques have also been combined with OFFGEL fraction as shown by Fraterman *et al.,* (2007), who were able to identify 61 proteins that were differently expressed between two extraocular muscle types. Salem *et al.,* (2010) were able to identify a total of 146 significantly modulated proteins in atrophying muscles in rainbow trout using the peak intensity label free method. Although a number of proteomic techniques have been used to investigate the mechanisms with muscle cells on a protein level, most only utilise

one technique per investigation. This may limit the data gained as each technique has limitations and may not be able to analyse the complete proteome. In this study more than one proteomic technique will be employed to hopefully increase the knowledge on the effect of salbutamol on muscle cells. These techniques will also be compared to each other in an attempt to ascertain how well the different techniques to correlate with each other.

## 1.4 Genomics

Genomics is able to give us an insight into the complex nature of all organisms, however, there is an absence of a direct functional correlation between gene transcripts and their corresponding proteins (Dhingra *et al.*, 2005). Along with proteomics data it is important to gain further knowledge on mechanisms within muscle cells and identify any differences and similarities in the transciptomic data. Relatively recently there has been a need to sequence entire genomes, which has driven innovative developments to allow for sequencing technology to be faster, cheaper and more accurate.

### 1.4.1 Next generation technologies

In the late 1970's several publications came out that set the groundwork for sequence-driven technologies. The most notable publication was by Sanger and colleagues in 1977 on the development of the chain termination method (Sanger *et al.*, 1977). Until recently Sanger sequencing has been the most commonly used DNA sequencing technique. However, in recent years there have been a lot more commercially available techniques. Next generation sequencing makes it possible to obtain a much more detailed picture of the transcriptome or genome of any species. Currently the high-throughput technologies available that are most commonly used include: Roche's 454, Illumina's Genome Analyzer and Applied BioSystem's SOLiD (Zhao & Grant, 2011). These technologies also allow for applications related to target region deep sequencing, epigenetics(ChIP-seq), transcriptome sequencing (RNA-seq) and megagenomics. These technologies offer opportunities to increase our understanding of the functions and dynamics of the genome.

#### 1.4.1.1 Sanger sequencing

This method is based on the DNA polymerase-dependant synthesis of a complementary DNA strand in the presence of natural 2'-deoxynucleotides (dNTPs) and 2'3'-dideoxynucleotides (ddNTPs), the latter of which serve as non-reversible synthesis terminators (Sanger, 1977). Fluorescent detection allows the four terminators (ddATP, ddCTP, ddTTP and ddGTP) to be combined into one reaction by having them labelled with different fluorescent dyes. Whenever a ddNTP is added to the growing oligonucleotide chain the DNA synthesis is terminated. This results in truncated products of various lengths with an appropriate ddNTP at their 3' terminus. The products are separated by their size using capillary gel electrophoresis and the terminal ddNTPs are used to identify the DNA sequence of the template strand.

This technique has a lot of advantages and has a relatively low error rate. However, it is an expensive technique which has allowed other technologies to be developed. Another limitation of Sanger sequencing is the requirement of *in vivo* amplification of DNA fragments that are to be sequenced, which is usually achieved by cloning into bacterial hosts. However, this step is prone to host-related bias and is quite a lengthy and labour intensive step. The 454 technology was marketed as a way of bypassing this.

In this study the next generation sequencing technology will be utilised to identify changes occurring within the genome under the influence of β-agonists. As previously stated it is extremely important to understand the mechanisms within a cell at both the genomic and proteomic level which is why both areas will be investigated.

### 1.4.1.2 454 sequencing technology

This technique avoids the cloning requirement by using a highly efficient *in vitro* DNA amplification method called emulsion PCR. In emulsion PCR the DNA is fragmented and attached to streptavidin beads using adapters. Each bead contains only one DNA fragment and is captured into separate emulsion droplets. Each droplet acts as an individual amplification reactor producing around $10^7$ clonal copies of unique DNA template per bead (Margulies *et al.*, 2005). The beads are then transferred into a picotiter plate and clonally related templates are analysed using a pyrosequencing reaction. Pyrosequencing is a sequencing by synthesis technique that measures the release of inorganic pyrophosphate (PPi) by chemiluminescence. The template DNA is immobilized and solutions of dNTPs are added one at a time. Whenever the complementary nucleotide is incorporated the release of PPi is detectable by light produced by a chemiluminescent enzyme present in the reaction mix. The DNA sequence of the template is determined from a "program" which corresponds to the order of correct nucleotides that that been incorporated. The pyrosequencing approach is prone to errors as a result of incorrectly estimated lengths of homopolymeric sequence stretches.

### 1.4.1.3 Illumina

The Illumina approach is able to achieve cloning-free DNA amplification by attaching single-stranded DNA fragments to a solid surface known as a single-molecule array and conducting solid-phase bridge amplification of single-molecule DNA templates (Bennett *et al.*, 2005; Bentley, 2006). Briefly, this process involves one end of a single DNA molecule being attached to a solid surface using an adapter. The molecule is then bent over and hybridized to complementary adapters forming

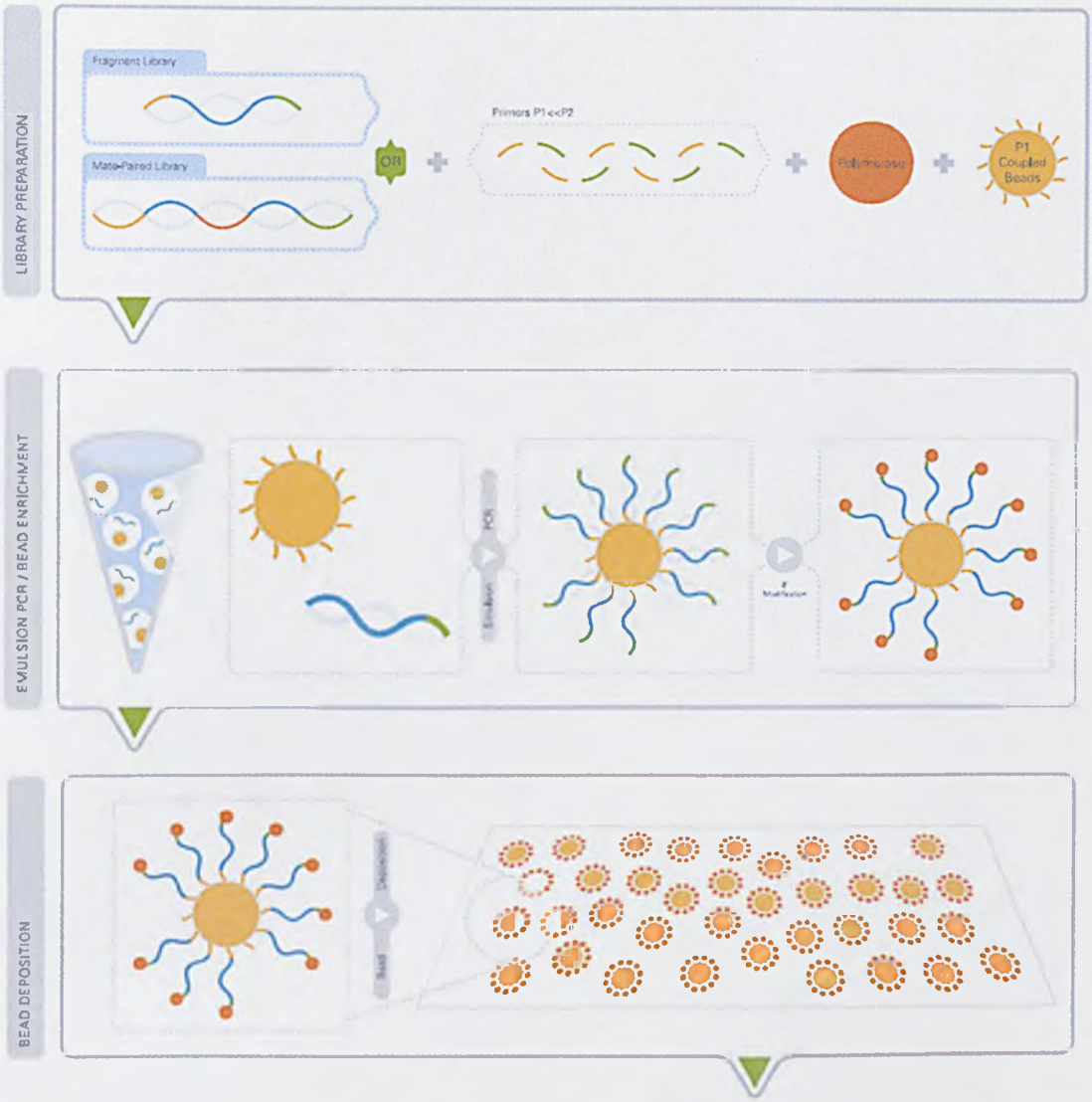
the "bridge" and so creates the template for the synthesis of the complementary strands. After the amplification step more than 40 million clusters are produced; each cluster is composed of approximately 1000 clonal copies of a single template molecule. The templates are sequenced in parallel using a DNA sequencing-by-synthesis approach that uses reversible terminators with removable fluorescent moieties and special DNA polymerases that can incorporate these terminators into growing oligonucleotide chains. The terminators are labelled with four different fluorescent dyes to allow the different bases to be distinguished at any given sequence position. The template sequence of each cluster is identified by reading off the colour at each successive nucleotide addition step.

This approach is more effective at sequencing homopolymeric stretches than pyrosequencing. However, it does produce shorter sequence reads so cannot resolve short sequence repeats.

#### 1.4.1.4 SOLiD sequencing

Massively parallel sequencing by hybridization-ligation, implemented in the supported oligonucleotide ligation and detection system (SOLiD) is the most recent technique available. The ligation chemistry used in this approach is based on the colony sequencing technique that was published in the same year as the 454 method (Shendure *et al.*, 2005). Constructing the sequencing libraries starts with an emulsion PCR single-molecule amplification step which is similar to the one used in the 454 method. The amplification products are transferred to a glass surface where sequencing occurs by sequential rounds of hybridization and ligation with 16

dinucleotide combinations labelled with four different fluorescent dyes. Each position is probed twice and the identity of the nucleotide is determined by the colour seen which results from two successive ligation reactions. The two-base encoding system enables the distinction between a sequencing error and a sequencing polymorphism. A diagram outlining the process undertaken for SOLiD sequencing is presented in Figure 1.4.


LIBRARY PREPARATION
Fragment Library
Mate-Paired Library
OR
Primers P1<<P2
Polymerase
P1 Coupled Beads
EMULSION PCR / BEAD ENRICHMENT
PCR
Emulsion
BEAD DEPOSITION
Bead Deposition
1. Prime and Ligate
PRIMER ROUND 1
POH
Ligase
Universal seq primer (n)
AT
3'
TA
P1 Adapter
Template Sequence
2. Image
Excite
Fluorescence
3. Cap Unextended Strands
Phosphatase
4. Cleave off Fluor
Cleavage Agent
HO
AT
P
TA
5. Repeat steps 1-4 to Extend Sequence
Ligation cycle 1 2 3 4 5 6 7 ... (n cycles)
6. Primer Reset
Universal seq primer (n-1)
2. Primer reset
1. Melt off extended sequence
7. Repeat steps 1-5 with new primer
PRIMER ROUND 2
1 base shift
-1
Universal seq primer (n-1)
AA CA CG TC AA TA CC
T GT GC AG TT AT GG

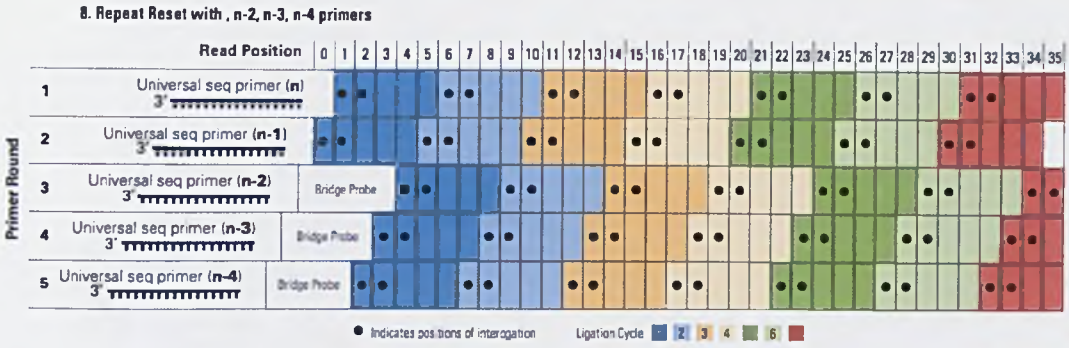



**Figure 1.4** ***Diagram describing the process of SOLiD sequencing.*** The sequencing is based on ligation chemistry. The process starts with constructing the sequencing libraries using emulsion PCR. The amplification products are transferred to a glass surface where sequencing occurs by sequential rounds of hybridization and ligation with 16 dinucleotide combinations labelled with four different fluorescent dyes. Each position is probed twice and the identity of the nucleotide is determined by the colour seen which results from two successive ligation reactions. Images taken from Life Technologies, Carlsbad, CA.

### 1.4.2 Transcriptome sequencing

When a gene is expressed in a cell, its code is transcribed to an intermediary messenger RNA (mRNA) which is then translated into a protein. The identification and quantification of mRNA in cells under different conditions or in different cell types has long been of interest. The sequencing of cDNA rather than genomic DNA allows the analyses to focus on the transcribed portion of the genome. Transcriptome sequencing can be used for gene expression profiling and genome annotation.

One method for achieving gene expression profiling and genome annotation is using microarrays where cDNA is hybridised to an array of complementary oligonuclotide probes which correspond to genes of interest. The abundance of a particular mRNA

is estimated from the hybridisation intensity to the probe (Schena *et al.*, 1995). However, the detection of genes on a microarray is limited due to the microarrays being designed for specific genomic annotations. Any non-annotated or any incorrectly annotated genes will be missed. A second method uses sequenced cDNA fragments to count the number of times a particular fragment is observed to measure abundance (Velculescu *et al.*, 1995). Direct sequencing is independent of any annotation allowing novel transcripts and slice variants to be measured. With many advantages including high resolution, the next generation RNA-seq is quickly replacing microarray methods for transcriptomics. The only downside to this new technology is the amount of data produced creating a challenge on a bioinformatics level to deal with the large files and extract the relevant data.

## 1.5 Aims and Objectives

Studies into β-agonists are increasing due to their interesting effects on muscle cells as well as their better known use as bronchodilators (Bresnick *et al.*, 1949; Emery *et al.*, 1984). Salbutamol is the $\beta_2$-agonist that is of interest in this study due to its ability to increase muscle growth at a lower dose than other $\beta_2$-agonists making it safe for human use (Price & Clissold, 1989). Although there is a large amount of knowledge about the effects of $\beta_2$-agonsits, the mechanisms by which they work within the cells are poorly understood. By understanding the mechanisms of the $\beta_2$-agonist; salbutamol, within skeletal muscle cells it is hoped to not only help combat the side effects of this drug but also to be used as a treatment for muscle wastage.

One aim of this project is to use proteomic and genomic technologies to model changes in the development of mammalian skeletal muscle cells *in vitro* in the presence of salbutamol, and to identify proteins and pathways within these cells that interact with, and could therefore be potential targets for these agents.

The second and most important aim of this project is to determine the ability for different techniques correlate to allow the system as a whole to be examined and not just on one biological level. There are a number of techniques available with the capacity to determine the protein and gene expression in a growing organism as well as a variety of down-stream bioinformatic software packages that allow for further analysis. However, it is unclear how successfully the results obtained using these techniques and packages correlate with each other, not just between different fields such as proteomics and transcriptomics but within the same field. This study hopes to

determine whether a systems biology approach can be undertaken with the techniques currently available.

This study of potentially very interesting aims will be achieved by:

- Using a variety of non-gel and gel-based proteomic techniques to analyse the protein expression, biological processes and pathways affected in response to exposure of salbutamol.

- Using next generation sequencing will to identify modulations within the transcriptome after treatment with the β-agonist.

- Comparing the transcriptomic and proteomic data to further understand the mechanisms of salbutamol as well as determining how efficient the correlation between gene expression and protein expression is.

- Undertaking a bioinformatic analysis of all data sets using a variety of programmes. The resulting identifications and quantifications from each software package will be compared to determine if any bias occurs by using different approaches.

It is hoped that this investigation will highlight the importance of utilising a number of techniques to gain a greater insight into the mechanisms of salbutamol within muscle cells as well as the consequences of not fully understanding how technologies integrate with each other.

# 2 Chapter 2

# DIGE analysis of the modulation of the primary rat cell proteome following treatment with salbutamol

## 2.1 Introduction

One of the main aims of this project is to identify modulation occurring within the muscle proteome due to the treatment of salbutamol. Gel based separation is a technique that is still frequently used within proteomics to identify proteins on a quantitative level (Choi *et al.*, 2010; Kosako & Nagano, 2011; Tafelmeyer *et al.*, 2008; Thomas *et al.*, 2011). Two-dimensional gel electrophoresis (2-DE) is a well established technique that separates proteins: first, according to their isoelectric point and then by molecular mass (Kenrick & Margolis, 1970). It is also an approach that when combined with other techniques can produce different types of data, as well as the relative abundance of differentially expressed proteins. For example, when combined with immunoblotting it can be used to identify the existence of multiple forms of a protein (Celis & Gromov, 1999). This method can also be used to reveal any post-translational modifications within the sample. However, there are several limitations to this method; often a lack of reproducibility is seen and the comparison of results between gels is often unreliable. No two gels are identical due to inhomogeneities in the polyacrylamide gels, electric and pH fields and thermal fluctuation, which affect the ability to accurately match spots between different gels.

A modified version of this method, called difference gel electrophoresis (DIGE), is more commonly used for observing differences in protein expression. In DIGE, treated and control samples are labelled with two cyanine dyes (Cy3, Cy5) to enable both samples to be run on the same gel (Unlu *et al.*, 1997). This diminishes some of the problems seen with 2-DE as both samples will experience the same conditions and there is no variability between gels. The dyes have been manufactured so that

they match the charge of the protein residue which has been modified and all three dyes have similar molecular masses and charge. They also possess distinctive fluorescence spectra allowing spots to be identified through fluorescence imaging which has equal sensitivity to silver staining. An internal standard, created using pooled aliquots of all the samples in the experiment and labelled with Cy2, is also used. This is run on the same gel, allowing the abundance of each spot on the control, or treated sample, to be expressed as a ratio compared to the internal standard. Computer software allows these ratios to be measured and statistical analysis to be performed. Four pairs of biological replicates, which are run on four different gels at the same time, are needed for each experiment to minimize any biological variations in the statistical test. (Karp *et al.*, 2004) This method relies on the ability to statistically determine the extent of differential protein expression between control and treated samples accurately. A specifically designed piece of software for DIGE using the internal standard experimental design called DeCyder is used to identify proteins of interest and perform the statistical analysis (Yan *et al.*, 2002). Figure 3.1 shows an overview to a DIGE method.

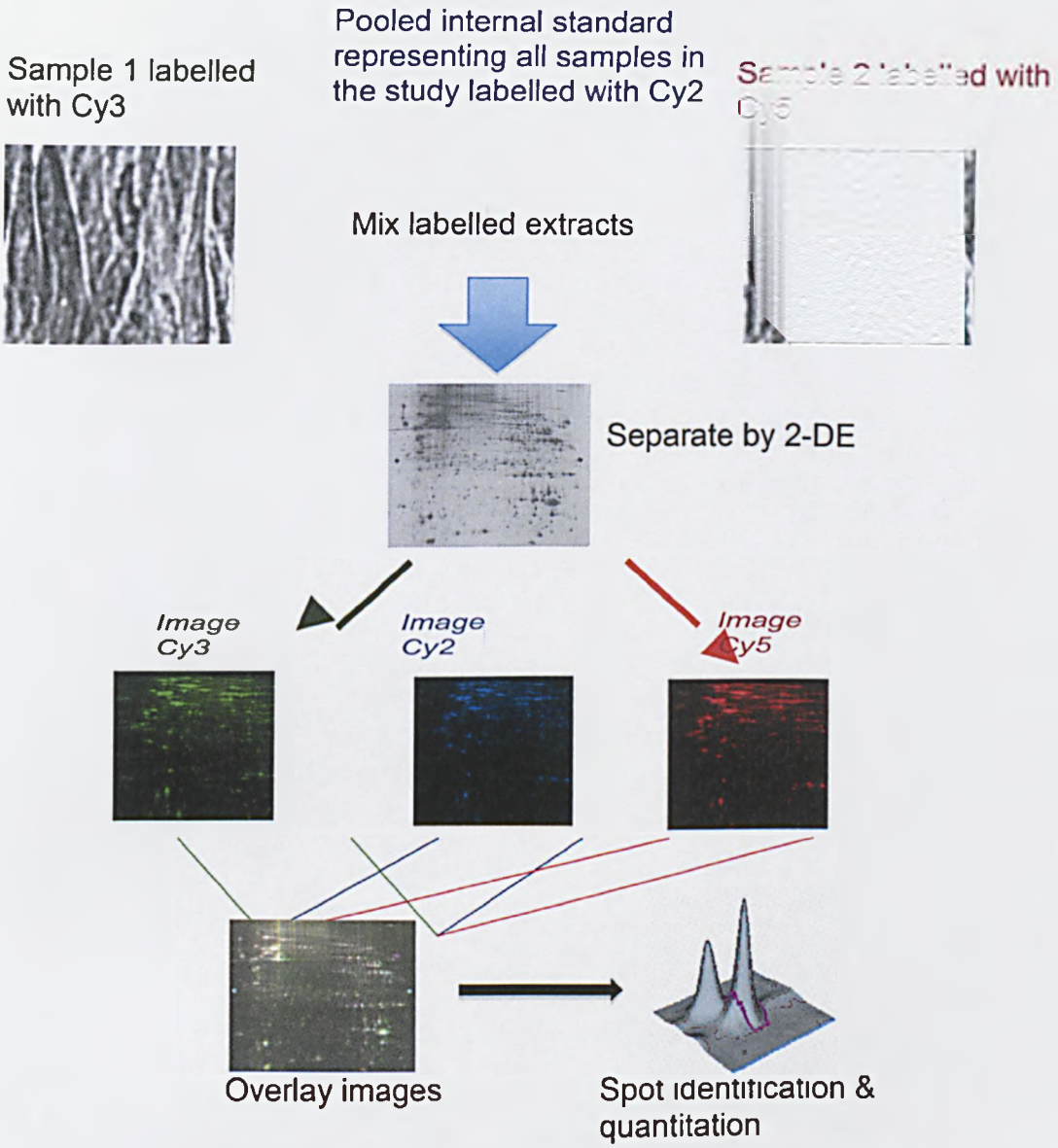



**Figure 2.1 Overview of DIGE workflow**

Once primary skeletal muscles cells have been grown and harvested at Pfizer they can be lysed and the protein labelled with one of the CyDyes. A different dye is used for control and treated. A pooled sample is labelled with Cy2. All three samples can be run on the same gel. Once the gel as been run and scanned a piece of software called DeCyder can be used to identify if any of the protein spots differ between conditions.

### 2.1.1 Beta agonists

Salbutamol and clenbuterol are two examples of $\beta_2$-adrenergic agonists. Studies have suggested that there is great potential in using $\beta_2$-adrenergic agonists in restoring muscle mass and strength in humans affected by muscle wastage especially elderly patients (Carbo *et al.*, 1997; Emery *et al.*, 1984; Herrera *et al.*, 2001; Maltin *et al.*, 1993). However, the dosage of clenbuterol needed to increase muscle mass in animal models, when scaled up for estimated human dosage, exceeds the estimated safe dose for humans (Bantscheff *et al.*, 2007). Salbutamol is most commonly used as a bronchodilator, but when its ability to increase muscle mass was investigated it was observed that much lower dose, which falls within the safe range, is needed to produce the same effects (Ricart-Firinga *et al.*, 2000). The underlying mechanisms of the $\beta_2$-agonist induced anabolism are unclear, though there is evidence to suggest there is direct involvement from the $\beta_2$-receptors (Hinkle 2002, Downie 2008). By identifying the precise cellular activity involved in muscle hypertrophy it is hoped that enhancement of drug development for this condition will occur. A greater understanding is needed into the effects of β-agonists within the cells before they can be used to their full potential within medicine.

It can be hypothesised that this study will highlight proteins associated with the cytoskeleton, cell growth and energy production as β-agonists are known to cause hypertrophy within muscle cells. The three most abundant muscle proteins are actin, myosin and titin (Fulton & Isaacs, 1991). As titin is a large protein that will not enter conventional 2-DE gels this protein will not be identified in the experiment, but the other two proteins will most likely be seen as modulated due to their role as structural proteins.

Previous experiments using 2-DE have been undertaken to analyse the mechanisms of β-agonists. Burniston *et al.,* (2007) identified five proteins modulated; heat shock protein 72, β-enolase, aldolase A, adrenylate kinase and phosphoglycerate mutase, within rat plantaris muscle after treatment with clenbuterol. Using SDS-PAGE gels and western blotting, Oishi *et al.,* (2004) also identified Hsp 72 as being modulated by clenbuterol in rat skeletal muscle. More recent studies have identified heat shock proteins, including Hsp 72, as having an important role in functional overload in muscles (Huey *et al.*, 2010). It is likely that heat shock proteins will be identified in this study due to the cells being put under stress while growing from the treatment. A recent experiment using 2-DE found that proteins within the functional classes of signalling, hydrolase, cell organization, transcription, translation, chaperone, $Ca^{2+}$-binding, energy metabolism and immune response are modulated in cardiomyocytes treated with isoproterenol (Hong *et al.*, 2011). Although this study is using heart muscle, the proteome of heart and skeletal muscle is very similar (Raddatz *et al.*, 2008), it can be hypothesised that similar proteins and functional classes will be affected by salbutamol, as the drug's mechanisms should resemble those of other β-adrenergic agonists.

Studies completed with a range of other methods have identified possible mechanisms of the β-agonists. It is already known that β-agonists interact with G-coupled β-receptors. The stimulation of the G-protein by the β-agonists leads to the production of cyclic adenosine monophoshate (cAMP) and activates adenyle cyclase. It is hoped that this experiment may shed some more light on the role of the receptor via modulations in related proteins. Shi *et al.,* (2007) determined the involvement of both the PKA signalling pathway, MAPK signalling pathway and ERK signalling in

the hypertrophic process in fast-twitch muscle fibres caused by clenbuterol. It is conceivable that proteins within these pathways will be detected within a DIGE experiment. Although testosterone is not a β-agonist it has the same side effects and mechanisms within muscles. Wu *et al.,* (2010) found that testosterone-induced hypertrophy is dependant also on the activation of *erk.*

### 2.1.2 Aims

In this chapter, a DIGE analysis of the effects of salbutamol on the rat skeletal muscle cell proteome was undertaken. The aim of using this approach was to give a simple overview of the effects on the cell, before going on to use other methods to give a deeper understanding and quantification of any proteins identified within this study. The first aim is to grow a cell culture system to determine the optimal concentration of salbutamol needed for an effect to be seen. This will be followed by identifying the differences in both cell morphology and at a protein level between control and treated samples. The last aim will involve analysing different pieces of software available for pathway analysis. Any results gained from this experiment would be used to direct further investigations as well as helping to gain insight into the mechanics of salbutamol.

## *2.2 Method and Materials*

### 2.2.1 Cell Culture

Primary rat skeletal muscle cells, prepared by Pfizer, were transferred into cell growth medium (DMEM, 20% (v/v) FCS, 1.4% (v/v) antibiotic-antimycotic solution). Treatment occurred after three days of incubation, when cell layers were approximately 90% confluent. First each well was washed with 5 ml phosphate buffer saline (PBS) (137mM NaCl, 2.7mM KCl, 10mM $Na_2HPO_4$, 2mM $KH_2PO_4$, pH 7.2). Following this, 4.5 ml of cell fusion medium (2.5% DHS (v/v), DMEM and antibiotic-antimycotic solution) was added to each well. On each six well plate 500 µl of $10^{-4}$ M salbutamol, dissolved firstly in dimethyl sulfoxide (DMSO) then added to a DMEM solution, the final concentration of DMSO being 1% (v/v), was used to treat three of the wells. The other three wells were used as control wells and had 500 µl of 1% (v/v) DMSO in DMEM solution added. Cells were incubated for different time lengths: 24 hr and 96 hr. For each time point six plates were collected and the wells washed with PBS three times and frozen at -80°C. Images of cells (Figure 2.1) at different time points under both conditions were provided by Adrian Thompson, Pfizer.

### 2.2.2 Sample Preparation

#### 2.2.2.1 Lysis of cells

One ml of lysis buffer (30mM Tris/H-Cl, 2M thiourea, 7M urea, 4% (w/v) CHAPS) was added to each well and scraped thoroughly before being transferred to a sterilized 1.5 ml tube. Samples were treated to a freeze/thaw process four times, and

then centrifuged at 16000 x g for 10 min and the pellet containing any insoluble protein was discarded. For each time point and condition four biological replicates were collected.

#### 2.2.2.2 Protein Assay

Protein concentration was determined using the Bio-Rad Protein Assay. Standards of 10, 20, 30, 40 and 50 µg/ml were made using a 1 mg/ml stock solution of bovine serum albumin (BSA). Samples were diluted 1 in 10, 1 in 50 and 1 in 100. A 100 µl aliquot of standards, blanks and samples were all made in duplicate and placed into the appropriate wells within a 96 well plate. Coomassie Plus Reagent was diluted 1 in 4 with water and 200 µl was added to each well. The plate was left at room temperature for between 5 min and 1 hr before using a Multiskan Ascent microplate photometer (Thermo Scientific) to read the absorbance at 540 nm. The software plotted a graph of the absorbance readings for the standards verses the concentration of protein in each standard, which was used to determine the concentrations of the samples using their absorbance readings.

#### 2.2.2.3 Acetone Precipitation

The required volume, to equal 100µg of each sample, was transferred to a 1.5 ml tube. Five volumes of ice cold acetone were added to the tube, vortexed and incubated overnight at -20°C. Samples were centrifuged at 2000 x g for 5 min at 4°C before carefully removing the supernatant and leaving the pellets to air dry.

## 2.2.3 Protein separation

### 2.2.3.1 Mini 2D-gel electrophoresis

#### *2.2.3.1.1 Rehydration*

To the precipitated protein pellet, 130 µl of rehydration buffer (8M urea, 2% (w/v) CHAPS and 0.002% (w/v) bromophenol blue) including 0.02% (w/v) DTT and 0.02% (w/v) IPG buffer pH 3-10 (GE Healthcare), was added. This was left at room temperature for 1 hr with occasional mixing. To remove any insoluble material the sample was centrifuged at 16000 x g for 3 min. An aliquot of 125 µl of rehydration buffer sample mix was run along the bottom of the rehydration tray and the 7 cm pH 3-10 non linear (NL) IPG strip (GE Healthcare), positioned with the gel face down, was placed on to the rehydration tray. The strip was covered with DryStrip cover fluid and the lid to minimize evaporation. This was then incubated over night at room temperature to rehydrate.

#### *2.2.3.1.2 First dimensional Isoelectric focusing (IEF)*

The rehydrated strip was placed on an Ettan™ IPGphor II™ (GE Healthcare) following the manufacturer's handbook. Strips were focused at 50 µA per strip following a five step protocol involving; 3 hr Step at 300 V, 3 hr Gradient at 600 V, 3 hr Gradient at 1000 V, 3 hr Gradient at 8000 V and 4 hr Gradient at 8000 V.

#### *2.2.3.1.3 Second Dimensional SDS-PAGE*

The strip was equilibrated in 10 ml of SDS-equilibration buffer (2% (w/v) SDS, 50mM Tris-HCL pH 8.8, 6M urea, 30% (v/v) glycerol and 0.002% (w/v)

bromophenol blue) plus 1% (w/v) DTT for 15 min with shaking. The strip was then incubated in 10 ml of SDS-equilibration buffer containing 2.5% (w/v) idoacetamide (IAA) for 15 min with shaking. A gel was cast following the recipe set out in Table 2.1 for a 15% (v/v) resolving gel.

**Table 2.1** ***15 % (v/v) Resolving Gel***

| Reagent | Volume | Additional information |
|---|---|---|
| 30 % (v/v) acrylamide | 10 ml | |
| 1.5 M Tris-HCl pH 8.8 | 5 ml | |
| Ultrapure $H_20$ | 4.6 ml | |
| 10% (v/v) SDS | 0.2 ml | |
| 10% (v/v) ammonium persulphate (APS) | 0.2 ml | Add fresh |
| TEMED | 0.01 ml | Add fresh |

Once the gel was set the equilibrated IPG strip was slotted on to the top of the gel with the acidic end to the left. The strip was overlaid with a 0.5% (w/v) agarose sealing solution and assembled in to a PROTEAN™ II Slab Cell Kit (Bio-Rad). Running buffer was prepared as a 10 x stock solution but diluted to 1 x (2.5mM tris-base, 19.2 mM glycine, 0.02% (w/v) SDS) before use. The gel was run at 150 V and 30 mA for about an hour or until the dye-front had run to the bottom of the gel.

#### *2.2.3.1.4 Coomassie blue staining*

After electrophoresis, the gel was removed from the gel plates and fixed (40% (v/v) ethanol, 10% (v/v) acetic acid) for 2 hr. The gel was washed twice for 10 min with ultrapure $H_20$ (resistivity 8.2 Ωm, Milli-Q grade), before staining for 1-7 days using a colloidal coomassie stain (1 part methanol, 4 parts colloidal stock (0.7M ammonium sulphate, 1.2% (v/v) phosphoric acid and 1.9% (v/v) of a 5% (w/v) coomassie stock)).

### 2.2.3.2 Large 2D-gel electrophoresis

#### *2.2.3.2.1 Rehydration*

A volume of 460 µl of rehydration buffer (8M urea, 4% (w/v) CHAPS and 0.001% (w/v) bromophenol blue) was added to the protein pellet. Before adding rehydration buffer to the samples, 3.4 mg dithiothreitol (DTT) and 5.5 µl IPG buffer pH 3-10 (GE Healthcare) were added to 1 ml rehydration buffer. Dissolved pellets were vortexed and left at room temperature for 1 hr. The sample was centrifuged at 6000 x g for 5 min and then 450 µl of the rehydration buffer and sample mix were loaded into one lane of the immobiline DryStrip reswelling tray (GE Healthcare). A 24cm pH 3-10 NL IPG strip (GE healthcare) was placed gel side down in to the same lane. DryStrip cover fluid was used to overlay the strip. Once the lid was in place the strip was left over night at room temperature to rehydrate.

#### 2.2.3.2.2 *First dimensional Isoelectric focusing (IEF)*

The rehydrated strip was placed on an Ettan™ IPGphor II™ (GE Healthcare) following the manufacturer's handbook. Strips were focused at 50 μA per strip following a five step protocol involving; 3 hr Step at 300 V, 3 hr Gradient at 600 V, 3 hr Gradient at 1000 V, 3 hr Gradient at 8000 V and 4 hr Gradient at 8000 V.

#### 2.2.3.2.3 *Second Dimensional SDS-PAGE*

The strip was equilibrated using the same method as the mini 2D electrophoresis (see section 2.2.3.1.3). Plates were treated with bind saline (80% (v/v) EtOH, 18% (v/v) dd-$H_2O$, 2% (v/v) acetic acid, 0.1% (v/v) γ-methacryloxypropyltrimethoxysilane) and left for an hour to dry before being assembled. A gel was cast following the recipe set out in Table 2.1 for a 12.5% (v/v) resolving gel and left overnight to set.

Once the gel was set, the IPG strip was slotted on to the top of the gel with the acidic end to the left. It was covered with 0.5% (w/v) agarose sealing solution before being assembled into the Ettan DALT*six* System. It was then run in 1 x running buffer. The system was set to run at 5W per gel for 30 min and then 4 hr at 17W per gel.

#### 2.2.3.2.4 *Sypro Staining*

Once the gel had run, the gel was fixed (40% (v/v) methanol, 10% (v/v) trichloroacetic acid) and left covered on a shaker for 3 hr. The gel was washed three times in ultrapure $H_20$, for 10 min each time, then covered with SYPRO® Ruby protein gel stain (Invitrogen) and developed overnight.

### 2.2.3.3 Difference In Gel Electrophoresis

#### *2.2.3.3.1 Sample preparation*

The first steps of preparation were undertaken following the methods set out in sections 2.2.2.1 and 2.2.2.2. After acetone precipitation the pellet was resolubilised in lysis buffer, described in section 2.2.2.1, to give a final concentration of 5μg/μl. The pH level was checked to ensure it was at pH 8. The rat cell lysates were labelled using CyDye$^{TM}$ DIGE Flour minimal dyes (GE Healthcare), according to the manfacturer's recommended protocol. 50μg of both the control protein sample and the treated protein sample were added to 1 μl of working (400 pmols/μl) Cy3 or Cy5 (depending on sample) and mixed by vortexing. Control and treated samples were alternately labelled with either Cy3 or Cy5 to normalize for dye specific effects. Equal amounts of all samples were mixed together to produce an internal standard, to minimise differences caused by sample preparation, which was labelled with Cy2. The labelled mixtures were incubated on ice in the dark for 30 min and the reaction terminated by the addition of 1μl of 10mM lysine, followed by further incubation on ice for 10 min. Samples were mixed for each of the four gels as shown in Table 2.2. An equal volume of 2x sample buffer (2M thiourea, 7M urea, 4% (w/v) CHAPS) was added to each of the four biological replicates. The volume of the samples was made up to 450 μl using rehydration buffer (2M thiourea, 7M urea, 4% (w/v) CHAPS, 0.002% (w/v) bromophenol blue, 2% (w/v) DTT and 2% (v/v) pH 3-10 ampholytes).

**Table 2.2** ***DIGE labelling of samples***

| | Cy2 | Cy3 | Cy5 |
|---|---|---|---|
| **Gel 1** | Internal standard | Control plate A | Treated plate D |
| **Gel 2** | Internal standard | Control plate B | Treated plate C |
| **Gel 3** | Internal standard | Treated plate B | Control plate C |
| **Gel 4** | Internal standard | Treated plate A | Control plate D |

#### *2.2.3.3.2 First dimension separation*

Immobiline DryStrips, pH 3-10 Non-Linear 24 cm (GE Healthcare), were rehydrated over night with the rehydration sample mix using a reswelling tray (GE Healthcare). IEF was carried out on an Ettan™ IPGphor II™ (GE Healthcare) following the instructions in the manufacturer's handbook. Strips were focused at 50 μA per strip following a five step protocol involving; 3 hr Step at 300 V, 3 hr Gradient at 600 V, 3 hr Gradient at 1000 V, 3 hr Gradient at 8000 V and 4 hr Gradient at 8000 V.

#### *2.2.3.3.3 Second dimension separation*

Low-fluorescence glass plates, 27 × 21 cm (GE Healthcare) were used to assemble the gels. The back plates were treated with bind-silane solution (80% (v/v) EtOH, 18% (v/v) dd-$H_2O$, 2% (v/v) acetic acid, 0.1% (v/v) γ-methacryloxypropyltrimethoxysilane), and left to dry for one hour at room temperature. Reference markers were placed on the back plates before gels were cast. Gel solution was made following the recipe in Table 2.1 for a 12.5 % (v/v) resolving gel. The solution was made without APS and left in the fridge to cool. Once plates have been assembled APS was added and the solution poured in to the caster. A

solution of 0.1% (w/v) SDS was used to overlay each gel and then left overnight to set.

Prior to the second dimension separation the strips were equilibrated in SDS-equilibration buffer (2% (w/v) SDS, 50mM Tris-HCl pH 8.8, 6M urea, 30% (v/v) glycerol and 0.002% (w/v) bromophenol blue) twice for 15 min. DTT 10 mg/ml was present in the first equilibration and 25 mg/ml of idoacetamide was included in the second round. The strips were immediately transferred to the 12.5% SDS-PAGE gel before being covered with agarose sealing solution (0.5% (w/v) agarose, trace bromophenol blue). Gels were run in the Ettan DALT*six* System which was assembled as guided by the handbook. The gels were run for 30 min at 5 W per gel and then 4 hr at 17 W per gel or until the tracking dye reached the bottom edge.

A preparative gel was prepared using 400 µg of a pooled sample made from all biological replicates of treated and untreated samples. This gel was run under the same conditions as the previous gels and at the same time. The gel was fixed in 40% (v/v) MeOH, 10% (v/v) acetic acid for 1 hr then washed twice with ultrapure $H_2O$ for 10 min before staining with SYPRO® Ruby gel stain (Invitrogen) overnight. The gel was then washed in 10% (v/v) MeOH, 7% (v/v) acetic acid for 1 hr followed by two 5 min washes in ultrapure $H_2O$. The preparative gel was compared to the DIGE gels and spots of interest were picked from this gel for MS analysis.

#### *2.2.3.3.4 Gel imaging*

All gels were scanned using Ettan™ DIGE Imager (GE Healthcare) following the manufacturer's guidelines. For the fluorescently labelled gels the scanner was set at specific wavelengths for each dye shown in Table 2.3

**Table 2.3** ***Recommended filters for each dye taken from the Ettan Dige System User Guide (GE Healthcare)***

| Dye | Excitation Filter (nm) | Emission Filter (nm) |
|---|---|---|
| Cy2 | 480/30 | 530/40 |
| Cy3 | 540/25 | 595/25 |
| Cy5 | 635/30 | 680/30 |
| Sypro | 540/25 | 595/25 |

#### *2.2.3.3.5 Gel image analysis*

DeCyder™ Differential Analysis Software (GE Healthcare) was used to analyse the gel images to allow spots of interest to be identified. The DeCyder programme works by firstly importing the gel images into the DeCyder database using the Image Loader tool. Twelve gel images of three dye channels from four gels were uploaded into the programme. Every spot on each gel is identified using the batch processor within the part of the software called DIA (Differential In-gel Analysis). The DIA module also allows a comparison from a pair of image channels (Cy2/Cy3/Cy5) on the same gel to be made. All the data is combined into a file which is uploaded to the BVA (Biological Variation Analysis) section. The next step was to compare all images from the different gels to gain statistical data on the protein expression levels of each spot, allowing modulated proteins to be identified. Gel to gel spot matching was undertaken using a match detection algorithm. A master gel was assigned to one

of the Cy2 images and each image was then matched to it either by using the batch processor or manually, enabling all common spots across the gel images to be identified. To decide whether differences in spot volume between the different gels were significant the Student's t-test was performed. Any protein spots which appeared in nine of the twelve gels and had a +/- fold change of 1.5 or above ($p<0.05$) was highlighted as a spot of interest. The preparative gel image was uploaded into the DeCyder software and assigned the pick gel status. This image was then matched to the master gel and then any spots of interest were identified on the preparative gel. These were manually checked before exporting the pick list.

#### *2.2.3.3.6 Identification of protein spots*

The pick list was uploaded to the Ettan™ Spot Picker (GE Healthcare) software and the robotic spot picker was used to remove the spots from the preparative gel. Before being able to identify the proteins each spot needed to be digested, this was undertaken using trypsin.

## 2.2.4 Protein Identification

### 2.2.4.1 Trypsin digest

Trypsin digestion was performed on all the spots individually. The spots were destained by incubating at 37°C with 10 µl of 50mM ammonium bicarbonate /50 % (v/v) acetonitrile (ACN) for 10 min. The destain solution was carefully removed and discarded and the destain process repeated. Following this, 10 µl of 100 % (v/v) ACN was added to each spot and again incubated at 37°C for 15 min, at which point the plug turned white showing dehydration. The solvent was removed and the

samples replaced into an incubator for 10 min to allow the remaining solvent to evaporate. Sequence grade trypsin (Roche) was diluted with 50mM acetic acid to give a stock solution of 100ng/µl. This was then diluted 1/10 with 25mM ammonium bicarbonate and 10 µl of this solution was added to each well. These were then incubated for 30 min to 1 hr at 37°C. A further 10 µl of 25mM ammonium bicarbonate was combined with the solution and left to incubate at 37°C overnight. To stop the reaction 2 µl of 2.6M formic acid was applied. Samples were stored at -20°C.

#### 2.2.4.2 LC-MS/MS

The LTQ (LC-MS/MS) (Thermo Fisher Scientific) was used to acquire protein identifications for all spots. The LTQ is a linear quadrupole ion trap, which is similar to a quadrupole ion trap, but it traps ions in a two dimensional quadrupole field. This ion-trap mass spectrometer is coupled to an external electrospray ionisation (ESI) source to introduce the gas-phase ions to the mass spectrometer. An electrospray is formed by applying a strong electrical field, under atmospheric pressure, to a liquid passing through a capillary tube with a weak flux. The field induces a charge accumulation at the end of the capillary, causing highly charged droplets to form. The droplets are then passed through either a curtain of heated inert gas or a heated capillary to remove the last of the solvent. Following this the ions are held in the trap and then expelled from it, separating the ions according to their m/z, due to the stability of their trajectory in an oscillating electric field (de Hoffmann, 2002). The ions expelled from the trap can be fragmented and subjected to MS/MS to give sequence data for the parent peptide.

Briefly, the method involves using a Dionex Ultimate 3000 HPLC system equipped with a nano C18 Pepmap reversed phase column. An injection of 10 µl of the tryptic peptides, diluted in 1% (v/v) formic acid (FA), was loaded on to a C18 TRAP, desalted and washed before being transferred on to the column. The peptides were eluted at a flow rate of 300 nl/min with a linear gradient of 0-50% (v/v) ACN/ 0.1% (v/v) FA over 30 min, followed by 80% (v/v) ACN/0.1% (v/v) FA for 5 min. The ionised peptides were analysed using the data-dependant "triple play" mode. This allows the mass spectrometer to acquire data as an automated process without the intervention of the user. After the initial MS scan a charge state is allocated to each ion in the "zoom scan" mode. The threshold for peptides to be further analysed was set at 20,000. The three most intense ions above the set threshold are selected to be fragmented and subjected to a final MS/MS scan. The LTQ was tuned before and after a run using a 500 fmol/µL solution of glufibrinopeptide (*m/z* 785.8). Once spectra (dta files) were collected for samples these were merged into mgf files using Proteome Discoverer 1.1 (Thermo-Finnigan) before submitting to MASCOT. MASCOT is a search engine that uses mass spectrometry data to identify proteins from primary sequence databases.

### 2.2.4.3 MASCOT

Mascot Daemon (v 2.3.2) was used to analysis the merge files. The search was set to use the ratIPI database (v 3.53) with parameters of; fixed modification of carbamidomethyl of cysteine residues, variable oxidation modifications, peptide charge of +1,+2 and +3, peptide tolerance of +/- 1.5 Da, a maximum of 1 missed cleavage and MS/MS tolerance of +/- 0.5 Da.

#### 2.2.4.4 Bioinformatics Analysis

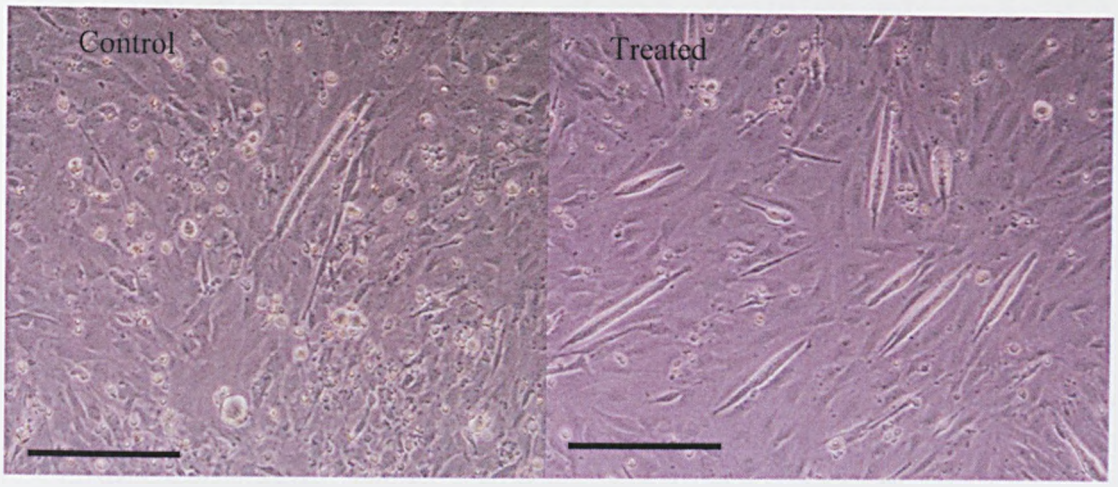
The Database for Annotation, Visualisation and Integrated Discovery (DAVID v 6.7) was also used in bioinformatics analysis to allow functional classifications as wells as identifying which biological processes are enriched. Several different pathway analysis software packages were used to identify any pathways the proteins of interest may be involved in. Ingenuity's Pathway Analysis software, Gene Go and Pathway Studio 6.1 (Ariadine) were all used by uploading the IPI gene id's of all proteins of interest.

## *2.3 Results*

### 2.3.1 Optimisation of Cell Culture

The initial part of the experiment involved establishing a primary rat skeletal muscle cell line which could be treated with salbutamol. Once cells were confluent, the cell medium was changed to fusion medium including either salbutamol for the treated cells or an equal amount of DMSO for the control cells. At the appropriate time, the medium was removed from the plates, wells washed with PBS and plates containing the cells frozen down at – 80 °C. Before completing this, images were taken of the cells using high magnification microscopy by Adrian Thompson at Pfizer. A clear difference can be seen between the control and treated cells when comparing images taken at Pfizer. Elongation and fusion of the cells can already be seen after 24 hr for both groups (Figure 2.2 A) but the effect is greater within the treated wells. The elongation and fusion becomes more pronounced after 96 hr as well as visible twitching of the cells (Figure 2.2 B). Again the effect is greater in the treated cells and no twitching can be seen in the control cells. These images support previous work using $\beta_2$-adrenergic agonists, which stated that the agonists can cause increased growth in muscle cells (Choo *et al.*, 1992).

A)




B)

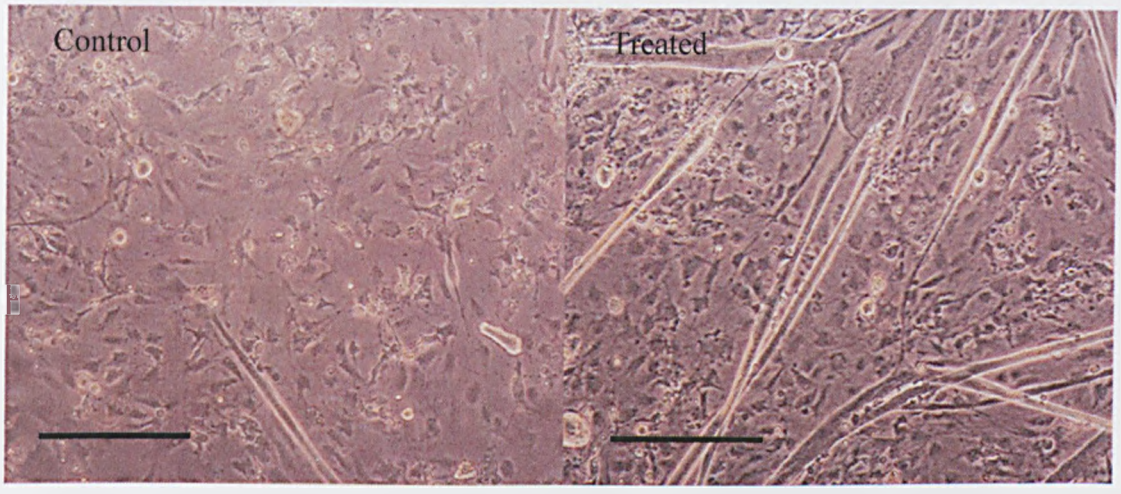



**Figure 2.2** ***Images of primary rat skeletal muscle cells after 24hr and 96hr.***

**A)** Images were provided by Adrian Thompson, Pfizer of control and treated primary rat skeletal muscle cells at after 24 hr incubation. Cells were grown in six well plates and either DMSO, for control samples, or a concentration of $10^{-5}$ salbutamol, for treated cells was added. Samples were not selected for muscle cells and contain fibroblasts.

**B)** Images of rat skeletal muscle cells under the same conditions as A) but after 96 hr of incubation.
Scale bar = 100μm *estimation

### 2.3.2 Identification of proteins modulated at 96 hr

Samples from the 96 hr sample set were used for the first DIGE experiment. After optimisation of the protocol, a successful DIGE run was completed; Figure 2.3 shows an example of one of the gels. Using DeCyder, each of the four gels were analysed separately then compared to each other to identify all spots. Each spot was compared between the four gels and to the internal standard before the fold change between control and treated samples was determined, as well as several other statistical measurements. Any spot that was seen in 9 out of the 12 gels was considered viable. A further requirement for the spot to be determined as significantly different was the need for the fold change to be > 1.5. The T-test values were also taken into consideration when deciding spot significance. Altogether 47 spots were determined to be statistically different. Unfortunately, the pick gel was unusable due to the gel not casting properly and causing it to run inefficiently. A second pick gel (Figure 2.4) was cast and run to repeat this analysis. However, as the gel was run at a different time to the original experiment, variations between the gels were likely. There is also a large amount of noise seen on the gel which makes it difficult to identify individual spots. When the two pick gels were compared, they seemed to be a good match, though they were not identical. Out of the original 47 spots identified for further analysis, only 35 of these spots could be matched with confidence to the second pick gel. All spots identified and matched with DeCyder were confirmed manually as being the same spots. Out of the 35 spots; 11 of the spots had a decrease in ratio compared to the control and are assumed to be down-regulated; the rest had an increase in ratio and are assumed to be up-regulated.

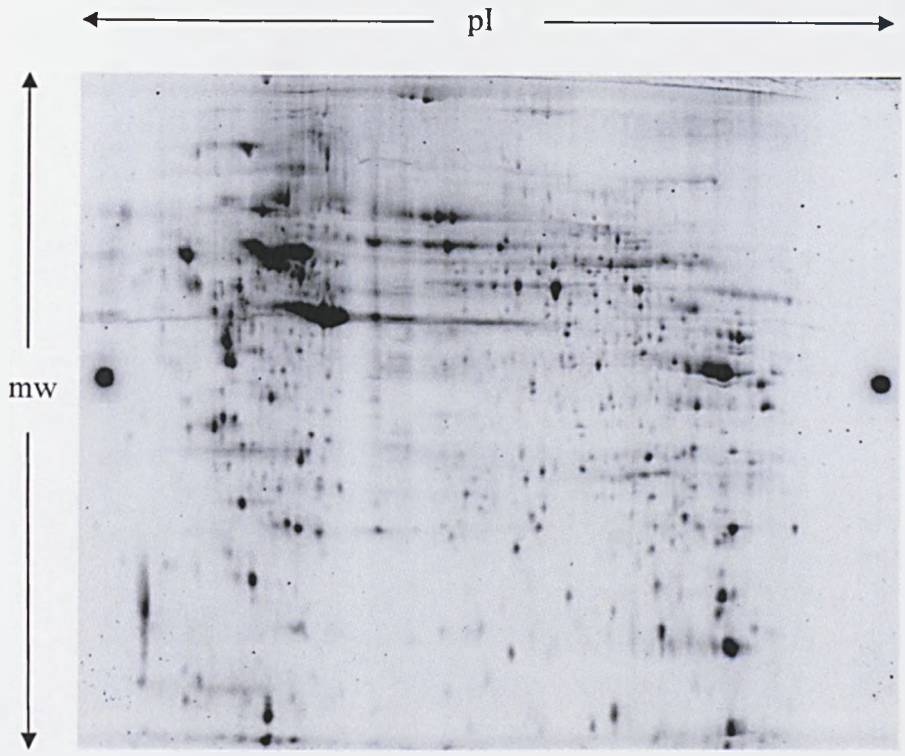



**Figure 2.3** ***2D Fluorescence Difference Gel Electrophoresis using 96 hr sample set.*** Protein from one biological replicate containing control and salbutamol treated rat skeletal muscles after 96hr of incubation were labelled with Cy3 and Cy5 fluorescent dye and separated two dimensionally using the GE Healthcare protocol for DIGE on 24 cm strips pH 3-10. An internal standard, labelled with Cy2, was run on each gel containing samples from all four biological replicates, both control and treated. Four labelled gels were run in total for a complete DIGE experiment. This image shows one of the internal standard images which was scanned with Ettan DIGE Imager (GE Healthcare) at an excitation wavelength of 480/30 nm and an emission wavelength of 530/40 nm.

The spots picked and identified are labelled on the pick gel in Figure 2.4. The 37 spots identified, were picked, digested and run through a LTQ mass spectrometer. The search engine MASCOT was used to identify the proteins, using the rat IPI database. The list of all the hits returned by MASCOT are shown in Appendix I Table 1, any hits with a score less than 50 and with only 1 peptide where excluded. A total of 110 non-redundant proteins were identified for this experiment.

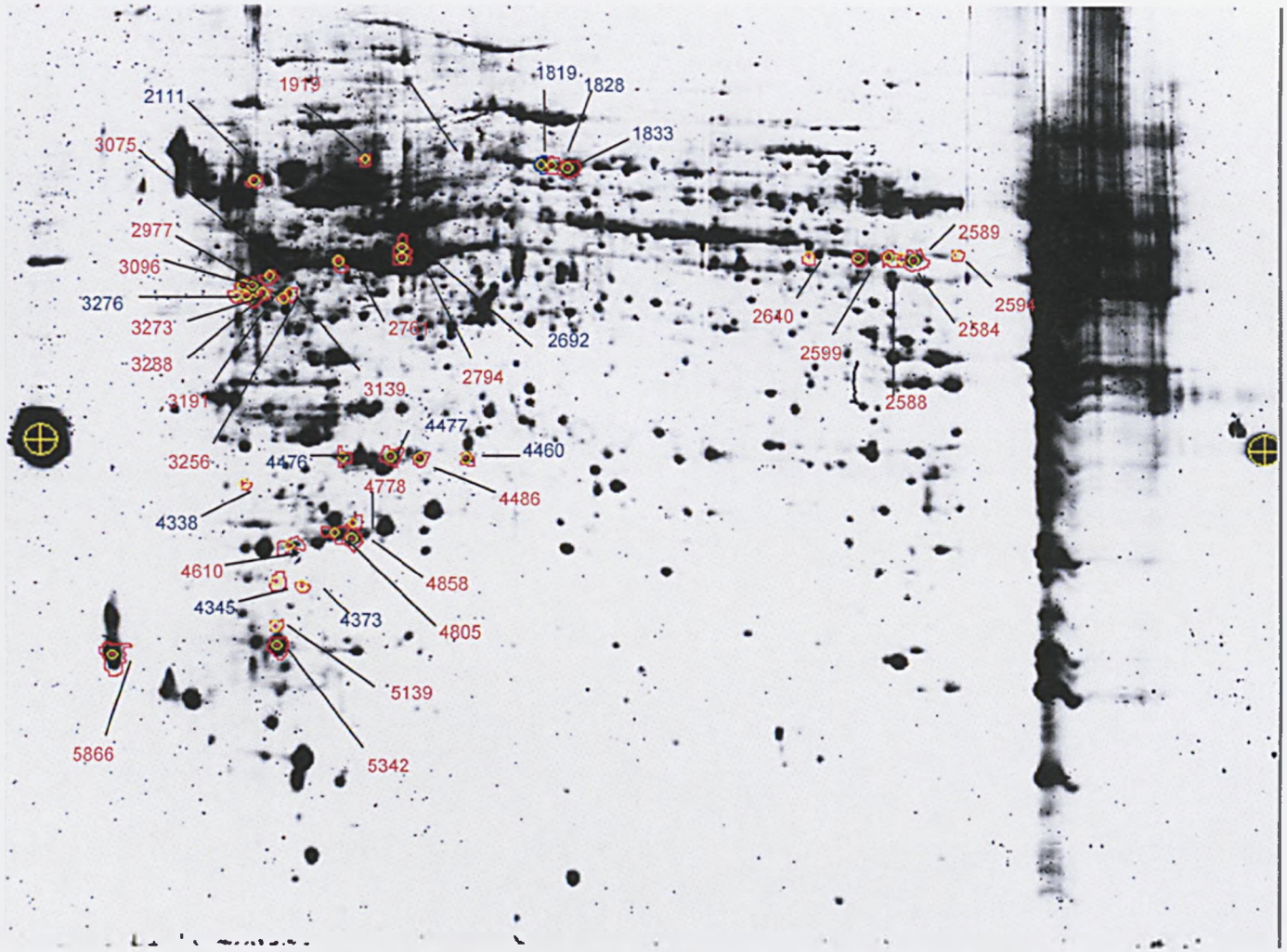



**Figure 2.4** ***2D gel electrophoresis stained with Ruby Sypro stain with differentially expressed spots labelled.***

Protein samples from both control and treated rat skeletal muscle cells (400 µg) was loaded on to the pick gel. This was separated and run under the same conditions as the fluorescently labelled gels. Gels were then fixed and stained using Ruby Sypro stain and then scanned with an Ettan DIGE Imager (GE Healthcare). Spots labelled in red have been identified with DeCyder as being up-regulated and those in blue as down-regulated.

From each spot analysed, more than three hits were returned using MASCOT. A likely reason for this is incomplete spot resolution. For example, spot 2111 is in an area where it is difficult to identify individual spots and had 10 identifications returned. However, this is not always the case, as spot 2588 also returned several hits but seems to be an obvious individual spot, when viewed as a fluorescent 3D picture. This may be due to the MS analysis being a more precise method for detecting individual proteins. Often the same proteins appeared in different spots across the gel, making it difficult for conclusions to be made, which is addressed further in the discussion. To attempt to overcome some of these problems the samples could have been separated further, either before running the gel or using different pH strips e.g. 3-5.6, 5.3-6.5, 6.2-7.5.

### 2.3.3 Identification of proteins modulated at 24 hr

An identical experiment was completed using the 24 hr sample set. Figure 2.4 shows examples of the gels run. Only 17 spots were identified as being statistically differentially modulated using DeCyder for this data set. The labelled pick gel is shown in Figure 2.6. These spots were analysed using the same methods as previously described and the hits returned from MASCOT for the 17 spots identified are shown in Appendix I, Table 2.

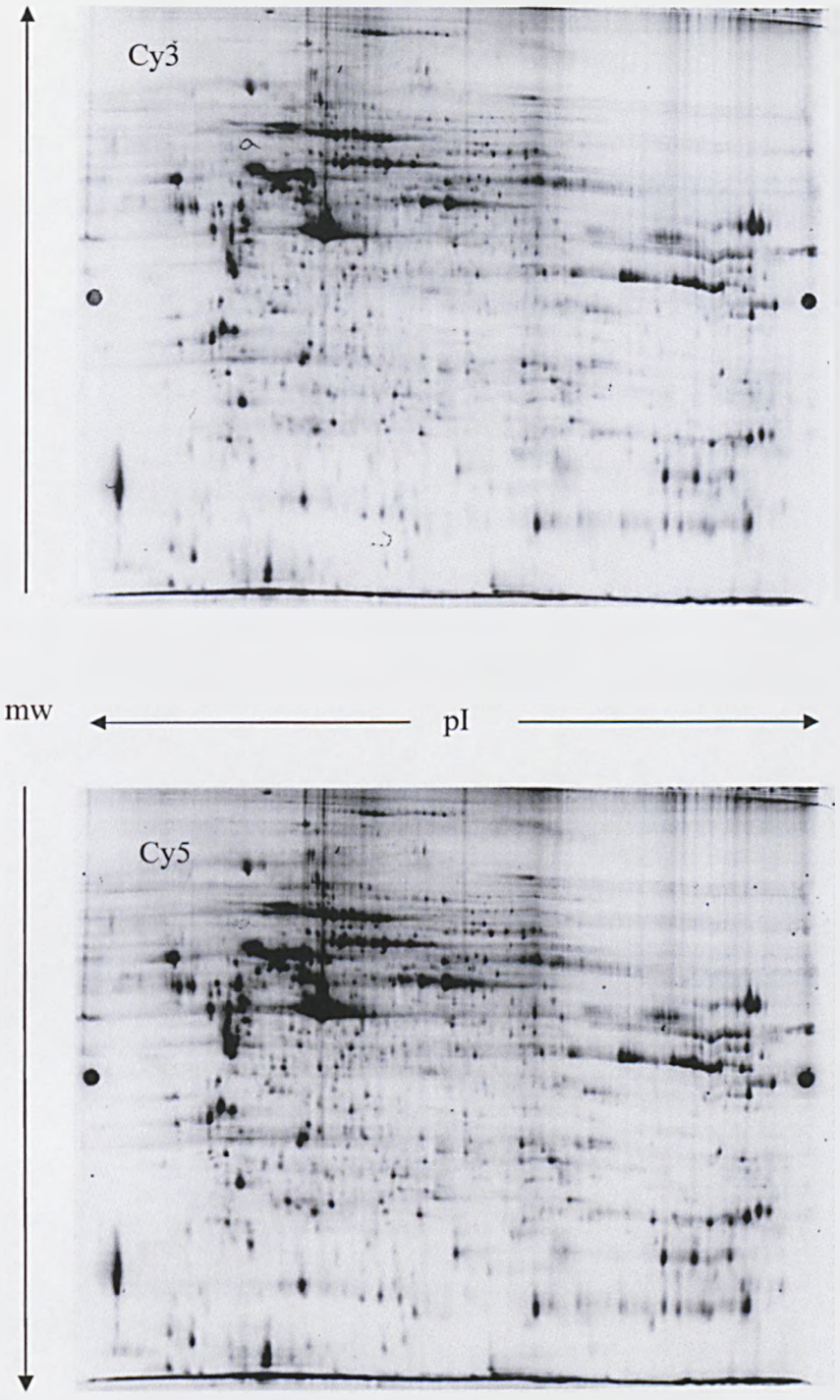



**Figure 2.5** ***2D Fluorescence Difference Gel Electrophesis of 24hr sample set.*** Protein samples from both control and treated rat skeletal muscle cells were labelled with either Cy5 or Cy3 fluorescent dye before being combined with an internal standard labelled with Cy2. This was then separated two dimensionally using GE Healthcare protocol for DIGE on 24 cm strips pH 3-10 (GE Healthcare). Ettan DIGE Imager (GE Healthcare) was used to scan the gels at different wavelengths for each dye.

From the MASCOT search most spots came back with between 1-4 identifications, though one spot returned seven. Only four of the spots identified were down-regulated. Altogether, 31 non-redundant proteins were identified and from this total 19 of the proteins identified are not found to be differentially expressed in the 96 hr sample set, which suggests that different pathways are activated at different times during the treatment. In both sample sets, spots are often dominated with cytoskeletal proteins, predominately vimentin and desmin which appear in most spots. It is unclear why certain proteins are constantly seen in several spots over different experiments. It may be due to these proteins always being post-translationally modified causing multiple forms, or poor separation is occurring during the experiments. Whatever the reason it makes a firm conclusion difficult, firstly, as it is not possible to determine exactly which protein is causing the modulation of the spots. Secondly, protein modulation may be missed due to a spot containing more than one protein being up and down modulated but when accumulated seems unchanged.

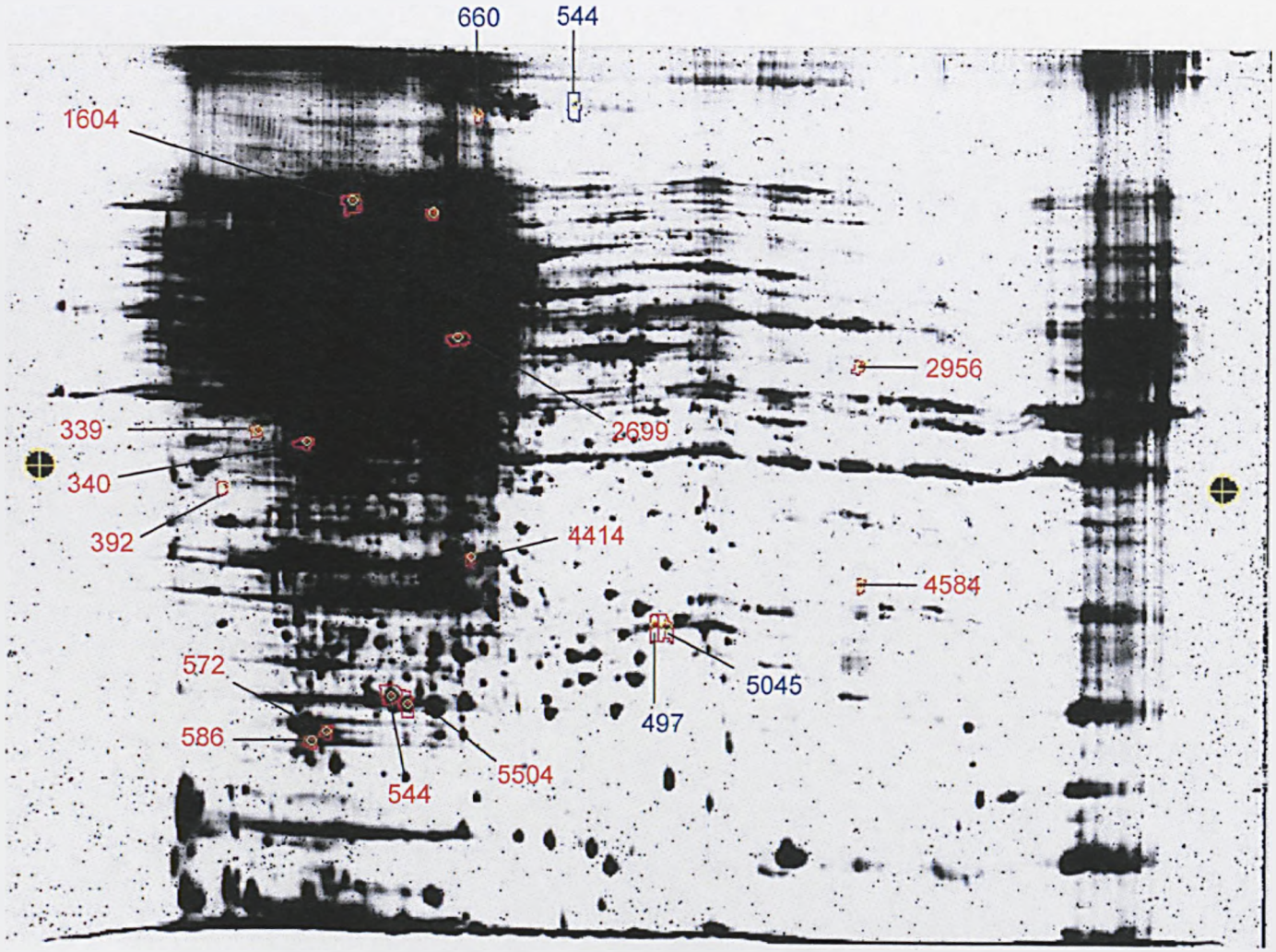



**Figure 2.6** ***2D gel electrophoresis using 24 hr sample set and stained using Ruby Sypro Stain.***

Protein samples from all four replicates of control and treated rat skeletal muscle cells (400 µg) were loaded on to the pick gel. This was separated and run under the same conditions as the fluorescently labelled gels. Gels were then fixed and stained using Ruby Syrpro stain and scanned using an Ettan DIGE Imager (GE Healthcare). Spots identified using DeCyder as being statistically different are labelled. Numbers in red correlate to up-regulated spots and those in blue are down-regulated.

### 2.3.4 Characterisation of proteins identified at 96 hr and 24 hr

Once a list of identified proteins had been produced, determining which pathways and functional groups they may be involved in was the next task. By grouping the proteins into their functional categories using their Gene Ontology (http://www.geneontology.org/) terms (Figures 2.7 and 2.8) it helped give a better understanding of which areas of cell biology were being affected by salbutamol. Any protein which was unidentified or within a functional category with only one protein identified was classed under "other". For both 24 hr and 96 hr samples the largest numbers of identified proteins were cytoskeletal proteins (25%, 26%) followed by proteins grouped in "other" (18%, 24%), then those proteins involved with transcription and translation (11%,11%).

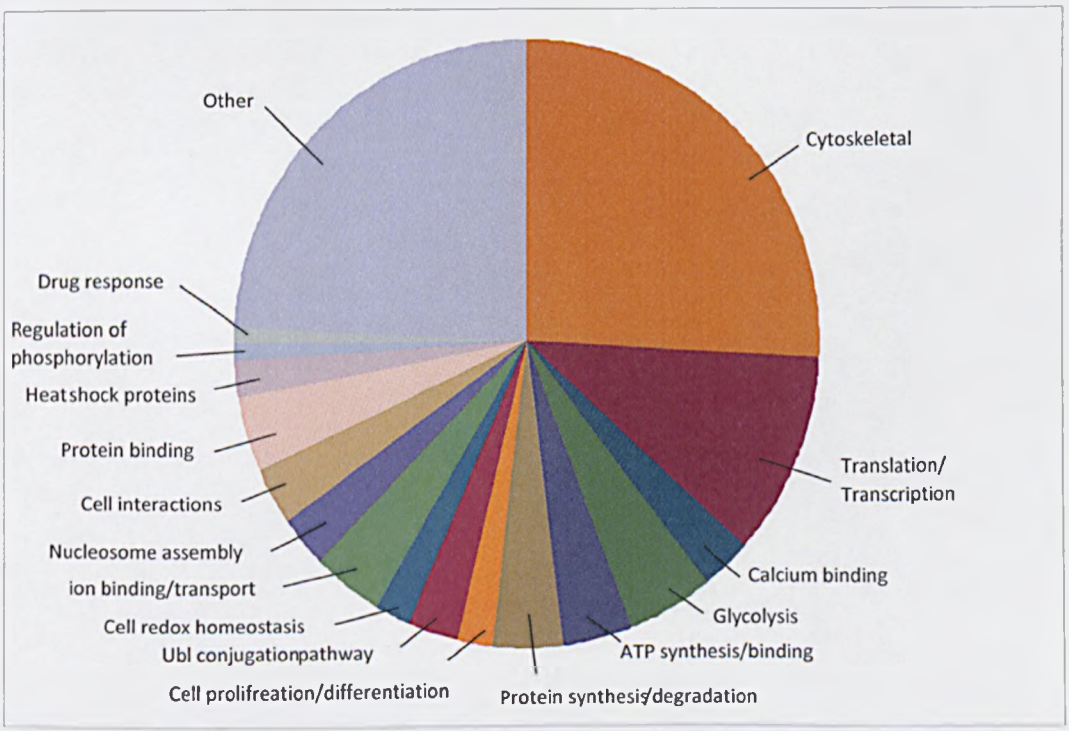



**Figure 2.7** ***Characterisation of proteins identified as differentially modulated in the rat muscle cells after 96 hr of treatment.*** Spots identified as being modulated after 96 hr of salbutamol treatment were analysed via the LTQ mass spectrometer. MASCOT was utilised to make protein identifications using the spectra gained by comparing peptides to the rat IPI database. The proteins identified were grouped according to GO terms functions. Percentages of each function compared to the total identified proteins are shown.

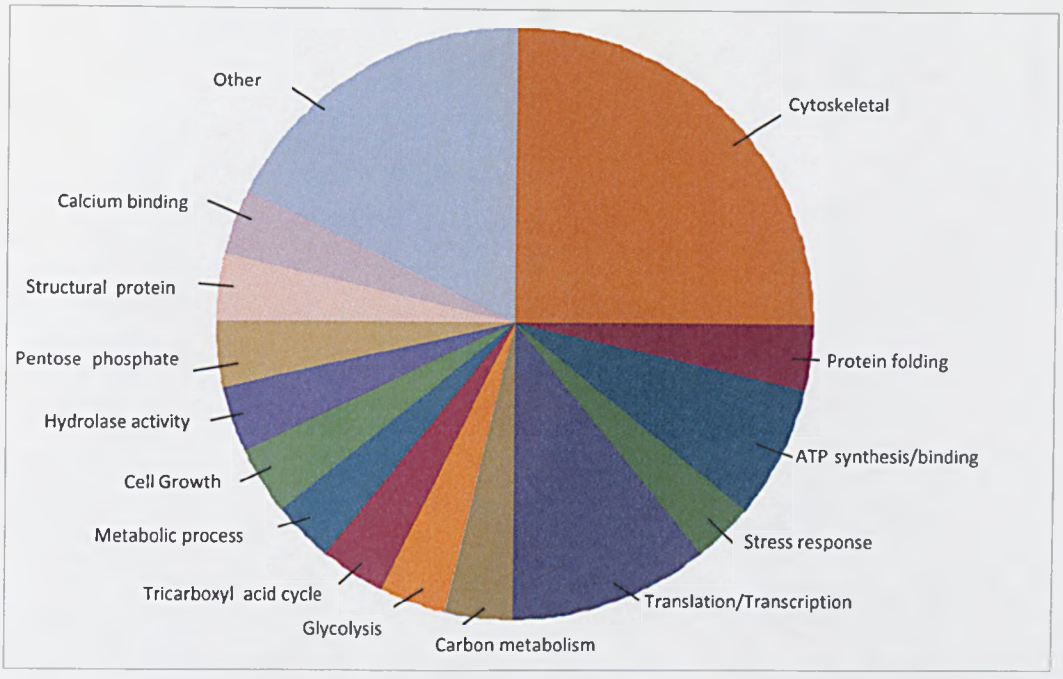



**Figure 2.8** ***Characterisation of proteins identified as being differentially modulated in rat muscle cells after 24 hr of treatment.*** Spots identified as being modulated after 96 hr of treatment were analysed via the LTQ mass spectrometer. MASCOT was utilised to make protein identifications using the spectra gained by comparing peptides to the rat IPI database. The proteins identified were grouped according to GO terms functions. Percentages of each function, compared to the total identified proteins, are shown.

However, at this point it was not possible to conclude whether the abundance of cytoskeletal proteins was due to contamination on 2D gels, possibly because of partial insolubility leading to smearing or, whether this reflected true modulation of structural proteins relating to cell differentiation. As each modulated spot comprised on average three protein ID's, distinguishing the protein responsible for the change in expression level is difficult. Assessing the MS data (the Mascot score, percentage coverage and peptide count) for each protein can give a rough indication of dominance within each spot, but in some spots it was still unclear. For some spots this analysis enabled a decision on which protein was dominate as some proteins were often identified with only one peptide or a very low coverage. However, in

other cases there was a number of proteins still with high coverage, peptide counts and scores. The fold change in spot also indicates the degree of modulation. Taken together, a clearer picture of which proteins were involved in salbutamol modulation emerges, however without complete spot resolution, the results must be viewed with a degree of caution.

When the percentage of proteins from each group compared to the total number of proteins identified are calculated for each sample set further conclusions can be made. In Figure 2.9 it is possible to identify that proteins involved in ion binding and transport, nucleosome assembly, cell interactions and protein binding seem to be modulated only at 96 hr, whereas structural proteins and energy production proteins are affected after 24 hr. However, the data set is small which needs to be taken into consideration when analysing the results. To confirm these finds, further experiments need to be undertaken. This could indicate that an immediate effect of the anabolic agent is to produce a structural effect which needs a great deal of energy.

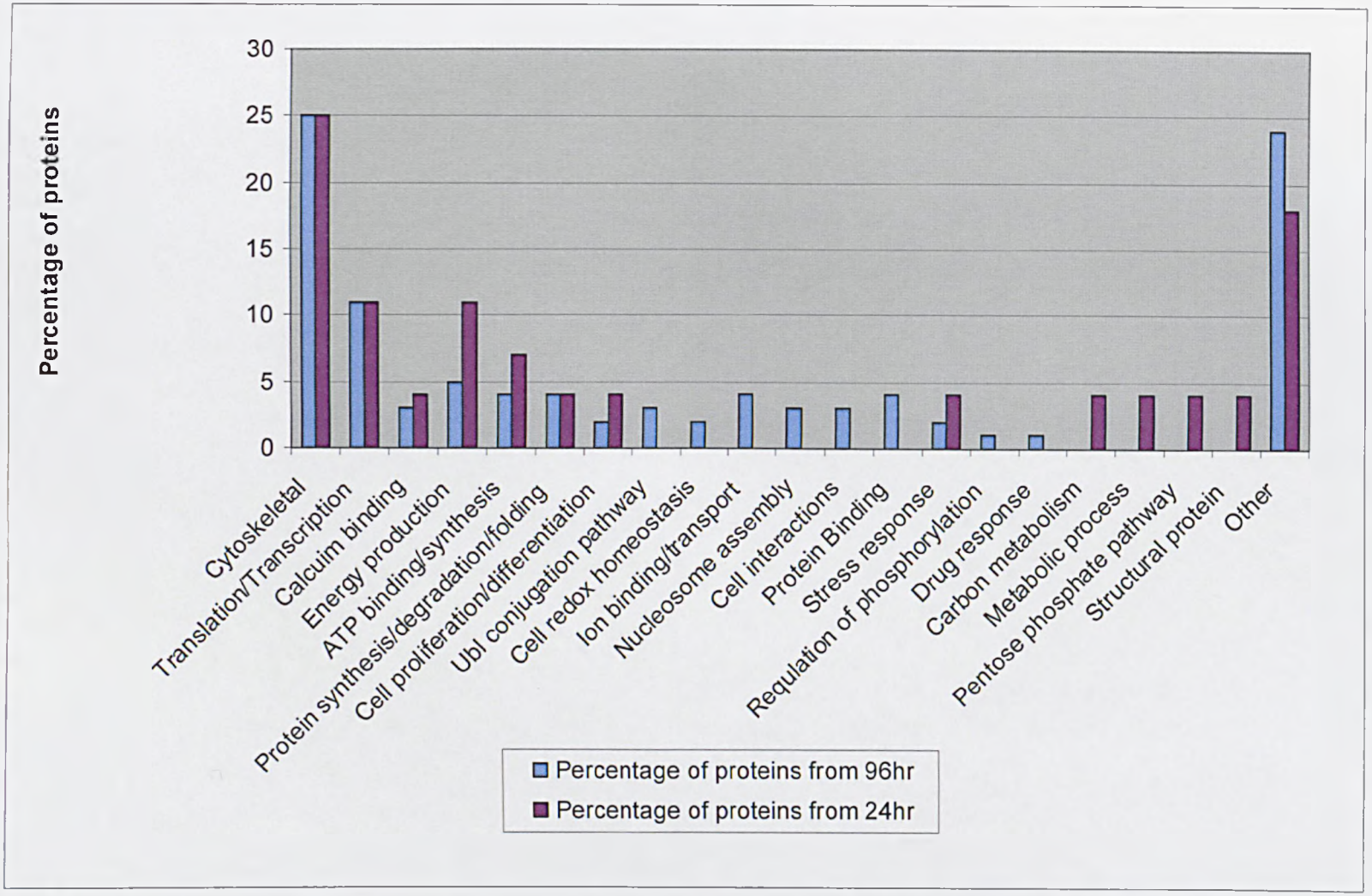



**Figure 2.9** ***Analysis of differences in percentage of proteins within certain functional groups that are modulated at 24 hr and 96 hr***
Functional groups were determined using GO terms. All proteins identified by MASCOT with a score greater than 50 and more than 1 peptide were used. Percentages were calculated by the number of proteins within a certain functional group compared to the total number of proteins identified for each time point.

### 2.3.5 Statistical Analysis of DIGE data

One of the biggest concerns with DIGE is its bias towards abundant proteins and the inability to distinguish the exact protein within a spot that has caused the change in ratio. To investigate the reliability of the data collected a separate pick gel (hereby referred to as the statistical gel) was run under the same conditions as the previous experiments and 50 random spots were picked and analysed in the same way. Figure 2.10 shows the functional groups identified from this experiment. As seen in the 24 hr and 96 hr data sets a large proportion of the proteins identified are cytoskeletal. Altogether, 200 identifications were made for the 50 spots but there were only 88 non-redundant proteins within this total (Appendix I Table 3). A comparison between the non-redundant proteins identified in the statistical gel, 24 hr and 96 hr sample sets highlighted 27 proteins (~30%) in the statistical gel data set that were found in either the 24 hr or 96 hr data sets. For only 25% and 19% of the proteins identified at 24 hr and 96 hr, respectively, to be detected in a random selection of spots suggests that the majority of proteins determined as significant have probably been modulated. Also, several of the proteins identified in both the statistical test and DIGE experiment were not always the top hit for their corresponding spot. If only top hits were compared the correlation would decrease. Further analysis using the data from this experiment is described in 2.3.6.

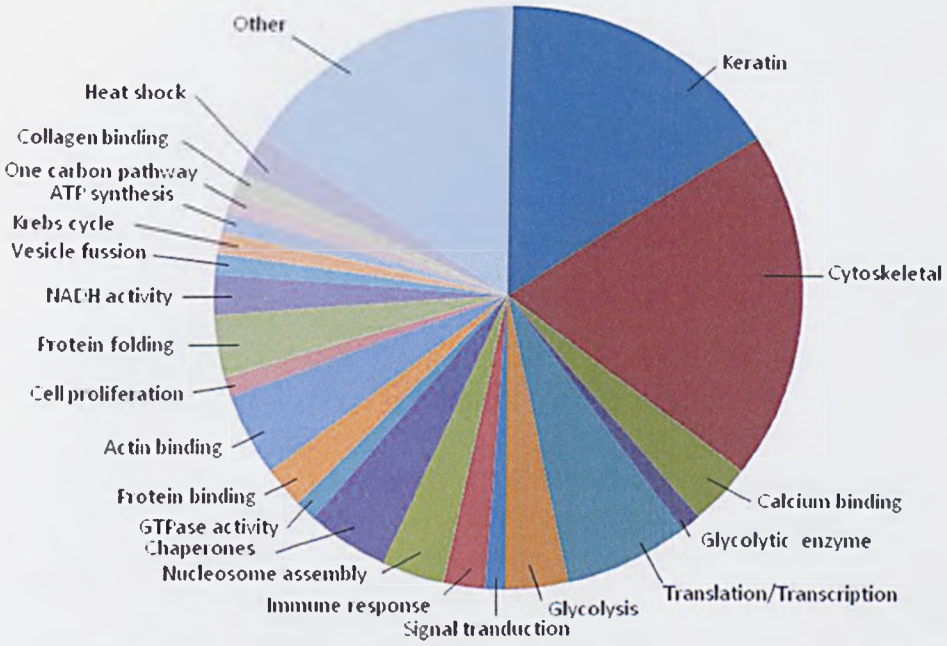



**Figure 2.10** ***Characterisation of proteins identified after random sampling.*** Fifty randomly picked spots from the statistical gel were picked and run through the LTQ mass spectrometer. The spectra gained from these samples were run through MASCOT using rat IPI database. The proteins identified have been grouped according to function using there GO terms. Percentages of each function compared to the total identified proteins are shown.

### 2.3.6 Analysis using the Database for Annotation, Visualisation and Integrated Discovery (DAVID)

DAVID provides a set of data mining tools that combine functionally descriptive data with graphical data. It also provides visualization tools that allow functional classification, biochemical pathway maps, and conserved protein domain architectures, which are linked to biological annotation (Dennis *et al.*, 2003). Gene ID's for both data sets were separately uploaded on to the DAVID site and data on functions and pathways were collected (Appendix I, Tables 4-5). DAVID is able to annotate the data using protein GO terms (http://www.geneontology.org/), the KEGG pathway database (http://www.genome.jp/kegg/) and INTERPRO

(http://www.ebi.ac.uk/interpro/), a protein domain database. Once the data had been compiled any result with a p value >0.01 was removed. The most significant information gained from the annotation chart for the 24 hr data set was: the most enriched biological processes involved the metabolic process; the cellular components most involved are non-membrane bound organelles and the cytoskeleton; the molecular functions are most likely to be structural activity and motor activity. However, for the 96 hr data set different areas were enriched: macromolecular assembly and protein assembly seem to be the most effected biological processes; again the non-membrane bound organelles and the cytoskeleton components are involved in the modulation of proteins after treatment but intracellular organelles and lumen are also highlighted; after longer treatment the molecular functions now involved according to DAVID are nucleotide and RNA binding as well as structural activity.

DAVID can also perform clustering of all annotated genes. At 24 hr the main functional clusters enriched were: "contractile fiber", "cytoskeleton", "calcium ion binding" and "collagen biosynthetic process". At 96 hr post treatment, the main enriched functional categories were "cytoskeleton", "tubulin", "microtubule-based movement", "GTPase activity" "cellular protein complex assembly", "contractile fiber" and "regulation of ATPase activity" amongst others.

DAVID usually annotates data against all the genes in the database for the species selected. As further analysis, the modulated genes for 24 hr and 96 hr samples were searched against a background uploaded of the genes identified from the random selection of 50 spots. By comparing with other proteins found within the gel, rather

than all proteins in the rat genome, a decision can be made about which clusters may actually be enriched. For the 24 hr experiment no clusters were found but at 96 hr nine clusters are found (Appendix I, Table 6), all of which were seen in the previous DAVID experiment.

### 2.3.7 Pathway coverage of differentially expressed proteins

To expand the examination of the data gained, pathway analysis software was used. Three pieces of software were tested to assess the identified list of modulated proteins gained from both the 24 hr and 96 hr sample sets. It was hoped that this software would increase the understanding of how the proteins identified interact not only with each other but within different pathways in the cell. The three different pieces of software have different features which allow a slightly different analysis to be undertaken.

The Ingenuity Pathway Analysis software is a well cited programme (Berry *et al.*, 2010; Kash *et al.*, 2006; Li *et al.*, 2007), using information extracted from PubMed to manually curate a database. The 96 hr sample set was applied to this programme to identify any pathways the proteins fell into. Using this software, five networks were identified. The pathways which had the highest scores and related to muscle cells included skeletal and muscular system development and function; tissue morphology; cell cycle; carbohydrate metabolism; and cellular development. The databases used involved all tissues rather than focusing on muscle cells which meant that pathways involved in other organs were highlighted e.g. neurological disease pathways, haematological disease and immunological disease pathways which

although not relevant to muscle cells are still interesting. Proteins within these pathways have been identified as modulated and these pathways may hold clues on how the mechanisms of salbutamol induced hypertrophy occur. Only two networks were identified using the protein list from the 24 hr sample set. The highest scoring network had pathways involved with skeletal and muscular system development and function, tissue morphology and cellular assembly and organisation (Figure 2.11).

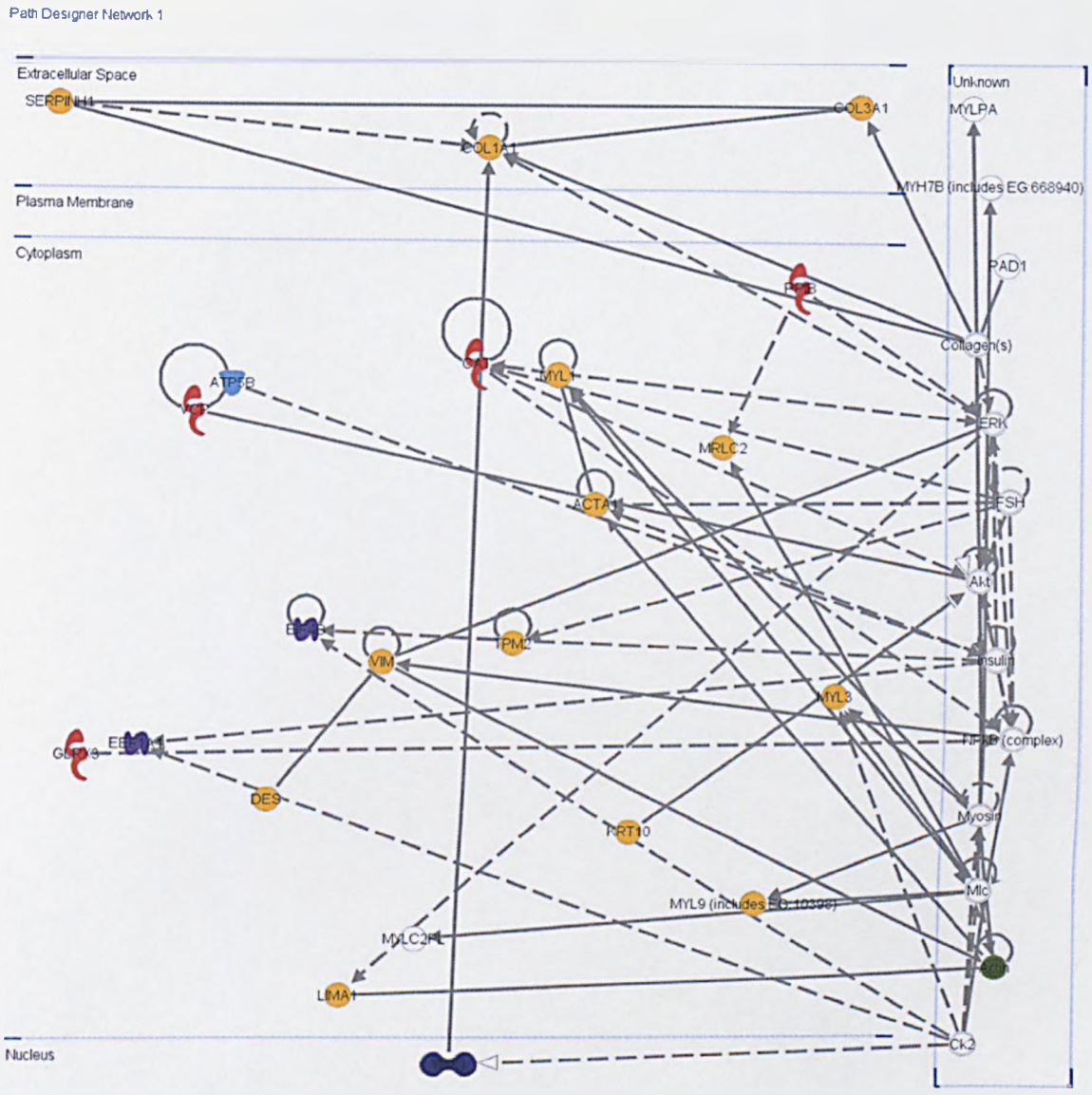



**Figure 2.11** ***Pathways affected by the identified proteins at 24hr mapped using Ingenuity's Pathway Analysis Software (Network 1).***

The network was generated using all hits returned from MASCOT for the 24 hr data set (Table 2, Appendix I). This image was the top scoring network with a score of 56. Pathways are scored according to the fit of that network to the user defined set of genes and derived from their p-value. The pathways shown are skeletal and muscular system development and function, tissue morphology and cellular assembly and organisation. Nodes represent proteins involved in the pathway. Different shapes are used to show the different classes of the proteins; yellow circles represent complexes, green circles are unknown, red spirals are enzymes, light purple sideways S shapes are translational regulators, dark purple binocular shapes are transcriptional regulators and blue upside down bells are transporters. Coloured nodes represent proteins identified in the DIGE experiment, whereas clear shapes represent the same groups but not identified. The gene names corresponding to proteins are listed in Table 8, Appendix I.

Pathway Studio 6 was also used to analyse the proteins identified from the 24 hr sample set. This piece of software uses text mining to build its databases. Abstracts and articles from PubMed are mined and can be referenced on the map. Figure 2.12 is an example of a map created using Pathway Studio 6. This map shows common regulators and targets for proteins identified at 24 hr. However, not all of the proteins identified are shown in this diagram as 13 of the proteins were not found to have common regulators or targets as any other proteins identified.

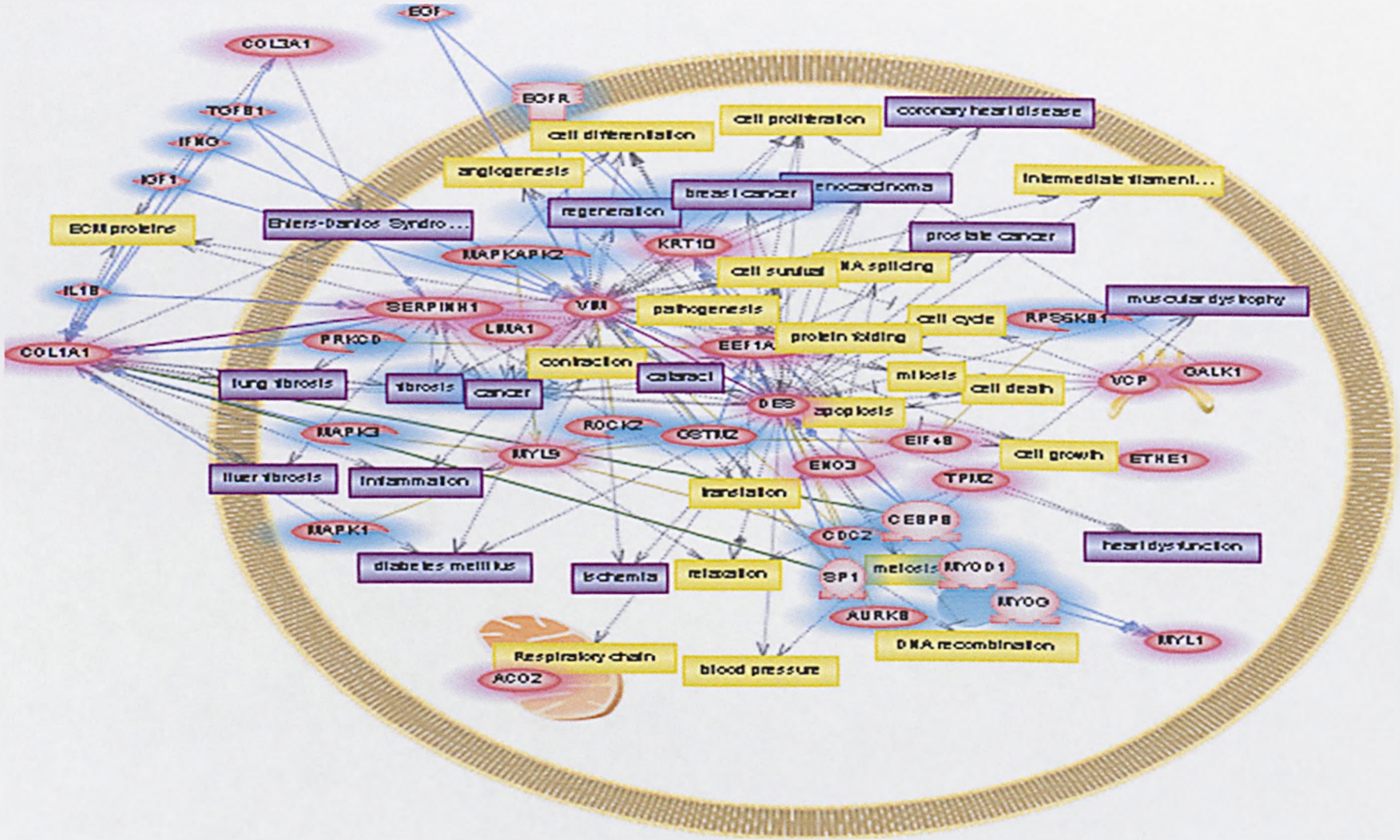



**Figure 2.12** ***Map created using proteins identified as being statistically different from 24 hr sample set.***
Proteins identified as being modulated at 24 hr were uploaded into Pathway studio 6 analysis software to determine any pathways of interest. Entities identified from the DIGE experiment are highlighted in pink. Common regulators of these proteins are highlighted in blue. All entities not highlighted are common targets. Pathways are mapped in a cellular manner and include the mitochondria, nuclei and endoplasmic reticulum.

The third piece of software used is called GeneGO, which is a manually created database of experimentally proven protein-protein, protein-DNA and protein-compound interactions. An example of the shortest pathway found when the 24 hr protein list was uploaded into the software is shown in Figure 2.13.

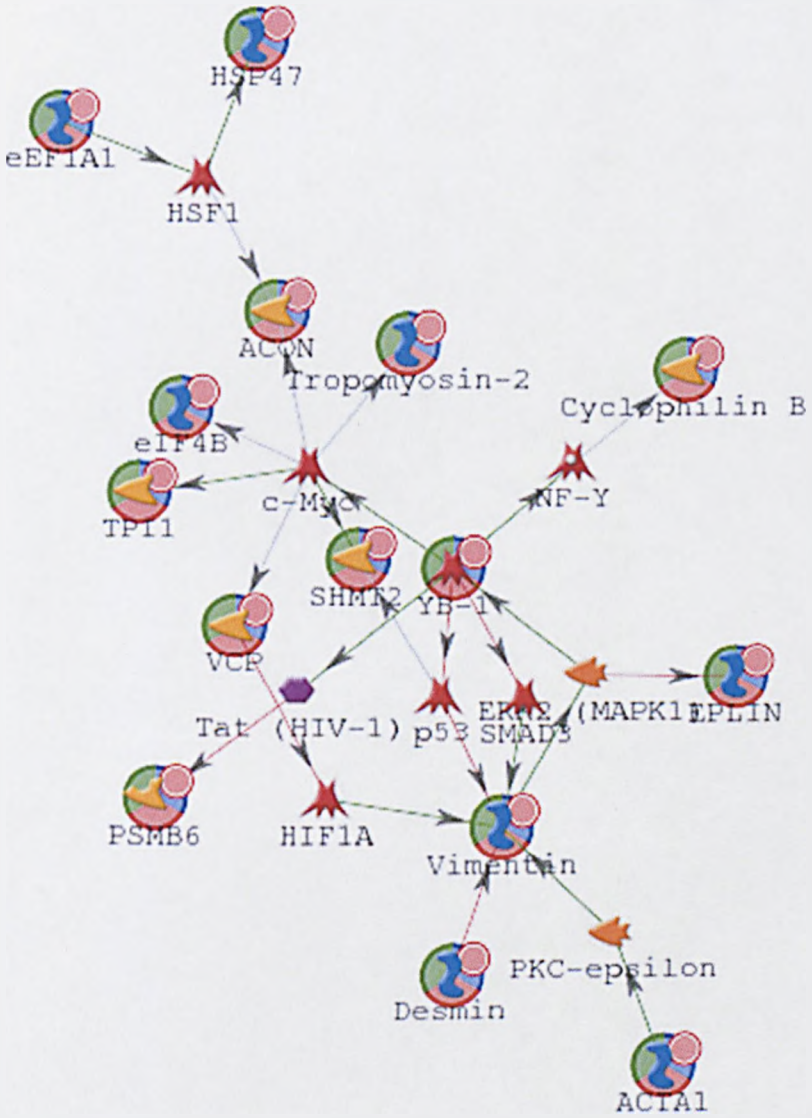



**Figure 2.12** ***Map of the shortest pathway found using 24 hr sample set and GeneGO.*** Genes identified at 24 hr to be significantly different have been mapped together, using GeneGO pathway analysis software, to only show those that interact with each other. Blue entities represent binding proteins, red are transcription factors, orange are enzymes and purple are compounds. Any entities that are circled where identified after 24 hr of treatment. The green arrows represent activation, those coloured in red are inhibition and those in grey corresponds to transport or consumption/production of metabolic intermediate.

Unfortunately limited biological information was gained using the pathway analysis software, mostly due to the functions not being designed for protein data but for genomic data. Each programme identified different pathways and different important genes, with no clear way of distinguishing between the reliability of each piece of software.

## 2.4 Discussion

This study was able to identifying several possible protein changes within muscle cells after treatment with salbutamol via DIGE. Salbutamol definitely causes muscle hypertrophy, as seen by other $\beta_2$-adrenergic agonists (Emery *et al.*,1984, MacLennan & Edwards, 1989)(MacLennan & Edwards, 1989). The images of the primary rat skeletal muscle cells (Figures 2.1) suggest that the response to this $\beta_2$-agonist is very rapid but continues for several days. Yang *et al.*, showed that using clenbuterol, another $\beta_2$-adrenergic agonist, *in vivo* increased muscle growth after 2 days, which reaches a maximum within 8 and attenuates after 14 days. Figure 2.1 clearly shows that even though the increase *in vivo* is not seen until the second day, the response starts well before this.

The initial results of the study show that several key pathways are modulated by treatment with salbutamol, with significantly more changes occurring at 96 hr post-treatment. A possible reason for the difference at the two time points may occur due to a greater expanse of pathways being involved at 96 hr as further differentiation occurs. A large proportion of the proteins identified have effects on skeletal and muscular development, as well as the cell cycle and glycolysis at both time points. However, there are several limitations to DIGE which mean that the results obtained from this study cannot be used to draw a definite conclusion relating to individual proteins.

Although there are limitations, this study has identified some potentially interesting proteins and pathways which are affected by the $\beta_2$-agonist, salbutamol, as well as confirming some findings from studies done in other groups. Burniston *et al.,* (2007) found that heat shock protein (Hsp) 72 was up-regulated when treating rats for two weeks with clenbuterol, a similar $\beta_2$-agonsist. However, in experiments undertaken by Oishi *et al.,* (2004), after four weeks the amount of Hsp 72 was significantly lower in treated rats compared to the control, though a larger dose was used. In the 96 hr sample set experiment, one of the spots included Hsp 70, which was seen to be up regulated. Hsp 70 seems to have a close relationship with calcineurin (CaN). This is a $Ca^{2+}$/calmodulin dependent protein serine/threonine phosphatase, which appears to be involved with the regulation of muscle fibre differentiation (Oishi *et al.*, 2004). In the experiment presented here, a serine-threonine kinase receptor-associated protein was up-regulated as well as calmodulin. Burniston *et al.,* (2007) found that β-enolase was also increased, which correlates with our data. Enolase is a glycolytic enzyme, however, it is thought it may have a role in other functions. It has been identified on the surface of eukaryotic and prokaryotic cells and found as an effective plasminogen receptor (Pancholi, 2001). It is also thought that it may play a role in the degradation of extracellular matrix which facilitates tissue growth (Pancholi, 2001). Several of the proteins, for example ATP synthase, triosephophosphate isomerise and phosphoglycerate kinase 1, found at both 96 hr and 24 hr to be differentially expressed have a role in either glycolysis or ATP synthesis. It is expected that an increase of energy would be needed for an increase in muscle mass. Downie *et al.,* (2008) concludes that it may be likely that the increase in protein caused by clenbuterol is due to a drug-induced increase in transcription as opposed to an effect

on translation, which correlates with the results presented here, where several translation initiation factors have been up-regulated.

### 2.4.1 Limitations of DIGE

One of the main disadvantages to DIGE is the inability to distinguish between different proteins in one spot. In nearly every spot picked (45 spots out of 52 picked from both experiments) more than one protein was identified. This makes data interpretation difficult as it is unclear which of the proteins identified will have caused the ratio change. Also, results may have been lost due to changes in low abundance proteins being hidden as high abundance proteins have co-migrated.

During this study it was observed that many of the proteins were identified in more than one spot. This fact needs to be taken into consideration when analysing the results for several reasons: firstly, the protein in question may have been separated unsuccessfully causing it to appear all over the gel and effecting ratios of spots. Secondly, it may reveal important information regarding any post-translation modifications that may have occurred causing this protein to have a different molecular mass or charge. For example, an isoform of troponin T was found in three spots in the 96 hr data set. Troponin undergoes regulation by phosphorylation within normal muscle cells (Brushia & Walsh, 1999), which would explain its positioning across the gel. Thirdly, some spots were hard to distinguish as individual spots and when sampling occurred adjacent spots may have contaminated the result. The pH range for separation was chosen to ensure no information was lost, however using a smaller pH range would have helped improve separation though would have

increased work load and time of experiment. However, in an unrelated collaborative study using primary rat heart muscle to investigate energy metabolism in low capacity runners compared to high capacity runners, using a DIGE study under the same conditions, the experiment was able to show good separation and identified 67 statistically significant proteins that differed between the two rat phenotypes (Burniston *et al.*, 2011).

When searching the literature these problems are not well discussed or analysed. By reading a selection of ten DIGE papers (Ali *et al.*, 2010; Briolant *et al.*, 2010; Dautel *et al.*, 2011; Foth *et al.*, 2011; Hamza *et al.*, 2010; Hannan *et al.*, 2010; Martinez-Esteso *et al.*, 2011; Oe *et al.*, 2011; Paasch *et al.*, 2011; Tewari *et al.*, 2011) from multiple journals from the last three years it became clear that these problems are common but often ignored. Most groups only list one protein per spot by only taking the protein with the highest MASCOT score or using other specifications to narrow down the proteins. This experiment shows that a large amount of information could be lost if post-translational modifications are not taken into account. It is also possible to over look information by only identifying the most abundant protein in a spot as the change in abundance could be due to one of the other proteins. In this experiment if only the top hits were considered only cytoskeletal proteins would be identified, whereas by analysing all the results we could identify proteins involved with glycolysis and ATP production, which can be supported by other literature (Burniston *et al.*, 2007, Downie *et al.*, 2008).

A recent study by Petrak, J. *et al.,* (2008) has shown that very similar proteins are constantly identified with 2-DE experiments. Several of these proteins which are highlighted as being commonly modulated appear within this study, such as vimentin, pyruvate kinase M1/M2, peroxiredoxin and ATP synthase beta subunit. In Petrak's study the most frequently identified protein out of the 186 2-DE experiments analysed was enolase 1. Although in this study enolase 1 was not found beta-enolase was identified six times over the two time points. Out of the list of 15 top identified proteins listed in this paper, eight where seen in this DIGE experiment, with two others seen as different isomers. Either these proteins are highly abundant and are more easily detected or this group of proteins are involved in many different functions and are far more important than originally thought.

Even with the limitations DIGE is able to identify a number of important proteomic discoveries, for example, a collaboration with Burniston *et al.,* was able to reveal that selection on low running capacity in rats is associated with cardiac energy metabolism and provided the first evidence that the cardiac proteome of low-capacity runners is exposed to greater oxidative stress using a DIGE methodology (Burniston *et al.*, 2011).

### 2.4.2 Interpretation of results

Although DIGE has its limitations it is still widely used in proteomics to give basic information about protein modulation and coverage. Within this study a smaller experiment was undertaken to discover the reliability of DIGE results. By running a typical pick gel under the same conditions and sampling 50 random spots it was

hoped to identify if the same proteins were being identified no matter which spots were picked. Encouragingly, less than 25% of proteins identified for both the 24 and 96 hr samples were identified in the statistical test. It has already been discussed that certain proteins were identified in more than one spot so it is not surprising that these proteins would be identified in random sampling. The majority of proteins fell in the category of cytoskeletal function, which mimicked the samples results. However, as muscle cells were growing during the study it was predicted that this functional group would be highly active.

A further analysis was undertaken using DAVID, which allowed clustering of the proteins identified. DAVID was used first to compare the identified modulated proteins to the rat proteome then to the list of proteins produced via random sampling. This second analysis produced no clusters for the 24 hr sample set and four times fewer clusters for the 96 hr sample set in comparison to the first DAVID analysis. These results help support the reliability of the DIGE experiments; as there is a difference in clusters found it shows that very different proteins are being identified for the two separate studies.

There are also other studies using β-agonists that highlight similar biological functions or processes as being modulated to those identified in this investigation. For example, Hong *et al.,* (2011) used a 2-DE method to identify the proteomic changes within hypertrophic cardiomyocytes after treatment with carndothelin-1 (ET-1) and isoproterenol (ISO). By performing a database search using SEQUEST the cellular functions of the differentially expressed proteins were obtained. They found that the major functional classes represented by the differentially expressed

proteins were signalling and hydrolase. They also identified other functional classes which the differentially expressed protein were part of including; cell organization, transcription, translation, chaperone, $Ca^{2+}$-binding protein, energy metabolism, and immune response. Many of these functions were also identified within this study using a similar agonist which produces the same hypertrophic response in muscle. Stella *et al.,* (2011) investigated the protein changes within cattle skeletal muscle after the treatment of 17β-oestradiol or clenbuterol. Using a DIGE methodology they were able to identify 19 non-redundant differentially expressed proteins. To determine which proteins were statistically different a fold change between treated and control groups of <0.80 and >1.20, and test *p*-value >0.05 was used. These values are a lot less stringent than those used in this study and produced less proteins identified as differentially expressed. However, several of proteins that were identified are also observed in this study, for example, actin, desmin, myosin, heat shock protein beta-1 and beta-enolase. As similar functions and proteins have been identified in previous studies, which in some cases use less strigent methods for determining statistically different changes, it enables the reliability of the results gained to be strengthened.

### 2.4.3 Overall conclusions

In summary, proteins involved in ion binding and transport, nucleosome assembly, cell interactions, protein binding and structural proteins seem to be modulated at 96 hr, whereas only structural proteins and energy production proteins are affected after 24 hr. An immediate effect of the anabolic agent is to produce a structural effect which needs a great deal of energy. It appears a more complex later effect is observed, involving a number of cellular pathways. It is easy to conclude that

salbutamol causes muscle hypertrophy, which suggests it may have a similar mode of action as other $\beta_2$-adrenergic agonists. This effect on the muscle cell is rapid, and is clearly seen using high magnification microscopy. As anticipated many of the differentially expressed proteins identified are cytoskeletal proteins. A significant number are also involved in transcription or translation. Skeletal development pathways are activated both at the early and later time points. Up-regulation of ATP synthesis, glycolysis and phosphoryalation also seems to be occurring, which is important for growth and differentiation.

There are some limitations of the DIGE results, not least that true quantitative ratios cannot be linked to individual protein identities, since the sensitivity of tandem MS revealed that most spots on the gels contained more than one protein. However, the ontology enrichment analysis is able to show clearly that many of the protein groups highlighted are likely to be direct or indirect targets for salbutamol, since the enriched functional categories fit plausible hypotheses of the effects of an adrenergic agonist. This study has enabled a good first insight into the mechanisms of this adrenergic agonist. A more in depth study of these interesting findings is discussed in the next few chapters. It has also come to light how important a technique is to producing reliable and useful data. Further analysis into proteomic techniques are also to be investigated.

# 3 Chapter 3

# A transcriptomic approach to investigating the effects of salbutmol on mouse skeletal muscle

## 3.1 Introduction

The transcriptome of a cell comprises the total complement of mRNA in that cell at any given moment. This transcriptome forms the template for protein synthesis, resulting in the corresponding proteome. Understanding the transcriptome is essential for interpreting the functional elements of the genome, understanding the working of cells and tissues, and also for understanding development and disease.

Transcriptomics is a robust, high-throughput technology capable of quantifying tens of thousands of mRNA species. Proteomics is more limited in breadth and depth of coverage due to variations in protein abundance, hydrophobicity, stability, size and charge. Being able to compare transcriptomic and proteomic data is beneficial in gaining greater knowledge of any overlaps or differences. However, this feat is not easy due to post-transcriptional gene modifications and the different stabilities and biological half-lives of RNA and proteins. Post-translational and transcriptional modifications of proteins and genes are common occurrences that are difficult to identify and measure. Post-translational mechanisms not only control gene expression through mechanisms such as translational control but also the control of protein and RNA half-lives (Varshavsky, 1996). It is evident that a bridge across these two disciplines is vital. However, while different scientific disciplines are still struggling to produce methods to analyse these modifications, it is important that all transcriptomic and proteomic data are collected with precision and accuracy to enable understanding of cell biology that is as accurate as it possible can be.

To investigate the workings of different cells and tissues gene expression is often observed. There are currently several techniques available that allow gene expression to be monitored by sequencing the whole genome. These techniques are often referred to as "next generation sequencing". Currently this high-throughput methodology is used on platforms such as Roche's 454, Illumina's Genome Analyzer and Applied BioSystem's SOLiD. Roche's 454 technology works by generating a single stranded template DNA library and amplifying each template read using emulsion PCR before sequencing the amplified templates through sequencing-by-synthesis (pyrosequencing) (Margulies *et al.*, 2005). This method enables deep sequencing up to 1000 bp read lengths and up to 1.1 Gb of sequence per run. This is useful when there is a limited genome due to the long reads meaning that there is more overlapping sequence, which helps in piecing a genome together. The Illumina approach uses a slightly different technique which enables cloning-free DNA amplification by attaching single-stranded DNA fragments to a solid surface known as a single-molecule array and conducting solid-phase bridge amplification of single-molecule DNA templates (Bennett *et al.*, 2005; Bentley, 2006). The Illumina approach gives reads of up to 150 bp with up to 6.5 Gb of sequence per 24 hr run. This approach is more effective at sequencing homopolymeric stretches than pyrosequencing. However, it does produce shorter sequence reads so cannot resolve short sequence repeats, as well as making it more difficult to assemble data if there is no genome available. The SOLiD system is based on sequencing by ligation. Whole DNA/RNA samples are used and fragmented with an enzyme before hybridising with SOLiD based adaptors. The fragments are reverse transcribed, size selected and amplified by emulsion PCR before the resulting bead with a clonal DNA colony is positioned on a glass slide ready for sequencing (Schuster, 2008). The SOLiD

sequencing process creates reads of 50 bp and up to 30 Gb of sequence per seven day run. This technology also allows up to eight samples to be run at the same time. For this experiment the Applied BioSystem's SOLiD will be used due to the ability to perform large scale sequencing, the high accuracy and due to a comprehensive genome available meaning that short reads will not cause a problem for sequence assembly.

### 3.1.1 RNA-Seq

RNA-Seq (RNA sequencing) is a technique that has been developed to use the novel high-throughput DNA deep-sequencing methods available. This provides a new method for both mapping and quantifying all transcribed RNA sequences within a given tissue. It is able to utilise next generation technologies to allow sequencing of millions of base pairs in a relatively short space of time (Meyerson *et al.*, 2010). This method has several advantages over current techniques for looking at the transcriptome. Figure 3.1 shows an outline to the method that will be utilised for this experiment. Briefly, once a population of RNA, either in total or fractionated form, such as poly(A), is collected from cell samples it is converted into a cDNA library. Each molecule is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing) (Wang *et al.*, 2009). After sequencing, the acquired reads can either be aligned to a reference genome or assembled *de novo* (without the genomic sequence), producing a transcription map consisting of both the transcriptional structure and/or level of expression for each gene (Bullard *et al.*, 2010; Griffith *et al.*, 2010; Marioni *et al.*, 2008; Wang *et al.*, 2009). Depending on the sequencing used RNA-Seq has the

ability to analysis genomes from organisms without a complete genomic sequence, though those producing short reads can encounter problems in data analysis due to the lack of gene annotations or the lack of a sequenced genome (Vera *et al.*, 2008).

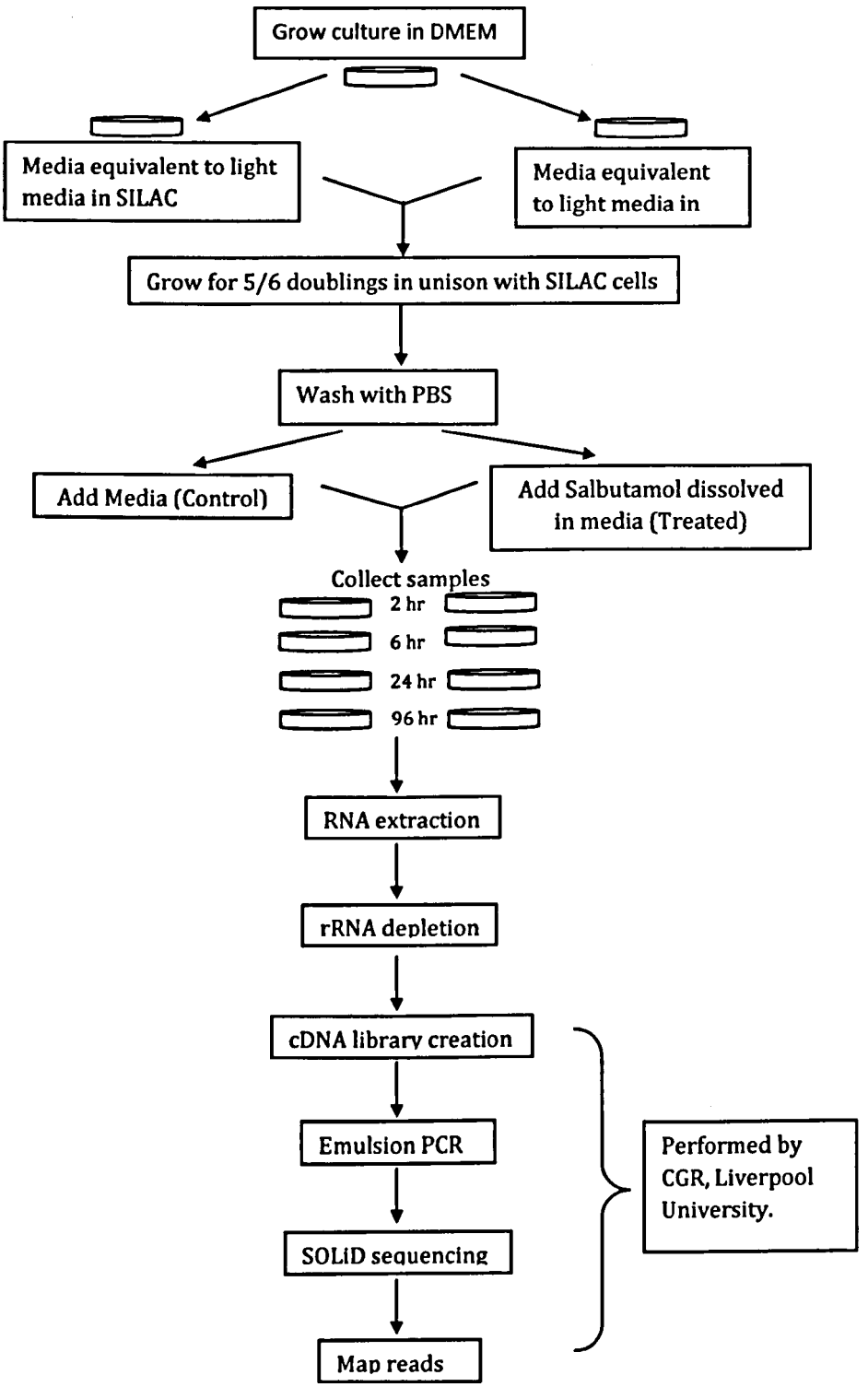



**Figure 3.1 Overview of transcriptomics workflow**

During the first half of the method the cells are grown at the same time as the SILAC and Lable Free cells in media equivelent to that of the SILAC light media to allow all three techniques to be compared. Once the cells have been grown and half treated with salbutamol samples can be collected at four different time points. Earlier time points are taken as a genomic change would be expected before a proteomic change. The creation of the cDNA library through to the sequencing and then the mapping of the reads obtained will be undertaken by the Center for Genomic Research at the University of Liverpool.

RNA-Seq can be used to quantify gene expression, as the number of mapped reads to a given gene or transcript is an estimation of the level of expression of that feature (Marioni *et al.*, 2008). As a quantitative technique, RNA-Seq is more accurate than microarrays as it has a large dynamic range which allows the quantification of very low and high expression levels (Mortazavi *et al.*, 2008). RNA-Seq can also give information about how two exons are connected using short reads, whereas using longer reads or pair-end short reads reveals information about connectivity between multiple exons. All of these factors make RNA-Seq useful for studying complex transcriptomes and informing genome annotations.

A treatment on any cell type can be accompanied by significant or subtle changes in the expression of many genes and/or their protein products. In general the transcript levels of genes are regulated through interplay of transcription factors, chromatin modifications and RNA degradation rate. By using this technique in this study, it is hoped that a better understanding of the effects of salbutamol on gene expression will be understood. This method will allow all genes to be relatively quantified between an untreated control and a treated sample highlighting any differences.

### 3.1.2 Muscle cell transcriptome

At the time of writing there was very little known about the effect salbutamol has on the mouse or rat muscle transcriptome. However, Sato *et al.*, (2008) showed that clenbuterol significantly increased the RNA concentration and the total RNA content of fast-twitch fibre rich extensor digitorum longus (EDL). It was also demonstrated that clenbuterol decreased the $\beta_2$-adrenoceptor mRNA expression of EDL muscle. It

has been found that the activation of cAMP response element binding protein (CREB) by protein kinase A (PKA) is related to β-adrenoceptor mRNA transcription (Collins *et al.*, 1990; Collins *et al.*, 1989; Mak *et al.*, 1995). Clenbuterol has also been seen to effect the expression of Cdkn1a and GHR, two genes that have well-established roles in muscle growth and development (Guo *et al.*, 1995; Mukherjee *et al.*, 2004). Spurlock *et al.*, (2006) additionally discovered three differentially expressed genes which are involved in critical steps of the biosynthesis of polyamines. Polyamines have been associated with cardiac hypertrophy resulting from treatment by β-agonists (Cubría *et al.*, 1998). Furthermore, they witnessed a general up-regulation of translational machinery occurring shortly after the administration of clenbuterol.

Pearen *et al.*, (2009) used Illumina BeadArray gene expression profiling to examine the gene expression in skeletal muscle in response to formoterol, another $\beta_2$-agonist. They found that the gene expression of *Stat3*, and *Smad3, Acvr2b* altered after the treatment. These three genes are directly associated with the regulation of muscle hypertrophy. Pearen *et al.*, (2009) also hypothesis that β-agonist treatment may enhance myogenesis, which could lead to proliferation, differentiation, and/or recruitment of satellite cells into muscle fibers to promote muscle growth. Significant changes are observed in *Itgb1bp3, Smad1, Smad3, FoxO1* and *Idb1*, all genes believed to have a role in regulating myogenesis. Their most interesting finding highlighted a limited number of significantly differentially expressed genes, which are regulated in the circadian transcriptome of adult mouse skeletal muscle. They then go on to implicate the β-agonist signalling pathway in coordinating communication between central and peripheral circadian clocks in skeletal muscle.

Although, there is limited information of the effects of β-agonists on a gene level there has been research published investigating the changes in transcription and translation on growing and differentiating muscle cells. Choi *et al.,* (2005) identified an increase in mRNA expression of *Fos, Jun, Cdkn1a, Ccna2, and Myod1* was shown to be associated with cell proliferation. Mancini *et al.,* (2011) have investigated the activation of the muscle-specific transcriptome using RNA interference (RNAi) during myoblast differentiation in C2C12 muscle cells. This cellular process involves expression of the muscle-specific transcriptional process and changes in cellular morphology. They also identified *Dbn1*, encoding drebrin, as a gene that is induced during myoblast differentiation. Their results in turn suggested that one or more p38-regulated transcription factors are involved, directly or indirectly, in the expression of *Dbn1* during myoblast differentiation. It is known that the p38 MAPK signalling pathway plays an important role in myogenic differentiation (Guasconi & Puri, 2009; Lluís *et al.*, 2006). Perdiguero *et al.,* (2007) demonstrated that p38α is the central p38 MAPK in myogenesis. In this chapter, fully differentiated cells will be investigated but it is possible that the β-agonist may cause further fusion of cells, which *Dbn1* and other genes involved in myogenesis, may be part of.

### 3.1.3 Aims

This chapter investigates the mouse transcriptome in an attempt to discover the effects of salbutamol in skeletal muscle cells at a gene level. Although not directly comparable it is hoped that the data collected may complement the data already collected using the DIGE technique, as well as identifying differences that may occur at a gene level compared to a protein level. These results will also be compared with

further proteomic work using the same cell line in other chapters. As a different cell system was used, the first aim was to optimise growth conditions for the immortalised cell culture system in a medium suitable for this experiment but also for the SILAC (Chapter 4) and label free (Chapter 5) proteomic analysis. The next aim involved producing enough depleted RNA to use for RNA-Seq. The last aim was to analyse the data and identify any differently expressed genes, what their biological meaning is and whether this correlates with any data already gained.

## *3.2 Method and Materials*

### 3.2.1 Cell Culture

#### 3.2.1.1 Growing C2C12 cells

C2C12 mouse cell line were grown as myoblasts in 25 $cm^2$ bottom (T25) vented cell culture flasks (BD Falcon™) and incubated at 37°C in a 5% $CO_2$ humidified incubator. Initially they were grown in filter sterilised DMEM medium (Sigma-Aldrich) supplemented with 10% (v/v) FCS and 1% (v/v) L-glutamine (Sigma-Aldrich) and 1% (v/v) penicillin-streptomycin (Sigma-Aldrich). A stock medium (DMEM low glucose minus arginine and lysine media (Sigma-Aldrich) including: 0.0159 g/L Phenol Red, 0.0001g/L proline, 0.105g/L L-leucine, 0.11g/L sodium pyruvate, 4.5g/L D-glucose, 1% (v/v) penicillin-streptomyocin, 10% (v/v) fetal calf serum) is initially assembled. Before the cells could be transferred L-arginine and L-lysine needed to be added to complete the medium. They were made up as stock solutions of 0.1g/ml in DMEM low glucose media. L-arginine 0.08% (v/v) and L-lysine 0.15% (v/v) from the stock solutions were combined with the stock medium. This media is suitable to use for other proteomic techniques (Chapters 4 and 5) to enable a fair comparison.

#### 3.2.1.2 Splitting Cells

C2C12 cells need to be split before they become confluent to sustain the culture, as they start to differentiate into myotubes once confluent. To split a 25 $cm^3$ flask of cells, the old medium was first removed and the flask washed with 5 ml of HEPES buffer (Sigma-Aldrich). The buffer was then disposed of and 5 ml of trypsin (Sigma-

Aldrich) was added and incubated for 2-7 min until the cells started to cleave away from the flask bottom. Cells were then extracted and placed in a universal tube. Next, 10µl cells were taken to use on a haemocytometer to count them. The remaining cells were spun at 1500 g for 5 min and then the trypsin is carefully discarded. Fresh medium in the volume of 5 ml was used to resuspend the cells. Another 5 ml of fresh medium was added to new flasks before the correct volume of cells were added to equal a concentration of $1 \times 10^4$ cells per flask.

A total of 64 flasks were prepared consisting of four biological replicates with eight flasks for the two conditions (treated and untreated). These eight flasks equalled two flasks for each of the four time points: 2 hr, 4 hr, 24 hr and 96 hr. Once the cells were confluent the medium was changed for lower serum medium (DMEM low glucose minus arginine and lysine media (Sigma-Aldrich) including: 0.0159 g/L Phenol Red, 0.0001g/L proline, 0.105g/L L-leucine, 0.084g/L arginine, 0.146g/L lysine, 0.11g/L sodium pyruvate, 4.5g/L D-glucose, 1 % (v/v) penicillin-streptomyocin, 2% (v/v) horse serum). By lowering the serum the cells were able to differentiate into myotubes. Cells were left for six doublings to become fully differentiated. Medium was changed every two days.

Fresh low serum medium was prepared and either salbutamol at a concentration of $1 \times 10^{-4}$ M, made in DMEM low glucose medium, or the same volume of just DMEM low glucose medium was added to the correct flasks. Treatment was staggered every 15 min to allow for collection, as described in 3.2.1.3.

#### 3.2.1.3 Harvesting cells

After 2 hr, 6 hr, 24 hr and 96 hr of treatment, cells were harvested from all four biological replicates. Harvesting cells from a 25 $cm^3$ flask of cells involved first removing the old medium and washing with 5 ml of HEPES buffer (Sigma-Aldrich). The buffer was then disposed of and 5 ml of trypsin (Sigma-Aldrich) was added and incubated for 4-7 min until the cells started to cleave away from the flask bottom. Cells were then extracted and placed in a universal tube and spun at 1500 g for 5 min before carefully discarding the trypsin. To wash the cell pellet 10 ml of 1 x PBS (0.14 M sodium chloride, 8 mM disodium phosphate, 1.5 mM mono-potassium phosphate and 2.7 mM potassium chloride) was added and the pellet resuspended. This was then spun for another 5 min at 1500 g at 4°C. The PBS was then removed and a fresh 1 ml added to resolubilise and transfer the cells to an ependorf tube. This time the cells were spun at 16000 g for 4 min. The PBS was removed and the pellet was snap frozen in liquid nitrogen and stored at -80°C.

### 3.2.2 Preparation of total RNA

#### 3.2.2.1 RNA extraction

RNA was extracted from the treated and control cell pellets using an RNeasy Kit (Qiagen) following the manufacture's protocol. Briefly, this involves first disrupting the cells by adding 300 µl of RLT buffer included in the kit. The lysate was directly pipetted into a QIAshredder spin column (Qiagen) and spun at full speed for 2 min, to help homogenize cells fully. One volume of 70% ethanol was mixed with the lysate and transferred to an RNeasy spin column provided within the kit. This is then centrifuge at 8000 g for 15 sec and any flow through discard. Next, 700 µl of wash

buffer RW1 is added to the spin column and spun for another 15 sec at 8000 g to wash the membrane. Again the flow through was discarded but being careful not to touch the column in the process. Then 500 µl of Buffer RPE is transferred to the spin column and spun for 15 sec at 8000 g. Once the flow through has been removed this wash step was repeated, but was spun for 2 min to ensure that no ethanol was carried over. The spin column was then placed into a new 2 ml collection tube and centrifuged for 1 min to eliminate any possible carry over. The spin column was then transferred to a 1.5 ml collection tube and 50 µl of RNase-free water was added to it. To elute the RNA the column was spun again for 1 min at 8 000 g. The eluted sample was then passed through the column again to gain a higher concentration of RNA. A 10 µl aliquot was removed to a separate tube to use to calculate the concentration as described below. All samples were stored at -80°C.

### 3.2.2.2 Agarose gel electrophoresis

A 1.5 % (w/v) agarose gel, containing ethidium bromide and 1 x TBE (0.4 M Tris-base, 0.44 M boric acid and 16 mM ethylenediaminetetraacetic acid (EDTA)) buffer, was used. 1 µl of each sample was mixed with 1 µl of sample loading buffer. The gel was run for 1 hr at 100 V to assess the integrity of the RNA.

### 3.2.2.3 RNA Quantitation and Qualitation

Quantitation of the total RNA for each individual sample, performed in triplicate, was conducted using 1 µl of each sample on a Nano-drop (Thermo Scientific) according to the manufacturer's instructions. Quality of individual samples was assessed with a Bioanalyzer (Agilent Technologies Inc.).

#### 3.2.2.4 DNase treatment

TURBO™ DNase (Ambion) was used to treat the sample before depletion to degrade DNA in the presence of RNA. This product has a greater efficiency than conventional DNase I and is RNase free. The sample was diluted to 10 µg nucleic acid/50 µl followed by adding 10 X TURBO DNase Buffer (Ambion) to form a 1X concentration with the sample. To complete the process, 1 µl of TURBO DNase for up to 10 µg RNA was mixed with the sample before incubating at 37°C for 30 min.

### 3.2.3 Preparation of depleted RNA for RNA-Seq

#### 3.2.3.1 Concentration of the RNA sample

To concentrate and clean the sample Invitrogen's RiboMinus™ Concentration Module was used. To the DNase treated sample Binding buffer (Invitrogen) of equal volume to the sample was added along with 1X volume of sample of 100% (v/v) ethanol. The solution was loaded on to the column and centrifuged at 12,000 g for 1 min at room temperature and the flow through discarded. Next, 200 µl of buffer (supplied) was transferred to the column and spun at 12,000 g for 1 min at room temperature. The wash step was repeated and then centrifuged again for 3 min at maximum speed to remove any residual wash buffer. The column was placed into a clean recovery tube and 50 µl of RNase-free water added to the centre of column before incubating at room temperature for 1 min. The column was finally spun at maximum speed for 1 min.

### 3.2.3.2 rRNA depletion

Ribosomal RNA needs to be depleted in order for other less abundant RNA to be detected and analysed via RNA-seq. Depletion was undertaken using the RiboMinus™ Eukaryote Kit for RNA-Seq (Invitrogen A10837-08).

For each time point equal amounts of the four biological replicates were combined to give a total of 10 µg in 10 µl. The manufacture's guidelines were followed to use the depletion kit, which starts with a hybridization step. The extracted RNA was combined with 10 µl of RiboMinus probe (15 pmol/µl) and 100 µl of hybridization buffer in a sterile, RNase free 1.5 ml microcentrifuge tube. This was incubated for 5 min at 70-75 °C to denature the RNA. The sample was then allowed to cool to 37°C slowly over a period of 30 min by being placed in a water bath to promote sequence-specific hybridization. Magnetic beads and a magnetic separator are essential to this procedure. The beads bind to 5S, 5.8S, 18S and 28S ribosomal RNA allowing separation. The beads need to be prepared before use, which involves taking 750 µl of beads per sample and washing them twice with DEPC water. The magnetic separator was used to isolate the beads from the supernatant allowing it to be discarded. The beads were resuspended in hybridization buffer and separated into 250 µl and 500 µl and kept at 37 °C until needed. The tube that contained 500 µl of beads was again placed on the magnetic separator to allow the supernatant to be discarded and the beads resuspended in 200 µl of hybridisation buffer. Once the sample had cooled it was transferred to tube containing 250 µl of the prepared beads and incubated at 37 °C for 15 min. Again the magnetic separator was used to allow collection of the supernatant which contains the RiboMinus RNA. The supernatant was then mixed with the tube containing 200µl of the prepared beads and the process

repeated to ensure complete depletion. Each time point was depleted separately. A total of two 10 µl aliquots of each time were depleted. Once depletion had occurred the supernatant from both rounds of depletion were mixed together before being concentrated down to ensure there was enough mRNA for SOLiD sequencing. To normalise the data equal amounts of mRNA from all four biological replicates are mixed together to be run as one sample on the SOLiD sequencer.

#### 3.2.3.3 Concentrating depleted RNA

To concentrate the 300 µl of depleted RNA into a smaller volume a method provided in the RiboMinus™ Eukaryote Kit for RNA-Seq was used. To each sample 1 µl of a 20 µg/µl solution of glycogen (Invitrogen), 1/10$^{th}$ of the sample volume of 3 M sodium acetate (Invitrogen) and 2.5 x the sample volume of 100% (v/v) ethanol was added. This mixture was incubated for a minimum of 30 min at -80°C. After incubation, samples were centrifuged for 15 min at 12,000 g at 4°C and then the supernatant discarded without touching the pellet. Samples were then rewashed in 500 µl of 70% (v/v) cold ethanol. This was spun for 15 min at 12,000 g at 4°C, the supernatant discarded and the process repeated. The pellet was then left to air-dry before being resuspended in 10 µl of DEPC-treated water and stored at -80 °C.

### 3.2.4 Creation of cDNA library

The creation of the cDNA libraries, emulsion PCR (ePCR) and SOLiD sequencing was performed by Miss Pia Koldkjaer and Mr John Kenny at the Centre for Genomic Research, School of Life Sciences, University of Liverpool. Briefly this involved firstly fragmenting the RNA using the SOLiD™ Total RNA-Seq Kit. RNase III from

this kit is used to fragment the RNA into a suitable size for SOLiD sequencing. Once the RNA was cleaned using a Ribominus Concentration Module (Invitrogen K1550-05), the Quant-iT™ RNA Assay Kit with the Qubit® Fluorometer (Invitrogen) and the RNA 6000 Pico Chip Kit with the Agilent® 2100 Bioanalyzer (Agilent) was used to analyse the yield and size distribution of the fragmented RNA.

Once fragmentation was completed sufficiently the RNA could then be hybridised and ligated. This was achieved using an adaptor mix from the SOLiD™ Total RNA-Seq Kit. The adaptor mix contains a set of oligonucleotides with a single-stranded degenerate sequence at one end and a defined sequence required for the SOLiD system sequencing at the other end.

This was followed by reverse transcription and purification of the RNA. A reverse transcription master mix was made for each of the samples comprising of nuclease-free water, reverse transcription buffer, dNTP mix, SOLiD™ RT Primer and ArrayScript™ Reverse Transcriptase from the SOLiD™ Total RNA-Seq Kit. The master mix was then added to the RNA ligation reaction and incubated to allow the reaction to take place. Once this process was completed the cDNA was purified using MinElute® PCR Purification Kit (Qiagen). In order to amplify the cDNA of the correct size the samples must be run on a gel and bands of the correct size for SOLiD sequencing excised. The cDNA needed resides between 150–250 nucleotides (nt) on the gel. Once this material had been collected the cDNA was amplified to ensure there was sufficient amount for SOLiD sequencing. Again more components from the SOLiD™ Total RNA-Seq Kit where used. To prepare sufficient cDNA for the emulsion PCR step of SOLiD sequencing two PCR reactions were performed for

each sample each using one of the gel pieces excised during size selection. The amplified cDNA was purified using the PureLink™ PCR Micro Kit (Invitrogen) to remove unincorporated primers which may have affected the final quantitation and emulsion PCR. At this point the two PCR reactions are combined. Again the NanoDrop spectrophotometer and the 2100 Bioanalyzer with the DNA1000 Kit (Agilent) was used to assess the yield and size distribution of the amplified DNA.

### 3.2.5 Emulsion PCR

When less than 20% of the amplified DNA was in the 25–200 bp range, the SOLiD™ System templated bead preparation stage was performed. This was where each library template was clonally amplified on SOLiD™ P1 DNA Beads by emulsion PCR. Emulsions are made up of an oil phase containing emulsifiers and an aqueous phase, which includes PCR components. The oil was used to create droplets of solution which aim to contain a single DNA bead and a single DNA fragment template. The emulsion was then placed on a thermal cycler and a standard PCR conditions run according to manufacturer instructions and during this amplification >30 000 copies of each template were amplified onto each SOLiD P1 DNA bead. After the emulsion PCR was complete, 2-butanol was used to break the micro-reactors in the emulsion. The templated beads and non-amplifying beads were then washed to clear away the oil and emulsifiers. An enrichment step was undertaken to isolate the templated beads from non-amplified or poorly amplified beads. In the enrichment step, polystyrene beads with a single-stranded P2 Adaptor attached were used to capture templated beads. The mixture of enrichment beads, enrichment bead-templated bead complexes, and non-amplifying beads were centrifuged on a 60%

(v/v) glycerol cushion. The enrichment beads were extracted and denatured using denaturing buffer to dissociate the templated beads from the enrichment beads. The enriched DNA beads were then prepared for the SOLiD sequencer by covalent bonding to a glass slide.

### 3.2.6 SOLiD sequencing

Primers hybridize to the P1 adapter sequence on the templated beads added during the ePCR step. A set of four fluorescently labeled di-base probes were then added. These probes compete for ligation to the sequencing primer. The fluorescent tag attached to the probe was cleaved once the di-base probe had ligated to the sequence. When the tag is cleaved it emits a fluorescent colour which is recorded. Multiple cycles were performed to repeat the process to extend the sequence. The extended sequence was then removed from the template sequence and a new primer added to the adapter sequence. However this primer attached at a position n-1 of the first primer. The di-base probes were again added and the ligation cycle starts over. As before, once the sequence had been extended it was again removed and the whole process repeated with another three primers that are positioned at n-2, n-3 and n-4 from the original primer. This allows for each base of the template sequence to be analysed twice to ensure any errors in sequencing are minimal.

### 3.2.7 Read mapping

When the reads had been generated by the SOLiD sequencer they were then mapped onto the mouse genome. This process was performed by Dr. Xuan Liu at the Centre for Genomic Research, School of Life Sciences, University of Liverpool. The mouse

genome (mm9) was downloaded from UCSC genome browser ((http://genome.ucsc.edu/cgi-bin/hgGateway?hgsid=241934113&clade=mammal&org=Mouse&db=0). The Bioscope pipeline was used to map the SOLiD reads onto the mouse genome.

### 3.2.8 Analysis

A list of counts of each gene for both control and treated samples at all time points was compiled. The treated counts were normalised using the log ratios of all genes with a count over 50. Once data had been normalised the fold change was calculated to identify which genes had been modulated significantly. Gene ID's for any gene that appeared in all four time points with a read count over two and a fold change greater than two were uploaded on to the DAVID site (http://david.abcc.ncifcrf.gov/) and data on functions and pathways were collected. A software package called R (http://www.r-project.org/) was used to create heatmaps of the differentially expressed genes using their log ratios at all four time points. There are several systems and programmes to determine differential expression including; poisson distribution (Oshlack *et al.*, 2010), DESeq (Anders & Huber, 2010) and Edge R (Robinson *et al.*, 2010). As the biological repeats were combined into one sample, therefore producing one technical run a different approach to the statistical analysis was performed using a recommended statistical package. The control count data for the four time points were used as replicates, allowing each treated sample set to be compared to the four controls which allowed an exact t-test to be performed using EdgeR (http://www.bioconductor.org/packages/2.3/bioc/html/edgeR.html) which determined the genes that were differential expressed. Once the files containing the

counts had been uploaded into the software the low counts were removed. Next, the normalisation factors were calculated, followed by estimating the common dispersion. An exact test was performed on this data and a list of differentially expressed genes with a p value <0.01 was compiled. A pathway analysis software KEGG (http://www.genome.jp/kegg/) allowed the lists of identified genes to be visualised in the pathways they belonged to.

## *3.3 Results*

### 3.3.1 Analysis using the Database for Annotation, Visualisation and Integrated Discovery (DAVID)

DAVID was used to allow visualization of functional classifications, biochemical pathway maps and conserved protein domain architectures, which are linked to biological annotations for all four time points.

Data gained from the RNA-Seq gave the number of read counts detected for every gene in the mouse proteome. These read counts were the basis of determining modulation of genes. Genes identified in all four time points, with a read count over two and a fold change > 2, were analysed via DAVID site first, as it is assumed that if the gene is modulated across all time points there is a greater chance this modulation has actually occurred rather than occurring by chance. However, when assessing the data there was little data with a p value < 0.01. The most significant information gained from the annotation chart was that phosphoproteins may be involved. When the criterion was tightened so that only genes with a count over 20 reads and a fold change >2 are included there are no significant results.

To examine any relevance the genes modulated have across the time points (2 hr, 6 hr, 24 hr and 96 hr) the analysis was repeated with a list of gene ID's that were seen in three or more time points with counts >2 and a two-fold change. This time DAVID identified, via the annotation chart, that: the most enriched biological processes involved the DNA metabolic process, replication and transcription; the cellular components most involved were the non-membrane bound organelles,

mitochondrion and the ribosome, as well as the nuclear lumen; the molecular functions most likely to be modulated were DNA binding, transcription regulator activity and damaged DNA binding. Clusters for this data set included: "non-membrane-bounded organelles", "ribosomal proteins", "DNA replication" and "regulation of transcription" with enrichment scores of 5.3, 3 and 2.4 respectively. With counts over 50 DAVID produced a stricter analysis of the modulated genes across the time points. Under these parameters, it showed that DNA replication was the main biological process. However, the molecular function involved DNA binding, which was the same as the previous parameters, but does also highlight phosphostransferase activity as being modulated. The clusters identified not only have much lower enrichment scores but are similar to the previous findings. "DNA replication" was the most enriched with a score of 1.6, followed by "ribosomal proteins" with enrichment scores of 1.3.

As it appeared that the modulation of the genes at the four time points were not necessarily related, individual times points were then analysed to access any modulation differences. At 96 hr, when analysing genes with a count > 20 and a fold change > 2, there were over 50 biological process affected with a p value < 0.01 (Appendix II, Table 1-4). This allowed the criteria to be further tightened by focusing on the results with a DAVID count greater than 100 as well as a p value < 0.001. These indicated that regulation of transcription was being effected, as well as protein localisation and transport, macromolecular catalytic process and the intracellular signalling cascade. The cellular components identified as being most abundantly involved using the same standards as above are; the lumen including membrane-enclosed, intracellular organelle and nuclear lumen, the mitochondria, non-

membrane bound organelles, the nucleoplasm and the golgi apparatus. Ion binding, nucleotide binding, DNA binding and adenyl nucleotide binding were displayed as the most modulated molecular functions for this data set. When compared to the data set at 2 hr, with the same criteria, very similar top hits for all three categories were recorded despite there being 423 more genes modulated at 96 hr. At 6 hrs analysis also identified transcription as the biological process modulated, DNA binding as a molecular function involved and non-membrane bound organelles used. Although a number of results were given at 6 hr the majority of them did not have a DAVID count >100. As DAVID will not allow over 3000 genes to be uploaded at one time the data set for 24 hr could not be compared, so the fold change criteria was increased to 2.5. This data demonstrated that again very similar functions were affected, though those with the lowest p value highlight catabolic process. If the limit of 100 DAVID counts is lifted then the results with lower DAVID counts are often different. For example, apoptosis functions were seen at 24 hr but there were more cell cycle processes at 6hr. A summary is listed in Table 3.1 of the top three DAVID results for time points 2 hr, 6 hr and 96 hr with parameters of read counts of 20, two-fold change, p value < 0.01 and DAVID counts >100 as well as 24 hr with parameters of read counts of 20, 2.5 fold change, p value < 0.01 and DAVID counts >100. As stated previously a full set of results are listed in the Appendix II.

**Table 3.1** ***DAVID results when examining the modulated genes found at different time points***

Genes seen to be modulated in mouse muscles cells, when treated with salbutamol and collected at different time points were uploaded in to DAVID. To create the list of modulated genes a parameter was set to only include genes with a read count >50 and a two-fold change between control and treated cells. An exception was made for 24 hr as this gave a list too large for DAVID. In this case the parameter was increased to read counts greater than 50 and a fold change of 2.5. A list of biological process, cellular components and molecular functions were obtained. A summary of these results are shown, comprising of the top three results for each section with a p value less than 0.01 and a DAVID count greater than 100. At 6 hr the majority of results had a count lower than 100.

| | Top biological Processes | | | Top cellular components | | | Top molecular function | | |
|---|---|---|---|---|---|---|---|---|---|
| | | P value | Count | | P value | Count | | P value | Count |
| 2 hr | Transcription | 1.398E-13 | 256 | Membrane-enclosed lumen | 1.36E-15 | 181 | Zinc ion binding | 1.45E-10 | 278 |
| | Regulation of transcription | 2.081E-07 | 276 | Mitochondrion | 1.845E-15 | 197 | Transition metal ion binding | 1.369E-10 | 331 |
| | Protein localisation | 0.000123 | 102 | Intracellular organelle lumen | 3.626E-15 | 175 | Cation binding | 1.583E-07 | 442 |
| 6 hr | Transcription | 0.000372 | 103 | Intracellular non-membrane-bounded organelle | 7.575E-08 | 117 | DNA binding | 2.684E-05 | 105 |
| | Regulation of transcription | 0.000952 | 122 | Non-membrane-bounded organelle | 7.575E-08 | 117 | | | |
| | | | | | | | | | |
| 24 hr | Protein localisation | 1.389E-07 | 124 | Membrane-enclose lumen | 2.544E-17 | 209 | Nucleotide binding | 5.364E-10 | 299 |
| | Macromolecule catabolic process | 2.557E-07 | 110 | Intracellular organelle lumen | 6.892E-16 | 200 | ATP binding | 3.958E-08 | 205 |
| | Cellular macromolecule catabolic process | 2.79E-07 | 104 | Organelle lumen | 9.162E-16 | 200 | Adenyl ribonucleotide binding | 6.292E-08 | 206 |
| 96 hr | Transcription | 2.121E-08 | 279 | Membrane-enclosed lumen | 5.103E-21 | 236 | Transition metal ion binding | 1.359E-12 | 400 |
| | Cell cycle | 1.245E-07 | 114 | Intracellular organelle lumen | 2.404E-20 | 228 | Zinc ion binding | 1.812E-12 | 335 |
| | Regulation of transcription | 1.02E-06 | 324 | Organelle lumen | 3.888E-20 | 228 | Cation binding | 3.707E-10 | 543 |

When the criterion was increased to only include genes with a read count greater than 50, the 24 hr data set could also be compared via DAVID using a 2 fold change. For this data any clusters with a p value greater than 0.01 were examined. The majority of the clusters were the same over the three time points suggesting the same pathways and functions were affected by the drug, but it may be affecting different genes within the one pathway explaining the difference in the number of genes modulated. There were a few clusters highlighted in each time point which were not seen in either of the other data sets. At 2 hr DAVID identified "DNA replication", "zinc fingers" and "purine nucleoside binding" as being unique. Whereas, at 6 hr there was an effect on the "ribosome", "protein localization", "DNA repair" and "apoptosis" which was not seen at 2hr or 96hr. When the same analysis was performed on the 96 hr data set DAVID suggests that "non-membrane bound organelles", "transcription regulation", "chromatin modification", "limb morphogenesis", "nucleotide binding", "transcription activator activity", "negative regulation of transcription" and "protein-tyrosine phosphatise" were only modulated at this time point after treatment with salbutamol.

### 3.3.2 Using Kyoto Encyclopedia of Genes and Genomes (KEGG) to link modulated genes in pathways

By using Kegg to analyse each individual time point a better understanding on which pathways were affected and if there is a correlation across the time points could be gained. Gene lists with counts greater than 50 and a 2 fold change were uploaded onto the KEGG Mapper. A total of 102 genes at 2 hr were involved in metabolic processes, whereas 41, 202 and 146 genes are identified as modulated in the same overview for 6 hr, 24 hr and 96 hr respectively. The MAPK pathway was seen at all

time points to contain a large portion of the modulated genes identified. Figure 3.2 illustrates the genes present in this pathway that have been identified as significantly changing using data from all time points. Both the JAK-STAT and the ErbB signalling pathways, which lead on to the MAPK pathway, are identified with high number of genes modulated by KEGG. Another pathway identified, which has 66 genes in the pathway modulated at one or more of the time points, is the regulation of actin pathway (Figure 3.3). This is another pathway that is expected to be highly modulated in changing muscle cells. A large number of other pathways that would be expected to be involved in growing muscles in response to a stimulus were identified with several modulated genes, for example; purine metabolism, ubiquitin mediated proteolysis, endocytosis, cell cycle, insulin signalling pathway, calcium signalling pathway and glycerophospholipid metabolism. There were several pathways highlighted that suggest the cells may also have been under stress such as; apoptosis, p53 pathway and wnt signalling pathway. There were also a significant number of pathways seen to be involved in cancer or infection.

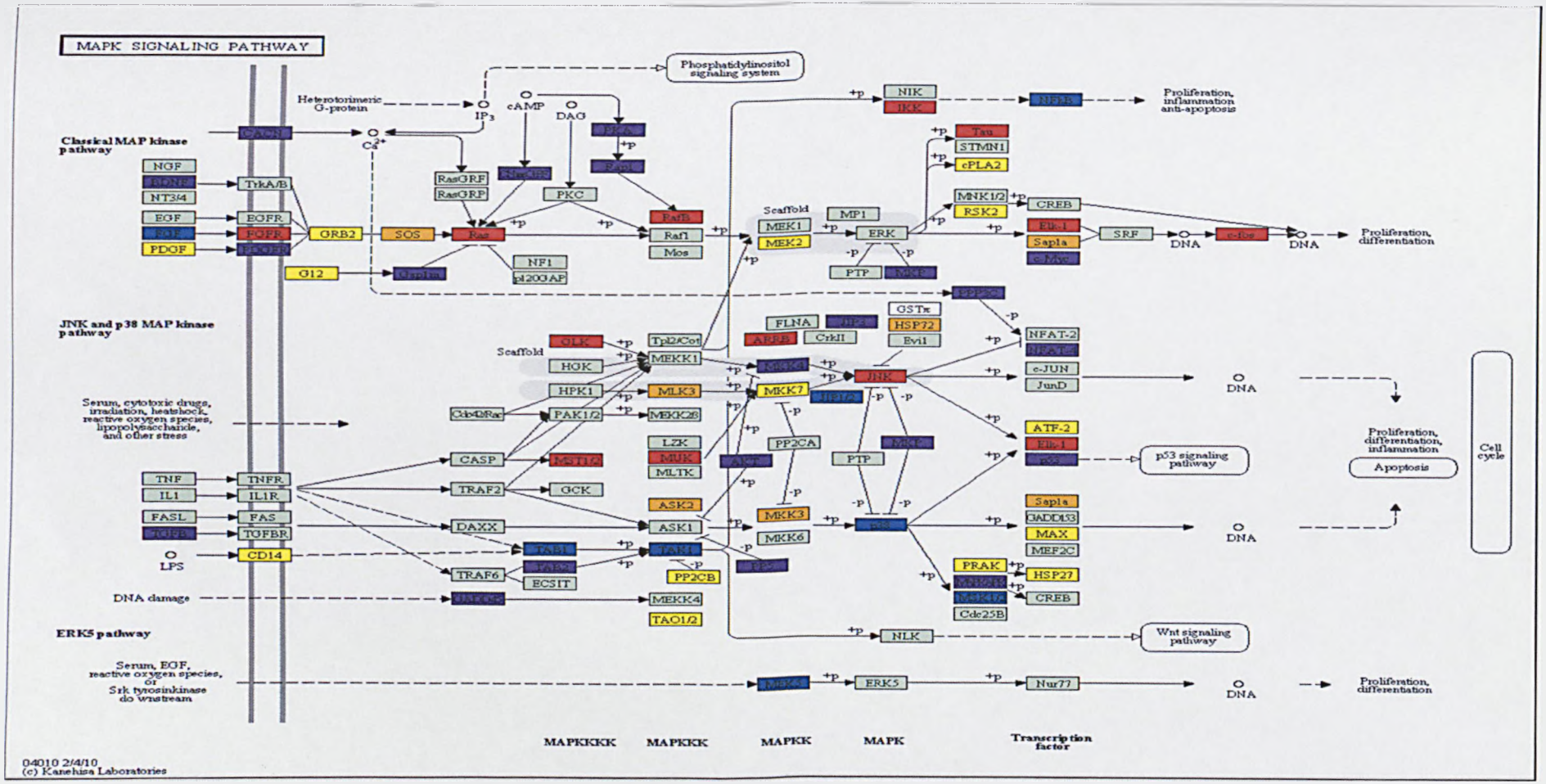



**Figure 3.2** ***MAPK signalling pathway identifying modulated genes using KEGG Mapper***

Using the KEGG Mapper a list of genes from all four time points (2 hr, 6 hr, 24 hr and 96 hr), with read counts over 50 and a two-fold change between the control and treated, were annotated to different pathways. A total of 82 genes were identified as being present in the MAPK signalling pathway. Genes are labelled corresponding to the four data sets they belong to; 2 hr – Red, 6 hr – Orange, 24 hr – Purple and 96 hr – Blue. Yellow nodes represent genes that were seen in more than one time point as changing.

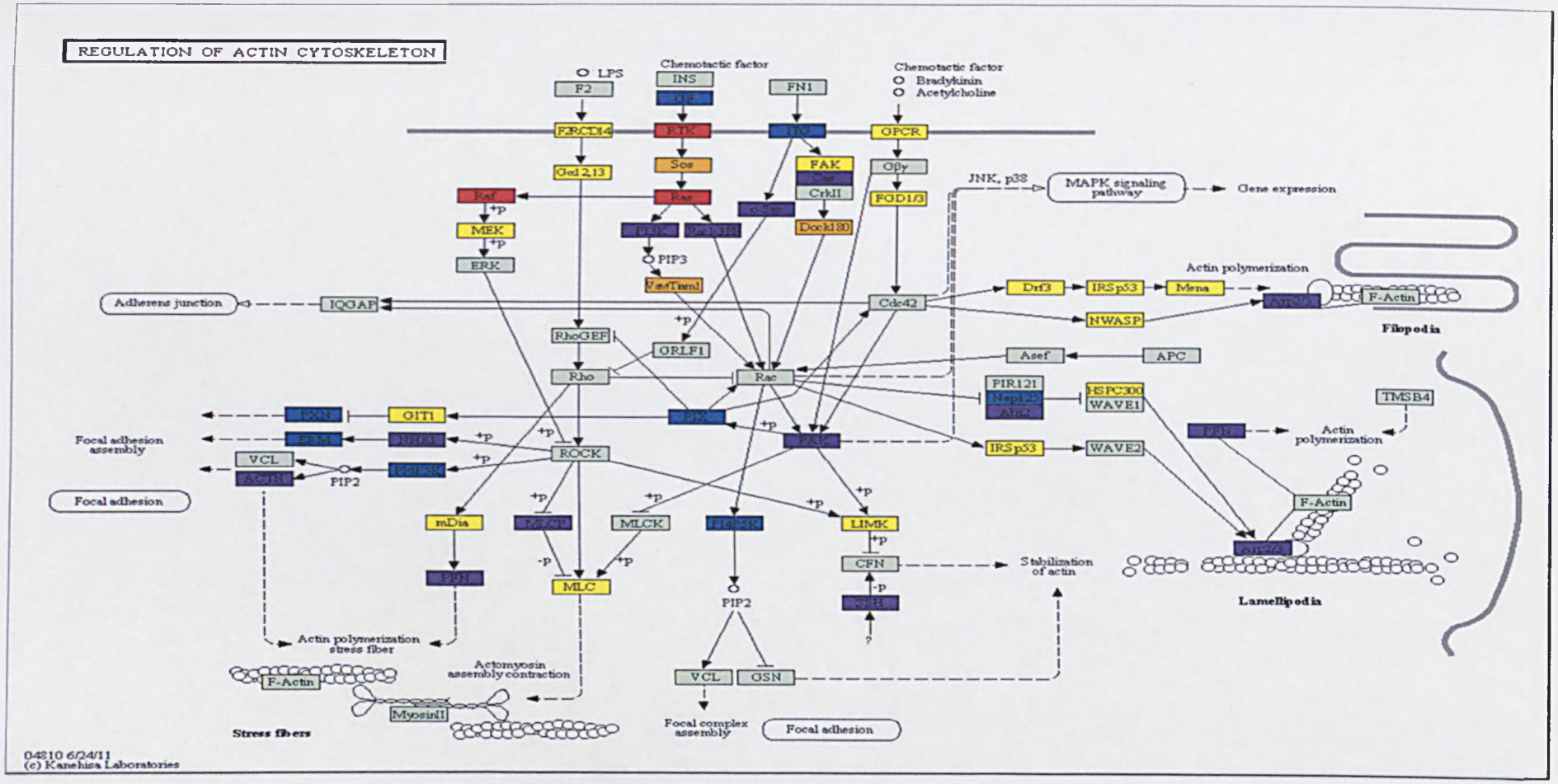



**Figure 3.3** ***Regulation of actin pathway identifying genes modulated in mouse muscle cells after treatment with salbutamol.***
Using KEGG Mapper a list of modulated genes (read counts over 50 and a two-fold change between the control and treated samples) from all four time points (2 hr, 6 hr, 24 hr and 9 6hr), were annotated to different pathways. A total of 66 genes were identified as being present in the actin regulation pathway. Genes are labelled corresponding to the data sets they belong to; 2 hr – Red, 6 hr – Orange, 24 hr – Purple and 96 hr Blue. Yellow nodes represent genes that are modulated in more than one time point.

### 3.3.3 Visualisation of genes modulated from all time points

To allow any patterns in gene expression over time to be identified, a heat map was constructed using the gene lists used in the previous analysis (count great than 50 with a two-fold change). There were 1399 genes identified as being modulated at 2 hr with a count of over 50 reads and with a two-fold change. With the same parameters there were 519 genes at 6 hr, 2600 genes at 24 hr and at 96 hr there were 1873 genes. A heat map (Figure 3.4) of all genes with a count greater than 50 reads, with a two-fold change and seen in two or more time points was created using R, a statistical analysis program. The genes up-regulated in the treated cells are shown in green, those down-regulated in red and any in black show no change.

From the heat map (Figure 3.4) it can be ascertained that there were very few patterns in modulation across the time points. There were a large group of genes highly up-regulated at 24 hr with a similar number of genes down-regulated but less intensely. There seems to be a similar number of down-regulated genes at 96 hr but they were different to the ones seen at 24 hr. This heat map only groups genes with similar log ratios so it is unclear whether there are patterns within genes involved in similar functions.

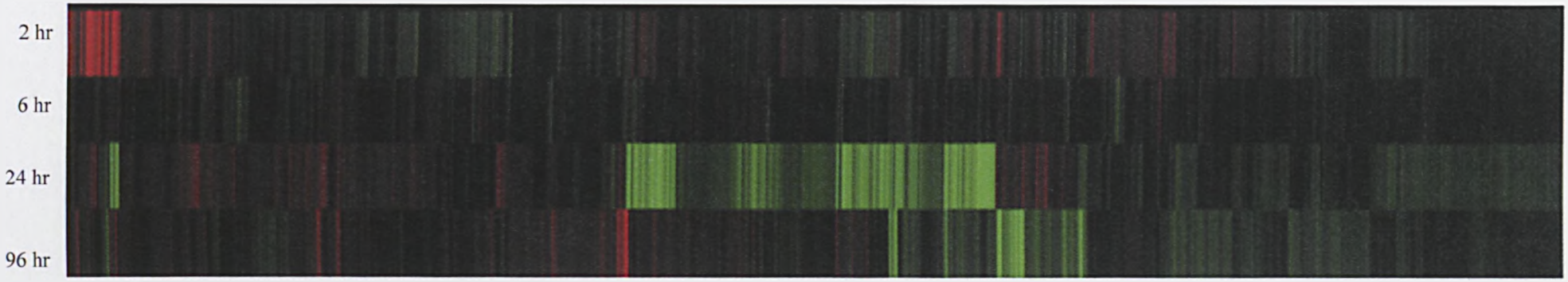



**Figure 3.4** ***Heat map displaying the modulation of all the genes identified at four time points after treatment with salbutamol.***
C2C12 mouse muscle cells were treated with salbutamol and collected after 2 hr, 6 hr, 24 hr and 96 hr after treatment. Samples were depleted, cleaned and concentrated in order to be sequenced using the SOLID sequencer at the University of Liverpool. The results were mapped to the mouse database and a list of identified genes and their corresponding read counts obtained. Fold changes were calculated for treated compared to control sample counts. Any gene with a fold change >2 with a read count >50 and seen in two or more time points are shown. The heat map was created using R and the log fold change for each of the genes. Those shown as green are thought to be up-regulated and those in red down-regulated. The genes are clustered due to their fold change.

### 3.3.4 Further statistical analysis using EdgeR

To gain better statistical validation of the results aquired from RNA-Seq the software package EdgeR was used. Normally a statistical analysis would involve comparing repeats with each other to discover the p value, allowing a decision to be determined on whether true differences are seen. As the four biological replicates were combined before running on the SOLiD sequencer it is not possible to do this. However, by comparing each individual treated data set against the four control data sets, one from each time point, a variant of the statistical analysis explained previous can be undertaken. As the control data sets should stay the same over the time course they were used in this analysis as repeats. Using this analysis for the four time points; 2 hr, 6 hr, 24 hr and 96 hr, a list of 161, 299, 2756 and 1346 differentially expressed genes respectively were found.

When comparing the lists from each time point with each other, there were two genes which are seen in all four time points. These genes code for iodothyronine deiodinase type I and G protein-coupled receptor 141. A total of 38 genes were identified in three or more of the time points. This gene list was also uploaded into DAVID to determine if similar results as previously described were identified. Only five clusters are identified for these 38 genes; "cell adhesion", "ion binding", "alternative splicing" "membrane" and "glycosylation and signal peptides". However, the enrichment scores for these clusters range between 1.18 and 0.05, which is very low. When the gene list was increased to include all genes seen in more than one time point, DAVID identified nine clusters but with higher enrichment scores. The top two clusters were "ion binding and transcription" and "DNA-binding" with scores of 9.97 and 5.57. These enrichment scores are higher than those found using a larger

data set in the previous DAVID analysis (Section 3.3.1.). Next, analysis of gene lists identified as being statistically different for the individual time points were completed. Using the 2 hr data set, biological processes with a p value < 0.01 involved those in the positive regulation of transcription. The biological processes identified as being the most significant (p value <0.01 and DAVID counts >100) at 24 hr and 96 hr also included transcription, whereas at 6 hr it is cell-cell adhesion which is flagged. Interestingly, DAVID identified (p value < 0.01) the molecular function of steroid hormone receptor activity at 2 hr. However, in all other time points the molecular functions identified were related to ion binding or DNA/RNA binding.

There were a total of 696 genes that were found in two or more time points. If genes are seen to be modulated at two consecutive time points, it may indicate the constant modulation of this gene after a certain time. There are 587 genes modulated at both 24 hr and 96 hr. At 2hr and 6hr there are 29 genes and between 6 hr and 24 hr there are only 12 genes modulated. Appendix II Table 5 indicates how the 785 genes are differently modulated across the four time points.

To determine whether the statistical analysis has had any effect on the protein pathways identified earlier, the new lists of modulated genes were uploaded to KEGG mapper. The MAPK pathway was again identified as a pathway being modulated. This time only 46 genes were modulated in the pathway (Figure 3.5). The majority of the modulated genes were seen at 24 hr.

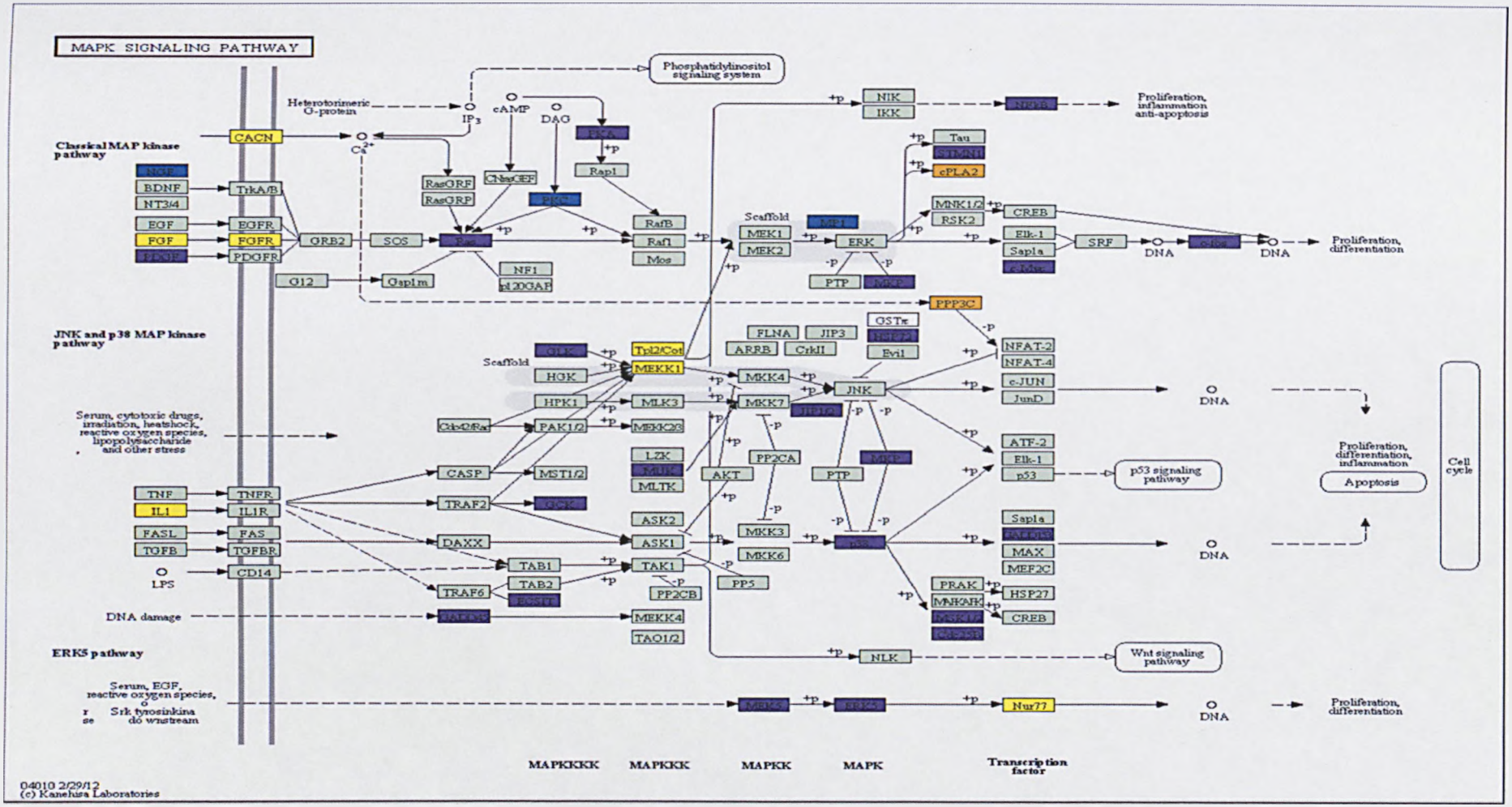



**Figure 3.5** ***The MAPK pathway with highlighted genes which are modulated in mouse muscle cells after treatment with salbutamol.***
Using KEGG Mapper a list of modulated genes (p value <0.01) from all four time points (2 hr, 6 hr, 24 hr and 96 hr), were annotated to different pathways. A total of 46 genes were identified as being present in the MAPK pathway. Genes are labelled corresponding to the data sets they belong to; 2 hr – Red, 6 hr – Orange, 24 hr – Purple and 96 hr – Blue. Yellow nodes represent genes that are modulated in more than one time point.

The KEGG results also suggest that a number of the genes code for proteins involved in cytokine-cytokine receptor interaction, another process expected to be seen in cells changing their morphology and being treated with a chemical. In this analysis, only 30 genes were modulated in the pathway responsible for actin regulation, whereas, 36 genes were seen in the cell cycle pathway (Figure 3.6). Once again the majority of genes identified were seen in the 24 hr data set. Many of the resulting pathways produced by KEGG were the same as the first analysis but with fewer genes identified as significantly different.

The overall the results from this experiment indicate that a change is occurring within the mouse muscle cells after treatment with salbutamol. It also seems that many pathways are involved with implementing this change in morphology, which would be expected in a large physical change.

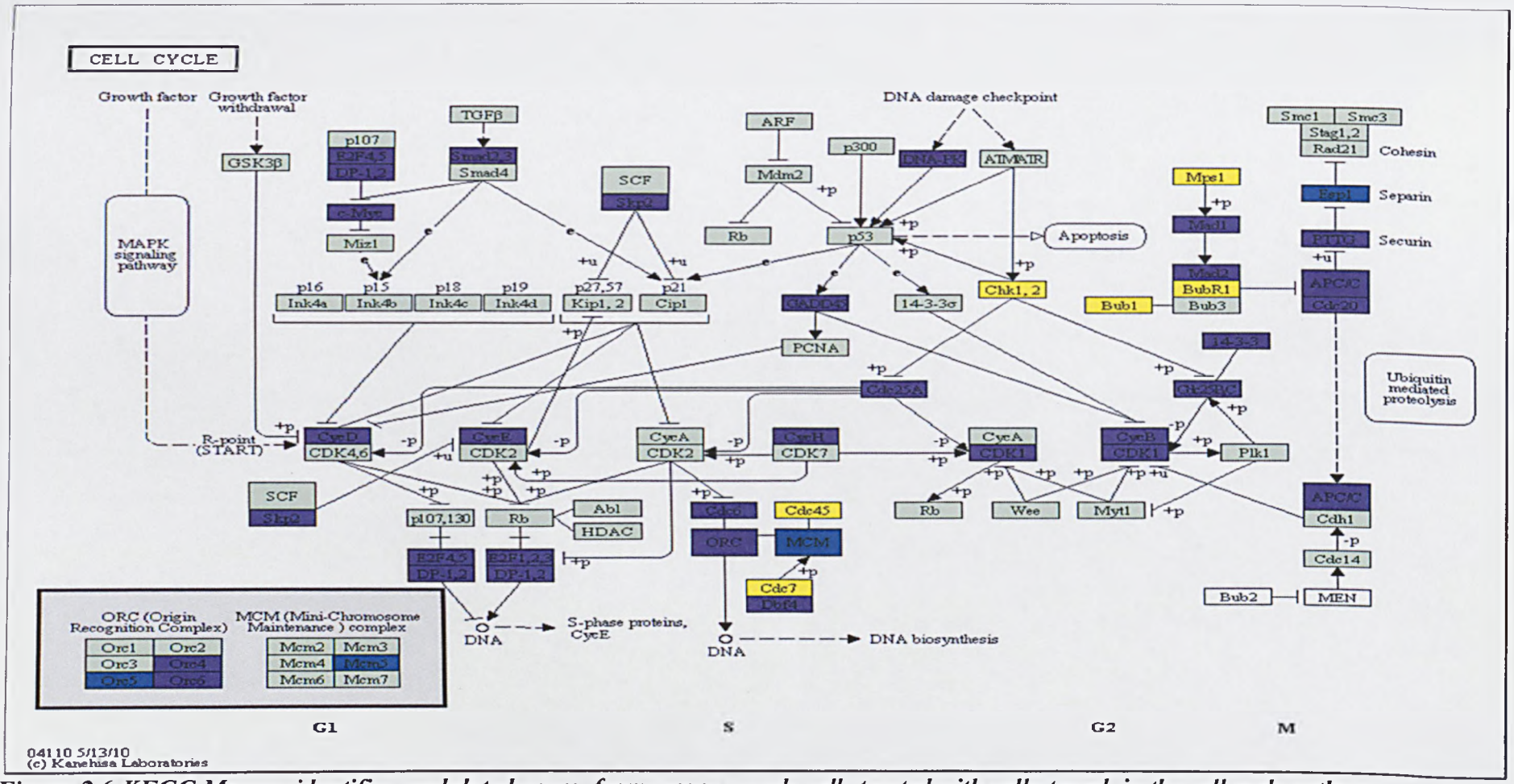



**Figure 3.6** ***KEGG Mapper identifies modulated genes, from mouse muscle cells treated with salbutamol, in the cell cycle pathway.***
Using KEGG Mapper a list of modulated genes (p value <0.01) from all four time points (2 hr, 6 hr, 24 hr and 96 hr), were annotated to different pathways. A total of 36 genes were identified as being present in the cell cycle pathway. Genes are labelled corresponding to the data sets they belong to; 2 hr – Red, 6 hr – Orange, 24 hr – Purple and 96 hr – Blue. Yellow nodes represent genes that are modulated in more than one time point.

## 3.4 Discussion

In this chapter, RNA-Seq has been used to characterise any changes in the transcriptome of mouse muscle cells after treatment with the β-agonist, salbutamol. By using the SOLiD sequencing platform it was possible to relatively quantify transcripts from C2C12 cells treated after differentiation with salbutamol, across four time points. This technique allowed the identification of several thousand genes being modulated within mouse immortalised muscle cells after treatment. Differential modulation was determined by comparing the number of reads counted for each gene within the whole mouse genome. Statistical analysis was performed which reduced the list of identified genes but still identified over 700 genes seen to be modulated at more than two time points ($p$ value $< 0.01$).

### 3.4.1 Identification of differentially expressed genes

Data gained after RNA-Seq comprises a list of all genes within the mouse genome and the number of identified reads for each gene under different conditions at four time points. It was assumed that similar number of reads would be mapped to its gene, under different conditions, if there was no difference between the cells. The *read count* has been found to be linearly related to the abundance of the target transcript (Mortazavi *et al.*, 2008). The aim of the experiment was to use these read counts to decide whether a given gene is differentially expressed. This means that the observed difference in read counts must be significant, for example, whether it is greater than what would be expected due to natural random variation. However, identifying which genes have true changes becomes difficult when no technical replicates have been undertaken. In this experiment, four biological replicates were

mixed prior to sequencing to help normalise any differences. Normalization enables a more accurate comparison of the expression levels between samples by removing systematic technical effects that occur in the data to ensure that technical bias has minimal impact on the results.

There have been several studies (Anders & Huber, 2010; Bullard *et al.*, 2010; Oshlack *et al.*, 2010; Robinson & Oshlack, 2010) investigating the importance of using the correct statistical analysis and normalising data after using next generation sequencing techniques. Two techniques were undertaken in this study to identify differentially expressed genes, including normalisation of the data sets. The original analysis of the data involved normalising the read counts of the treated cells using the log ratio between the two conditions. This allowed the normalised fold change to be calculated and therefore identify any changes between conditions. Deciding on whether all genes should be evaluated or only those with a read count above a certain threshold for both conditions was the next problem encountered. When both samples have zero reads, it is clear that nothing is differentially expressed. Problems are encountered when analysing the data where the gene has zero reads for one condition and a value for the other. It is possible that this represents an interesting biological phenomenon, where a gene in one condition is completely unexpressed according to sequencing. It could be that this phenomenon is due to technical error which prevented the gene from being detected. By completing a repeat of the experiment it would allow further analysis to be undertaken to see if these genes continued to have 0 counts. In that case, these genes may be extremely interesting as a large change is occurring. However, as it is difficult to determine the meaning of the 0 counts and the statistical test fail with data including zero in one condition, all genes including

zero counts were removed. Bullard *et al.,* (2010) observed that some statistical analysis methods were failing to detect many cases of differential expression, such as, genes with large differences in expression between the two conditions. Before any further statistical analysis was completed, any genes with a count over 50 were recorded as being of interest due to the lower count genes being difficult to determine significance.

As this first analysis did not include statistical validation, a further statistical analysis was undertaken. Due to the lack of technical runs choosing a statistical test was more difficult. Using the control count data for the four time points as replicates EdgeR was able to perform a statistical analysis. Each treated sample set was compared to the four controls and a list of gene with a p value $< 0.01$ was compiled. Again, a large amount of genes were identified as being modulated. Over 4000 genes were identified as being differential expressed in total, with 696 being seen in two or more time points. There was also a large proportion of each gene list that was duplicated in the time points either side of the original data set. Altogether, these results support that the changes being seen are true changes and not due to technical variations.

However, one thing that is not taken into account is the length of each gene. Bullard *et al.,* (2010) determined that a gene length-related bias can occur in RNA-Seq differential expression results due to longer genes tending to have more significant differential expression statistics. The resulting lists compiled could favour long genes with small underlying effects as compared to short genes with large effects, which means that the results here may not be the complete picture.

### 3.4.2 Identification of possible pathways involved

KEGG mapper is a very useful online software programme which allows a list of genes to be uploaded and mapped onto different pathways. All genes not identified in any of the pathways provided are highlighted and any nodes of interest can be colour coded. In this experiment KEGG identified numerous pathways, ranging from one to over 60 genes of interest present within them. Those with only a few genes of interest present within them were avoided due to the likelihood of these pathways being regulated by one gene is extremely small. The pathways identified are involved in many different tissues which means some are not as relevant, however, they are still interesting as these genes have been identified so similar pathways may be occurring. Two of these ambiguous pathways are involved in Alzheimer's and Huntington disease. Although not relevant in this study, it would be interesting to discover whether these β-agonists can cross the blood brain barrier and if the genes modulated in this experiment are the same in different tissues.

One of the pathways with a considerable amount of genes of interest present was the mitogen-activated protein kinase (MAPK) pathway. It is thought that G-protein-coupled receptors may activate the MAPK cascade (Malarkey *et al.*, 1995) which supports the finding that β-agonists, a G-protein-coupled receptor agonist, encourage growth via the MAPK signalling pathways. Markou *et al.* (2008) investigated an $\alpha_1$-adrenergic agonists called phenylephrine, which has a similar mode of action to β-agonists. This agonist, as well as similar agents, has been seen to initiate a programme of changes that suggest a hypertrophic phenotype (Chien *et al.*, 1991).In Markou's study, the hypertrophy of cardiac muscle was related to signalling mechanisms activating c-jun. It was also identified that ERK1/2 and JNKs are the

principal kinases responsible for phosphorylation of c-Jun. Although c-Jun is not identified in this experiment, other genes in the MAPK pathway which can regulate the expression of c-Jun were identified as being modulated. Before statistical analysis JNK was seen to be up-regulated in the treated samples at 2 hr. However, in Markou's study they only analysed the effects between 0-60 min which means that other proteins in the MAPK pathway may have been identified if early time points had been taken in this study. One gene which is identified both before and after statistical analysis as being significantly changing is P38 MAPK. The activation of p38 MAPK is a major event during myogenic differentiation in myoblasts (Cabane *et al.*, 2003; de Angelis *et al.*, 2005; Puri *et al.*, 2000; Wu *et al.*, 2000). Although the cells used were fully differentiated, salbutamol may work by encouraging the cells to further fuse and elongate. Bogoyevitch *et al.*, (1996) showed that simulation of the adrenergic receptor activates the MAPK cascade in cardiac myocytes via at least two distinct signalling pathways. Vaniotis *et al.*, (2011) studied the involvement of β-adrenergic receptors, which are activated by β-agonists, on gene expression in the heart. They discovered that the MAPK signalling pathways, although not directly activated by the β-adrenergic receptors in the nucleus, modulate the ability of the receptors to mediate transcriptional events. Overall, there is much evidence to suggest a role of MAPK in the regulation of gene expression as well as its ability to phosphorylate a number of transcription factors (Seth *et al.*, 1992). NFκB is also identified in the pathway analysis as being modulated within the MAPK pathways. Previous work by Vaniotis *et al.*, (2011) suports this as a target for β-agonists as it was seen to be down-regulated in response to treatment with isoproterenol. In this study, *Nfkb1* was identified as being down regulated at 96 hr whereas *Nfkb2* was down-regulated at 24 hr.

Although there is a lot of supporting evidence that this signalling pathway plays a role in the effects of salbutamol, most of the genes identified as statistically significant were from the 24 hr data set. This suggests that this is later process rather than the initial response. However, when considering the raw data counts it can be seen that at 2 hr there are several genes from the MAPK pathways that have a two-fold change. There is also a lack of a pattern across the pathway over the time points, possibly suggesting this pathway is constantly being modulated with the continuous addiction of the β-agonist but with greater changes happening at 24 hr. As well as the evidence that the MAPK pathway is being modulated there were also several pathways highlighted by KEGG that can lead on to the MAPK pathway, such as parts of the calcium signalling pathway, WNT signalling pathway and the Jak-STAT pathway.

A pathway which was not displayed in the previous section but has been identified with KEGG was the ubiquitin mediated proteolysis. Protein ubiquitination plays an important role in eukaryotic cellular processes. It mainly functions as a signal for 26S proteasome dependent protein degradation. During this degradation process ubiquitin is added to those proteins being degraded, which is performed by a reaction cascade consisting of three enzymes; named E1 (ubiquitin activating enzyme), E2 (ubiquitin conjugating enzyme), and E3 (ubiquitin ligase). Members of the ankyrin repeat and SOCS-box (Asb) containing gene family have been shown to act as an E3 ligase to target specific proteins for degradation through the ubiquitin-proteasome degradation pathway (Chung *et al.*, 2005; Wilcox *et al.*, 2004). McDaneld *et al.*, (2006) identified that the over-expression of *Asb15* could stimulate muscle hypertrophy *in vivo*, and that its over-expression caused a delay in differentiation and

an increase in protein synthesis within C2C12 cells. Spurlock *et al.*, (2006) then provided further support for the hypothesis that Asb family members participate in the regulation of intracellular signalling by showing that multiple members of the Asb family are coordinately regulated at the mRNA level at 10 and 24 hr. In their experiment they derived that the Asb family members act via β-adrenergic receptors. Results obtaining in this Chapter show that although *Asb15* was not seen to be regulated, *Asb1* was down-regulated at two time points. A total of 21 genes were statistically differentially expressed within the ubiquitin mediated proteolysis pathways supporting previous findings that this process plays a part in muscle hypertrophy.

It has been reported that the JAK/STAT pathway is activated by hypertrophic agonists such as cardiotraphin-1 and angiotensin II (Kunisada *et al.*, 1996; Pennica *et al.*, 1996). As β-agonists seem to have a hypertrophic effect within muscle cells it is not surprising that this pathway was identified as having modulated genes. As discussed early, β-agonists have an effect on myogenesis. Pearen *et al.*, (2009) identified several genes (*Stat3, Idb1, Smad1, Smad3, Hk2, Pdk4, Sorbs1, Pgc1α, Lipin1α, FoxO1, Ucp3, Nfil3, Dbp, Nurr1, Crem, and Cebpb*) at 1 and 4 hr after treatment associated with regulation of myogenesis but in this study only one, *FoxO1,* was identified as being differential expressed at 96 hr. This could suggest that once differentiation is complete the β-agonists no longer encourage myogenesis but increase growth in other ways. There were also many pathways involved in cancer and apoptosis suggesting the cells were under stress. However there has been some research to indicate that β-agonists can cause apoptosis and cell cycle arrest in cancer cells (Liao *et al.*, 2010). There have been several interesting pathways and

genes identified that need further work investigating their involvement to fully understand the mechanisms of salbutamol.

### 3.4.3 Overall conclusions

In summary, the addition of the β-agonist salbutamol had an effect on the mouse muscle cells within the transcriptome. Several hundred different genes have been identified as being differentially expressed between control and treated samples at four different time points. It appears that the same pathways were involved at all time points but usually different genes. The pathways that seem to have had the most genes modulated are the MAPK pathway and others linked with this pathway. However, several other pathways not related to the MAPK pathway were also identified with a high number of genes modulated. This either suggests that the effect of the addition of the β-agonist effects numbers pathways, each involved in allowing the cells to change and grow or that this transcriptomic experiment has limitations, such as the read counts do not fully relate to the abundance of the gene. To further investigate these results a repeat of this experiment would be necessary, including more time points to gain a greater insight. This would allow for the statistical analysis to be performed with several actual repeats rather than using the controls. Further work would also include an in-depth analysis of the relationship between gene length and read counts. A lot of the previous work performed took samples at 24 hr or later. As it can be seen from the data collected in this study there is a dramatic increase in genes modulated between 2 hr and 6 hr which suggest there are some interesting changes occurring during this time. To explore this future work may also include more time points before 24 hr and after 96 hr to see if the high number of modulated genes is still observed.

It is still unclear how well genomic data links with that of the proteome. To further investigate the mechanisms of salbutamol experiments need to be undertaken to look at the proteomic effect. This will hopefully give a better understanding on the correlation between genes and proteins, as well as helping to identify important targets for the increase in hypertrophy caused by β-agonists.

# 4 Chapter 4

# Investigation of the effect of salbutamol on a mouse muscle cell line using a SILAC approach

## 4.1 Introduction

Stable Isotopic Labelling of Amino acids in cell Culture (SILAC) is a method that allows quantification of proteins and post-translational modifications, as well as being applicable to the measurement of protein turnover (Amanchy *et al.*, 2005; Doherty *et al.*, 2009; Ong *et al.*, 2002). Labelling occurs by growing the cells in a medium deficient of certain essential amino acids that cells cannot synthesis themselves and only gain from medium. Isotopically labelled amino acids can be synthesized separately and added to the medium. As these labelled anologs are supplied instead of the natural abundant amino acids, they will be incorporated into each newly synthesised protein chain. As there is no chemical difference between the labelled amino acids and the natural ones, the cells should behave identical to a control cell population (Ong *et al.*, 2002). The most commonly used labelled amino acids are arginine and lysine as this allows every peptide to be labelled when digested with trypsin. Labelling with these amino acids can be achieved using isotopes containing heavy carbon ($^{13}C$), nitrogen ($^{15}N$) or hydrogen ($^{2}H$). Using $^{13}C$ and $^{15}N$ can simplify data analysis because the labelled and unlabelled pairs do not separate as frequently during LC-MS compared to deuterium (Ong & Mann, 2006). For this experiment arginine and lysine will be used with both carbon and nitrogen labelling to enhance the quality of labelling.

Due to there being no difference between the cell populations, half of the cells can be treated with a stimulus, such as a β-adrenergic agonist used in this study, allowing a comparison between treated and control populations. By labelling only the control or only the treated cells these populations can then be harvested, and due

to the label being incorporated directly into the amino acid sequence of every protein, the extracts from both conditions can be mixed. Purified peptides will maintain the exact ratio of the labelled to unlabelled protein, as protein synthesise is no longer occurring, allowing for quantification by comparison of these ratios..

Figure 4.1 shows an outline of a SILAC method. Once the labelling has occurred and samples collected from cell culture, the next task is to prepare the protein and peptides for mass spectrometry analysis. The sample preparation is an extremely important part of the SILAC methodology. Sample preparation will often include separation of the sample to simplify the mixture before mass spectrometry analysis. This can be achieved by either separating the proteins themselves before digestion, using for example; 1D gel electrophoresis or separating the peptides using techniques such as isoelectic focusing. However, protein separation may introduce sample losses through either additional handling steps or through the limitations of the separation technologies, for example; 1D gels show favouritism for abundant proteins. There are also many arguments for separating peptides due to the lack of clean up steps which occur when separating proteins due to detergents being utilised for solubilisation. These detergents can damage the mass spectrometer and overwhelm the spectra resulting in complications in the analysis so must be removed. By investigating separation of proteins and peptides as well as the effects of no separation, a greater understanding of the effects these procedures have on proteomics would be gained.

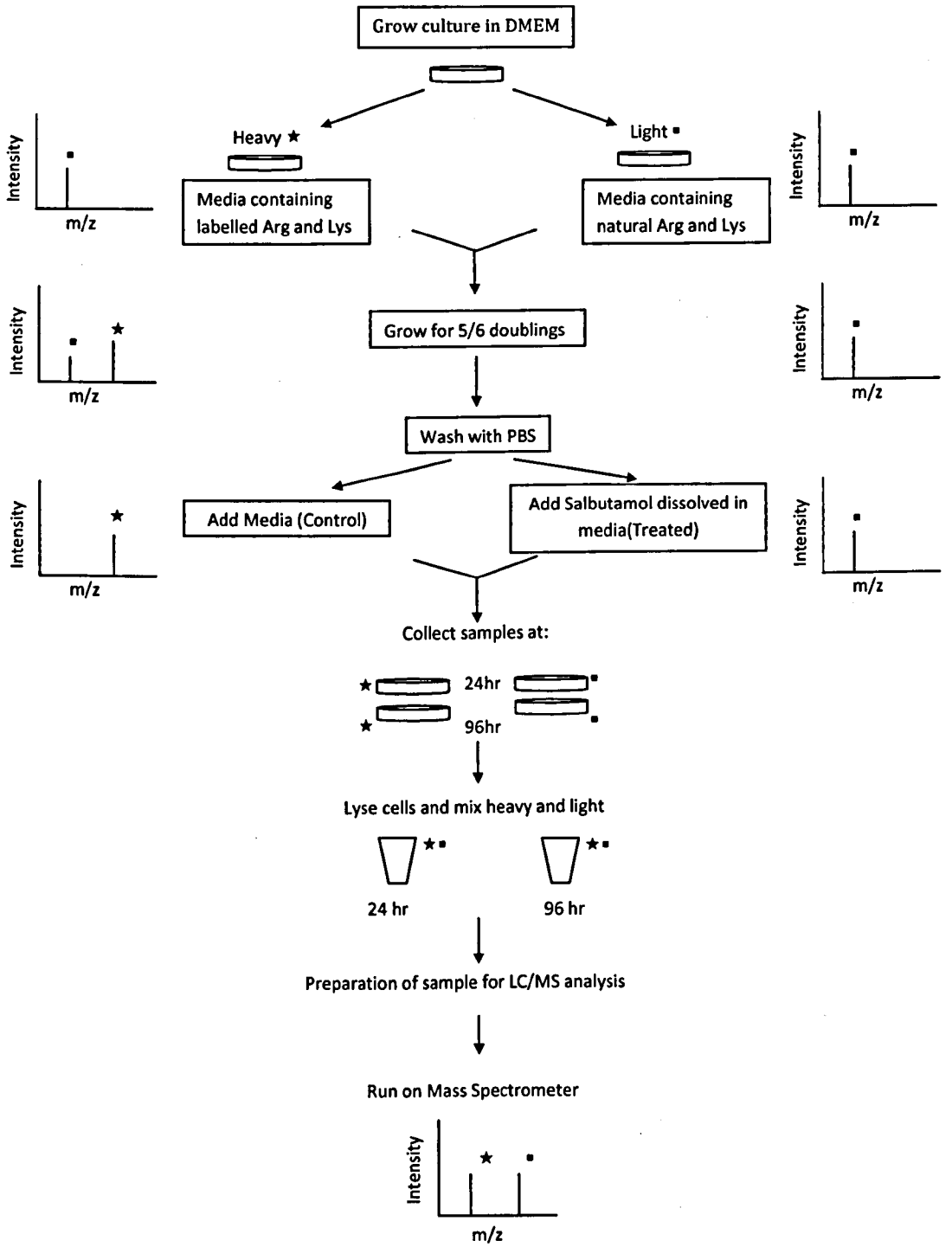



**Figure 4.1 Overview of SILAC workflow**

During the first half of the SILAC method the cells are grown in heavy and light SILAC media until the heavy cells have fully incorporated the heavy amino acids (★). The spectra show how the labelling can be seen throughout the growth of the cells. This allows the cells to be pooled and when MS data has been acquired, to distinguish between the two (★and ■ indicate the heavy and light peptides, respectively). Once the heavy amino acids are fully incorporated, the cells can be differentially treated to identify changes in the proteome. The cells can be collected at different time points to gain information of changes in the proteome over time. A variety of different sample methods can be used before mass spectrometry analysis.

The Agilent 3100 OFFGEL fractionator employs a fairly new technology which has been developed to allow separation of both proteins and peptides using the same principles as IEF separation but avoiding the disadvantages of SDS gels by keeping samples in solution. Isoelectric focusing is performed on proteins or peptides in an immobilized pH gradient (IPG) gel strips. By applying a high voltage to the ends of the gel strip the protein or peptide molecules migrate through the gel until they reach the position where the pH equals the pI of that molecule and sit in the liquid of the corresponding well.

To enable protein quantification of the label and unlabelled samples an LTQ Orbitrap Velos mass spectrometer (Thermo Scientific) was used. This machine combines the separation techniques of the LTQ with orbitrap technology (Makarov, 2000). The LTQ is capable of detecting MS spectra at very high sensitivity but low resolution and mass accuracy. The Velos comprises of two identical linear quadrupole cells separated by a centre lens. Ions accumulated in the LTQ are then transferred into the C-trap, the RF-only quadrupole, which accumulates and stores the ions. The C-trap refers to the shape of the RF section. In the C-trap, ions are collisionally damped by a low pressure of nitrogen, and they come to rest in the middle of the trap. The storage quadrupole is required to couple the continuous electrospray ion source with the orbitrap, which operates using high voltage electric pulses. Once injected into the orbitrap, these ion packets start to circle around the inner electrode and spread into rings that oscillate along the electrode without the need for any additional excitation. All ions have exactly the same amplitude, but those that have different mass/charge ratios will oscillate at their respective

frequencies. Data from the orbitrap is collection by customized control and acquisition software.

The next stage, once samples are prepared and run on a mass spectrometer, is analysing the data. MaxQuant (http://maxquant.org/) is a software package that can be used for the analysis of the SILAC spectra. It is able to produce peak lists, quantitation, false positive results determined from search engine results, peptide to protein group assembly and data filtration and presentation (Graumann *et al.*, 2008). The MaxQuant workflow is compatible with SILAC experiments involving prefractionation of proteins, as well as double or triple labelling with heavy amino acids. In addition to these, the MaxQuant software can handle a variety of SILAC comparisons including replicate measurements, time courses with common reference points, label switches and multi-sample comparisons. Briefly, the pipeline for MaxQuant involves acquiring the raw files from the mass spectrometer and loading them into the 'Quant' module. In this module, advanced three-dimensional peak and isotope pattern detection is performed as well as assembling the SILAC pairs. Andromeda is used to identify the peptides. The 'Identify' module uses the search engine results and the raw files to perform statistical validation, as well as converting peptides into proteins and quantifying the proteins. This data is displayed in several tab-separated text files. These can then be used for further downstream bioinformatic analysis (Cox *et al.*, 2009).

### 4.1.1 Proteomic analysis of the growth of skeletal muscle using C2C12 cells and labelling techniques

It is known that muscle is formed from small cells called myoblasts which fuse together under myogenesis to produce myotubes. These fibers contain multiple nuclei as they are formed from many myoblasts. Once fusion is complete, skeletal muscle loses the ability to undergo mitosis and any further growth is through hypertrophy. C2C12 is a subclone obtained by Blau *et al.,* (1983) from a myoblast cell line derived from regenerating adult mouse skeletal muscle (Yaffe & Saxel, 1977). C2C12 cells are used as a model system to study different types of muscular diseases and functions often with the focus on differentiating muscle cells (Andrés & Walsh, 1996; Pieper *et al.*, 1987). When deprived of serum, C2C12 myoblasts cells undergo cell cycle arrest and fuse together to form multi-nucleated myotubes. As myotubes serve as the building blocks for skeletal muscle, understanding both the differentiation process of skeletal myoblasts and the growth of myotubes are equally important. Skeletal muscle growth and differentiation is a complicated process which is continuing to be understood.

Although the genomics-based analyses of muscle cell growth have been highly informative, the molecular mechanisms by which proliferating myoblasts become fused multinucleated myotubes, as well as further muscle growth initiated from exercise or steroids, cannot be fully understood from transcriptomic data alone. This is because genes can encode for several proteins which can then go through different types of modification which means gene expression is just the beginning of the story. Therefore, it is desirable to complement the differential gene expression identified in the transcriptomic analysis with studies examining the proteomes of differentiated C2C12 cells undergoing treatment with salbutamol.

Once myogenesis has completed there is a second wave of proliferation and differentiation when muscle fibres are formed. The molecular mechanisms regulating these later stages of muscle development are still not fully understood. It is known that certain signalling pathways and transcriptional factors are implicated in controlling slow versus fast fibre phenotypes. As this study will be utilising a cell line that progresses through myogenesis to give myotubes before being treated it is possible that the involvement of some of these pathways may be highlighted in this experiment. There are many previous studies looking at the proteome of muscles undergoing hypertrophy, for example, Olson & Williams (2000) have shown that a calcineurin-activated $Ca^{2+}$-dependent pathway acting through the nuclear factor of activated T cells (NFAT) transcription factor plays an important role in controlling gene expression in slow fibres. Another study has found that Ras signalling through a MAPK/ERK pathway is another important mediator of slow muscle specific gene activation (Murgia *et al.*, 2000). This pathway and those adjoining were identified as the most highly modulated in the transcriptomic analysis carried out in this study (Chapter 3).

Most work using C2C12 cells investigate any modulation within differentiating muscle cells. As this experiment plans to use fully differentiated cells it is difficult from the literature to hypothesise what proteins may be identified. However, Tannu *et al.*, (2004) identified that myotubes accumulated muscle-specific actins and proteins which associate with actin. Tropomyosin, binds to actin and along with troponin regulates calcium-dependent contraction. The pathway involved in regulating actin was identified in Chapter 3 so proteins that interact with actin may

be identified in this study. As well as tropomyosin, serine/threonine protein kinase Akt2/PKB also accumulated in C2C12 myotubes (Tannu *et al.*, 2004).

There is limited research into muscle growth using SILAC. However, Cui *et al.*, (2009) used SILAC to identify differentially expressed proteins between rat myoblasts and myotubes. They were able to identify around 400 proteins differentially expressed with two or more peptides identified for each protein. Although this study focused on the difference in differentiating muscles, several proteins identified are in common with proteins and genes identified in Chapters 2 and 3, for example: eukaryotic translation elongation factor, actin and pyruvate dehydrogenase. It is most likely that proteins involved in differentiation will be identified when the cell goes through hypertrophy as there will again be cytoskeletal growth and the use of energy. Toigo *et al.*, (2005) used a different labelling method, ICAT, to look at a subproteome of atrophying mouse skeletal muscle. Using this technology they were able to identify 62 altered proteins with association to atrophy. In the cytosolic subproteome there seems to be proteins involved in alcohol catabolism, carbohydrate metabolism and energy pathways among others. Another similar labelling technique is iTRAQ, which was discussed in Chapter 1.The iTRAQ technique has been a useful for examining the proteome in muscles. Hakimov *et al.*, (2009) used this form of labelling to confidently catalogue 542 proteins in porcine muscle. When investigating the effects of propranolol, a chemical which also acts through the β-adrenoceptors, Sui *et al.*, (2008) were able to identify the secretion of T-kininogen in treated smooth vascular muscle cells by employing iTRAQ. Warren *et al.*, (2010) combined iTRAQ with OFFGEL fractionation to delve into the cardiac proteome. Although this study is unable to help with hypothesising what proteins

may be seen in this experiment it is a very good example about how prefractionation, when used in conjunction with labelling method, can be of great use in investigating the proteome of many cells. When Warren *et al.*, (2010) compared data gained from samples analysed with no fractionation prior to LC-MS/MS with those fractionated via the OFFGEL the identifications of proteins improved by approximately four fold.

### 4.1.2 Aims

The aim of this chapter is to use SILAC to understand the proteomic effect of salbutamol on an immortalised mouse cell line. The cell line C2C12, are mouse myotubes that can differentiate into fibroblasts once confluent. These cells are a good model for muscle tissue. The first aim was to fully label the mouse cells without effecting their growth. The second was to perform method optimisation to allow for efficient sample collection and preparation. Lastly, to determine the most suitable software package to be used for data analysis. This experiment should give additional information on the data already gained from the transcriptomic study. The results, compared with the findings made in chapter 3, would give further knowledge to the workings of this anabolic agent as well as allowing further analysis on the correlation between transciptomics and proteomics.

# *4.2 Materials and Methods*

## 4.2.1 Cell Culture

### 4.2.1.1 Growing C2C12 cells

The C2C12 cell line was used and grown as myoblasts in 25 $cm^2$ bottom (T25) vented cell culture flasks (BD Falcon$^{TM}$) and incubated at 37°C in a 5% $CO_2$ humidified incubator. As previously described in chapter 3, the cells were initially grown in filter sterilised DMEM medium (Sigma-Aldrich) supplemented with 10% (v/v) FCS, 1% (v/v) L-glutamine (Sigma-Aldrich) and 1% (v/v) penicillin-streptomycin (Sigma-Aldrich). Once the culture was established in this media several flasks were split into heavy or light labelled media to be used for the SILAC experiment. The original cells were kept growing to harvest for preliminary experiments. The separate heavy and light labelled medium was made by first making a stock solution of media (DMEM low glucose minus arginine and lysine media (Sigma-Aldrich) including: 0.0159g/L phenol red, 0.0001g/L proline, 0.105g/L L-leucine, 0.11g/L sodium pyruvate, 4.5g/L D-glucose, 1 % (v/v) penicillin-streptomyocin, 10% (v/v) fetal calf serum) without the arginine or lysine. At this point the medium was split and supplemented with either light or heavy labelled arginine or lysine at 0.08% (v/v) and 0.15% (v/v) respectively from a stock solution of 0.1g/ml, made in DMEM low glucose media.

### 4.2.1.2 Splitting Cells

As stated in chapter 3, cells need to be split before they become confluent. To split a 25 $cm^3$ flask of cells, the old medium was removed, the flask washed with 5 ml of

HEPES Buffer (Sigma-Aldrich), buffer disposed of and 5 ml of trypsin (Sigma-Aldrich) added before incubating for 2-7 min. Cells were then extracted and placed in a universal tube. Next, 10 µl of cells were used on a haemocytometer for counting. The remaining cells were spun at 1500 g for 5 min and the trypsin then carefully discarded. Fresh medium in the volume of 5 ml was used to resuspend the cells. Another 5 ml of fresh medium was added to new flasks before the correct volume of cells were added to equal a concentration of $1 \times 10^4$ cells per flask.

#### 4.2.1.3 Labelling of cells

Cells were split from the original media into light media and grown for one passage. After a week the cells were ready to be split again, this was undertaken using the method described above. However, when splitting cells for the labelling experiment, the cells were resuspended in DMEM stock medium before being transferred into the correct flasks. This allowed the same cells to be used for both the heavy and light experiments. There were four biological replicates for the SILAC experiment and each biological replicate had two time points. Figure 4.1 describes how the flasks were labelled.

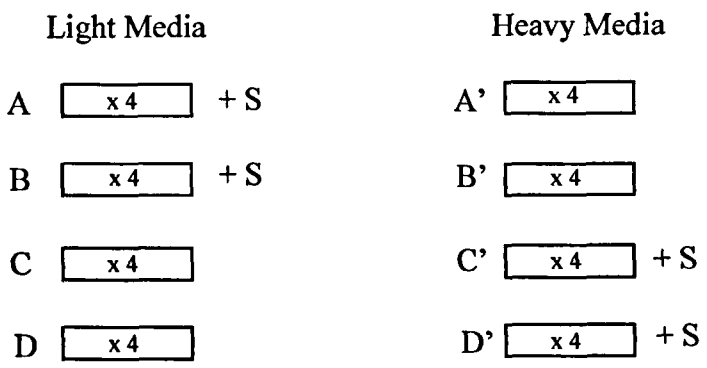



**Figure 4.2** ***Schematic showing flask labelling and treatment.***
A-D represents the four biological replicates containing either light amino acids arginine and lysine or the heavy variations of these amino acids. The image depicts the need for four flasks per biological replicate; +S symbolises which flasks were treated with salbutamol. Two biological replicates from each media were treated with the salbutamol to normalise any variation that may occur due to the labelling.

All flasks were left for six doublings, at which point, all medium was removed and either a heavy or light low serum media (DMEM low glucose minus arginine and lysine media (Sigma-Aldrich) including: 0.0159g/L phenol red, 0.0001g/L proline, 0.105g/L L-leucine, 0.084g/L arginine, 0.146g/L lysine, 0.11g/L sodium pyruvate, 4.5g/L D-glucose, 1% (v/v) penicillin-streptomyocin, 2% (v/v) horse serum) was added to the correct flasks. Cells were then left for another six doublings to allow them to become fully differentiated. Medium was changed every two days to prevent any effect on the cells growth due to limited nutrients or salbutamol. Treatment was staggered over 15 min intervals to allow for collection at a later stage (4.2.1.4). Fresh heavy or light medium was prepared and either salbutamol at a concentration of $1 \times 10^{-4}$, made in DMEM low glucose medium, or the same volume of just DMEM low glucose medium was added to the correct flasks. Fresh control or treated medium was added each day.

#### 4.2.1.4 Harvesting cells

After 24 hr and 96 hr of treatment, cells were harvested from all four biological replicates. Harvesting cells from a 25 $cm^3$ flask of cells involves first removing the old medium and washing with 5 ml of HEPES buffer (Sigma-Aldrich). The buffer was then disposed of and 5 ml of trypsin (Sigma-Aldrich) was added and incubated for 4-7 min to cleave the cells. Cells were then extracted and placed in a universal tube and spun at 1500 g for 5 min before carefully discarding the trypsin. To wash the cell pellet 10 ml of PBS was added and the pellet resuspended. This was then spun for another 5 min at 1500 g at 4°C. The PBS could then be removed and a fresh 1 ml added to resolubilise and transfer the cells to an ependorf tube. This time the cells were spun at 16000 g for 4 min. The PBS was removed and the pellet was stored at -20°C.

### 4.2.2 Sample preparation for protein separation

#### *4.2.2.1* Lysis of cells

An aliquot of 500 µl of lysis buffer (8M urea, 2M thiourea, 2% (w/v) CHAPS, 1% (w/v) DTT) was added to each sample pellet. Samples were treated to a freeze/thaw process five times, which involved; freezing with liquid nitrogen followed by quickly thawing to room temperature and vortexing for 1 min. A protein assay was undertaken on the lysed sample before precipitation. Samples were stored at - 20 °C.

#### 4.2.2.2 BSA assay

The Thermo Scientific Pierce Coomassie Plus Protein Assay was used to measure the protein concentration of the samples. A set of protein standards, ranging from 2.5 – 25 µg/ml BSA, were prepared in ultrapure water. An aliquot of 150 µl of each standard and the samples were transferred to different wells on a 96 well plate. An equal volume of Coomassie Plus reagent was added to all the standards and samples before being left to incubate at room temperature for 10 min. To determine the concentration of the samples a Multiskan Ascent microplate photometer (Thermo Scientific) was used to read the absorbance at 595 nm.

#### 4.2.2.3 Acetone precipitation for protein separation

An acetone precipitation was performed to remove any salts or detergent left by the lysis buffer. Five volumes of ice cold acetone was added to the sample, mixed and left at -20 °C overnight. The sample was then spun at 16000 g for 20 min at 4 °C and the supernatant removed. A further 200 µl of ice cold acetone was used to wash the sample before being centrifuged at 16000 g for 20 min at 4 °C. The supernatant was again removed and the pellet stored at –20 °C.

#### 4.2.2.4 TCA precipitation

An alternative to the acetone precipitation is to use trichloroacetic acid (TCA), which performs the same function as the acetone precipitation but can sometimes be more effective. Six volumes of 100% (v/v) TCA were combined with four volumes of sample. This was left for 2 hr at 4°C before being centrifuged at 16000 g for 5 min at 4°C. The supernatant was removed leaving the protein pellet intact and then

washed with 200 µl of ice cold acetone. The samples were centrifuged again at 16000 g for 5 min at 4°C. After the acetone washes were repeated twice more, pellets were dried in a heat block at 95°C for 5-10 min to evaporate any remaining acetone.

## 4.2.3 Sample preparation for peptide separation

### 4.2.3.1 Lysis of cells

Two different lysis buffers were used for preliminary experiments to identify which provided the best optimisation for protein recovery and clean up.

#### *4.2.3.1.1 Lysis buffer 1*

An aliquot of 300 µl of lysis buffer 1 (8M urea, 2M thiourea, 0.6% (w/v) CHAPS, 1% (w/v) DTT) was added to the sample pellet, mixed thoroughly and subjected to a freeze/thaw process five times. A protein assay was undertaken on the lysed sample before precipitation. The samples were kept at –20 °C.

#### *4.2.3.1.2 Lysis buffer 2*

The pellet was resolubilised in 1 ml of lysis buffer 2 (8M urea, 2M thiourea, 0.6% (w/v) CHAPS in 100mM HEPES buffer pH 8.5) and vortexed thoroughly before a BSA protein assay was performed ahead of precipitation. No freeze/thaw step was used here to help prevent any detergent dropping out of solution. Samples were stored at -20°C.

#### 4.2.3.2 BSA assay and mixing of heavy and light samples

The BSA assay was undertaken following the same method described in section 4.2.2.2. Once concentrations of both the heavy and light samples had been determined an equal amount of protein for the differently labelled samples, for each time point, were mixed together. This was repeated for all biological replicates. The total amount of protein combined equalled 500 µg.

#### 4.2.3.3 Acetone precipitation for peptide separation

Before precipitation the samples were diluted in double distilled water to reduce the concentration of urea from 8M down to 2M. Five volumes of ice cold acetone were mixed with each sample, vortexed and left over night at -20°C. Centrifuging the samples at 16000 g, 4°C for 20 min allowed the supernatant to be carefully removed. Consequentially samples were washed with 500 µl of ice cold acetone, spun at 16000 g for 10 min at 4°C and finally the supernatant decanted. This was repeated for a second wash before leaving to air dry to remove any remaining acetone.

#### 4.2.3.4 In solution trypsin digestion

Each sample was resuspended in 180 µl of 25mM ammonium bicarbonate after precipitation. An aliquot of 10 µl was removed to run on a 1D gel as a control (section 4.2.3.5). The addition of 10 µl of a 60 mM dithiothreitol (DTT) solution occurred once the aliquot was taken. This was vortexed and left to incubate for 10 min at 60 °C before being allowed to return to room temperature and then briefly spun. After reduction, an alkylation step was performed by adding 10 µl of a 180 mM solution of idoacetamide (IAA) and vortexing. Incubation followed at room

temperature in the dark for 30 min. Sequence grade trypsin was diluted in 25 mM ammonium bicarbonate to give a concentration of 1 µg/µl before being mixed with the sample in a 50:1 protein:trypsin ratio. This was then incubated overnight at 37 °C.

#### 4.2.3.5 SDS- PAGE 1D gel electrophoresis

A 10µl aliquot of the digested sample and the 10µl undigested aliquot were run on a 1D electrophoresis gel to check for complete digestion. Samples were run on a 16 cm wide and 1 mm thick SDS-PAGE gel using Mini-PROTEAN$^{TM}$ Tetra cell kit (Bio-Rad). A 12% (v/v) SDS-PAGE gel was cast following the recipes in Table 2.1 Chapter 2 for the resolving gel and Table 4.1 for the stacking gel.

**Table 4.1** ***5 % (v/v) Stacking Gel***

| Reagent | Volume | Additional information |
|---|---|---|
| 30 % (v/v) acrylamide | 10 ml | |
| 1.5 M Tris-Cl pH 8.8 | 5 ml | |
| ultrapure $H_20$ | 4.6 ml | |
| 10% (w/v) SDS | 0.2 ml | |
| 10% (w/v) ammonium persulphate (APS) | 0.2 ml | Add fresh |
| TEMED | 0.01 ml | Add fresh |

Running buffer was prepared at 1 x concentration (2.5mM tris-base, 19.2mM glycine, 0.02% (w/v) SDS) before use. The aliquots of both samples were mixed with 20 µl of sample buffer, Laemmli 2x concentrate (Sigma Aldrich) before being heated at 100°C for 5 min. The protein was separated by running the gel at 200 V for

1 hr or until the dye had reached the end of the gel. The gel was then removed from the plates and fixed for 2 hr in 40 % (v/v) ethanol, 10% (v/v) acetic acid. Once fixed the gel was washed twice for 10 min in ultrapure $H_20$ and stained in colloidal Coomassie blue (1 part methanol to 4 parts colloidal stock (Chapter 2 Table 2.2)) for 1-7 days.

## 4.2.4 Separation using an Off-Gel Fractionator

### 4.2.4.1 Protein Separation

The protein pellet formed after acetone precipitation was resuspended in 1.44 ml of the protein OFFGEL stock solution (7M urea, 2M thiourea, 0.07M DTT, 10% (v/v) glycerol and 1% (v/v) 3-10 pH ampholytes) and 360 µl of dd-$H_2O$. The OFFGEL trays were assembled following the manufacturers guidelines. Proteins were separated using 12cm 3-10 pH IPG strips (GE Healthcare). Each strip was rehydrated for 15 min with 40 µl rehydration buffer (80% Protein OFFGEL stock solution) in each well, before 150 µl of sample was loaded into each well. Proteins were either separated by the standard protocol; 20 KVhr at 4500 V, 50 µA and 50 mW or a by a programmed method of 500 – 1000 V over 4 hrs then up to 8000 V for 50 kVhr.

#### *4.2.4.1.1 Collection of samples*

##### *4.2.4.1.1.1 Acetone precipitation for fractionation*

Precipitation was undertaken using the method described in section 4.2.3.3. but using 1ml to wash the samples and only completing one overnight precipitation to minimise the protein loses.

##### *4.2.4.1.1.2 In solution digestion*

Digestion was performed as described in section 4.2.3.6. Again aliquots of the samples before and after digestion were run on a 1D gel electrophoresis to check for complete digestion.

##### *4.2.4.1.1.3 Clean up of samples using C18 stage tips*

A second clean up step was performed to ensure the samples were suitable to be run on a mass spectrometer. C18 stage tips (Proxeon) with a capacity to bind 10µg of protein in a max volume of 200 µl, were used for this process. To start the tips were spun at 300 g for 1 min before following with an initializing step, this involved passing 20 µl of 80% (v/v) ACN, 5% (v/v) formic acid through the tips by centrifuging at 300 g for 1 min. To re-equilibrate the tips 20 µl of 5% (v/v) formic acid was added to the tips and spun for 1 min at 300 g. The sample was then transferred to the tips, to which 20µl 5% (v/v) formic acid was added, before being spun at 700 g until all the liquid had passed through. The tips were then washed with another 20 µl of 5% (v/v) formic acid and centrifuged at 300 g for 1 min. To elute

the peptides 20 µl of 80% (v/v) ACN, 5% (v/v) formic acid was passed through the tips twice by centrifuging at 300 g.

#### 4.2.4.2 Peptide Separation

The peptide sample was added to HPLC $H_2O$ to give a total volume of 360 µl. This was then combined with 1.44 ml of the OFFGEL peptide stock solution (0.1% (v/v) glycerol, 1.2% (v/v) 3-10 pH ampholytes). The OFFGEL trays were assembled following the manufactures guidelines. Again, 12 cm pH 3-10 IPG strips from GE Healthcare were used to separate the peptides. The strip was rehydrated for 15 min by applying 40 µl of the rehydration solution (80% (v/v) peptide OFFGEL stock solution) to each well before pipetting 150 µl of the sample into each well. The peptides were separated using the standard protocol: run for 20 kVhr at 4500 V, 200 mW and 50 µA then holding at 500 V, 20µA and 50 mW.

##### *4.2.4.2.1 Clean up of peptide samples*

Each well was collected separately. No acetone precipitation is needed in this separation method but the glycerol still needs to be removed from the samples. To complete the clean up, each well was applied to a C18 stage tip following the protocol set out in section 4.2.4.1.1.3.

### 4.2.5 Sample preparation involving no separation

To enable the comparison of data gained by either the separation of peptides before mass spectrometry analysis via the OFFGEL, to data from a whole cell lysate run on

a mass spectrometer, the samples must be the same. For this sample preparation the samples were taken from the same preparation explained in section 1.2.3. An aliquot was taken from the digest which equalled 10 µg of protein and was cleaned up via C18 stage tips (section 4.2.4.1.1.3) before analysed on the mass spectrometer.

## 4.2.6 Sample preparation using RapiGest

### 4.2.6.1 Lysis of cells

Samples were collected as stated in section 4.2.1.4. A 10 mg vial of RapiGest (Waters MS technologies) was reconstituted in 1ml of 50mM ammonium bicarbonate to give a 1% (w/v) stock solution. A sample of this was taken and diluted down to 0.05% (v/v) in 25mM ammonium bicarbonate before 500µl was added to each sample. Samples were mixed thoroughly before being subjected to five rounds of freeze/thawing with intense votexing between each round. The samples were then heated for 10 min at 80°C before being spun for 10 min at 16000 g. The supernatant was carefully transferred to a new tube and an aliquot was taken for a protein assay (section 4.2.2.2).

### 4.2.6.2 Digestion of samples

Equal amounts of control and treated sample were taken and mixed together to give a total protein amount of 100µg. To this, 25mM ammonium bicarbonate was used to adjust the final volume to 160µl. To this 10µl of a 9.2mg/ml DTT solution was added, votexed and incubated at 60°C for 10 min. After briefly centrifuging the sample, 10µl of a 33mg/ml IAA solution was added. The samples were votexed and

incubated at room temperature in the dark for 30 min. Sequence grade trypsin (Sigma-Aldrich) was diluted in 25 mM ammonium bicarbonate to give a concentration of 1 μg/μl before being mixed with the sample in a 50:1 protein:trypsin ratio. This was then incubated overnight at 37 °C. After running a 1D gel to check for complete digestion the samples were analysed by mass spectrometery.

## 4.2.7 Protein identification

### 4.2.7.1 Mass spectrometry analysis

#### *4.2.7.1.1 LC-MS/MS*

Two types of MS were used for the analysis of the SILAC data, LTQ and Orbitrap. The LTQ, as described in Chapter 2 section 2.2.4.2, was used for two reasons; a) for a preliminary check of the samples, b) as a comparison to the Orbitrap.

#### *4.2.7.1.2 LTQ Orbitrap Velos (Thermo Scientific)*

Samples were run on the Velos by Dr Stuart Armstrong, University of Liverpool. Peptide mixtures were analyzed by on-line nanoflow liquid chromatography using the nanoACQUITY-nLC system (Waters MS Technologies) coupled to an LTQ-Orbitrap Velos (ThermoFisher Scientific) mass spectrometer equipped with the manufacturer's nanospray ion source. The analytical column (nanoACQUITY UPLC™ BEH130 C18 15cm x 75μm, 1.7μm capillary column) was maintained at 35°C at a flow-rate of 300nl/min. The gradient consisted of 3-40% acetonitrile in 0.1% formic acid for 90 min then a ramp of 40-85% acetonitrile in 0.1% formic acid

for 3 min. Full scan MS spectra (m/z range 300-2000) were acquired by the Orbitrap at a resolution of 30,000. Analysis was performed in data dependant mode. The top 20 most intense ions from MS1 scan (full MS) were selected for tandem MS by collision induced dissociation (CID) and all product spectra were acquired in the LTQ ion trap. A total of 500ng per sample was injected on to the column.

#### 4.2.7.2 MaxQuant

Raw files of Orbitrap data were processed using version 1.1.1.36 of MaxQuant (http://www.maxquant.org/). For protein identification the ipi.MOUSE protein database v3.87 (http://www.ebi.ac.uk/IPI/IPImouse.html) was uploaded to MaxQuant. Carbamidomethylation was set as fixed modification and variable modifications as oxidation (M). Initial peptide mass tolerance was set to 20ppm and fragment mass tolerance was set to 0.8 Daltons (Da). Two unique peptides per protein were required for high-confidence protein identifications. The peptide and protein false discovery rates (FDR) were set to 0.01. The maximal posterior error probability (PEP), which is the probability of each peptide to be a false hit considering identification score and peptide length, was set to 1. Raw files were uploaded separately but all data was retrieved in one file to compare the fold changes between experimental runs and time points.

## 4.3 Results

### 4.3.1 Sample pre-fractionation

Before determining if pre-fractionation before mass spectrometry analysis had any benefits, a decision on whether to separate in protein or peptide space needed to be achieved. Test samples of light labelled cells were used for both separation methods. To test the suitability of the focusing method for this experiment, the samples were separated in protein space originally. This method allowed the resulting samples to be inspected on a gel after separation (Figure 4.3). From this figure, separation of the different proteins present in different pI ranges can be witnessed. However, in several fractions the same band could be identified suggesting that focusing could be optimised further or that several of the proteins have isoforms which cause them to appear at different pI values. When attempting to change the focusing options to allow better separation the OFFGEL fractionator had trouble reaching the intended higher voltages. This was most likely due to the high salt content in the samples. When focusing the samples for the DIGE experiment a similar problem was encountered resulting in separation occurring via a low voltage for a longer time which can also cause incomplete separation.

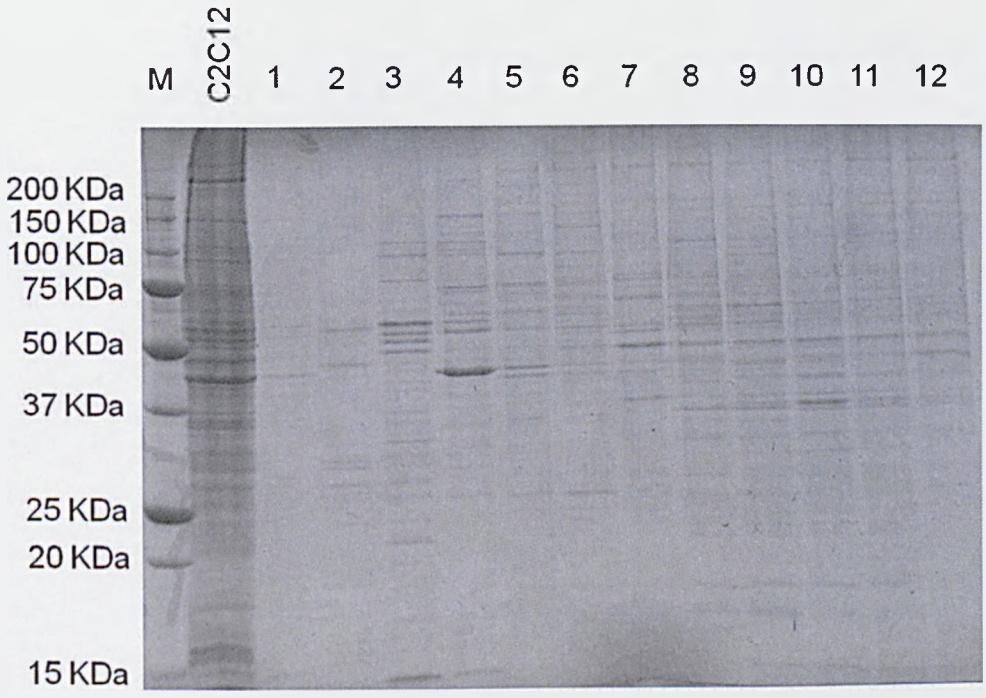



**Figure 4.3** ***SDS-PAGE 1D gel of protein samples collected after separation on an OFFGEL fractionators.*** A 13 cm pH 3-10 IEF strip was used to separate a lysed sample of C2C12 cells into 12 fractions (lanes 1-12). The lane labelled C2C12 contains a sample before fractionation. Gel is stained with commassie plus reagent to allow visualisation of the proteins present in each fraction.

Once the protein samples had been fractionated they needed to be recovered efficiently and cleaned to be compatible with mass spectrometry analysis. Substances, such as CHAPS and glycerol, can cause problems in the analysis by damaging the column or the trap. The samples were originally precipitated using acetone prior to being separated on the OFFGEL. However, this still resulted in a large amount of CHAPS being present (Figure 4.4). To allow for efficient precipitation, each well was washed with distilled water after the sample had been removed. This was added to the original sample to dilute the concentration of urea from 8M to 2M.

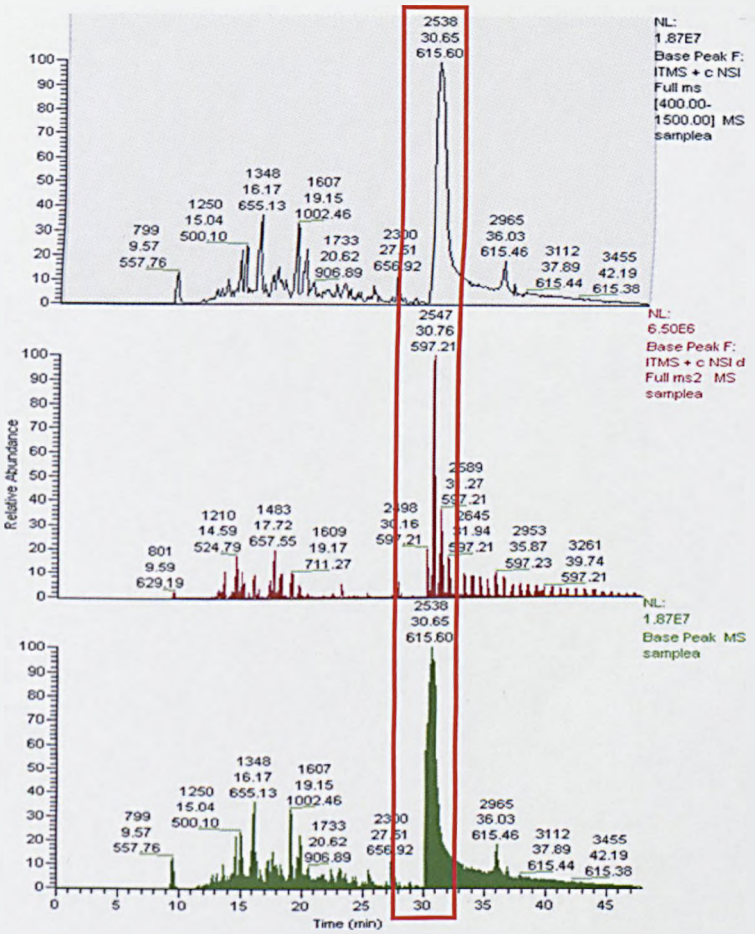



**Figure 4.4** ***Spectra showing contamination in a OFFGEL separated sample of lysed C2C12 cells.*** A sample of C2C12 cells grown in light amino acids were lysed in a buffer containing CHAPS. It was then separated on the OFFGEL fractionator for 20 kVHr at 4500 V. Acetone precipitation was performed on each well after being diluted four fold. The samples were then digested using trypsin and run on an LTQ. The highlighted peak indicates CHAPS (615 m/z).

When repeating the experiment with peptide separation CHAPS contamination was again identified but it was not as significant a problem as in the protein separation. Not only was this an advantage but the time to prepare, separate and clean the samples were significantly reduced when using peptides. Separation of peptides also produced more protein identifications in MASCOT. Due to the reasoning stated,

peptide separation was chosen as the preferred method of separation for the SILAC experiment.

## 4.3.2 Protein modulation

### 4.3.2.1 SILAC 1

At 24 hr the mean identifications recorded was 656 ±52 across the four replicates. However, only an average 338 ± 15 proteins were identified with a heavy to light ratio score by MaxQuant. Those proteins identified with only one unique peptide or with a heavy to light count <2 were discarded before ratios were calculated. At 96 hr the average number of proteins identified was also 656 ± 65, with 347 ± 35 proteins fitting the criteria to produce heavy to light ratios.

**Table 4.2** ***Table listing the number of proteins identified and considered modulated using MaxQuant.*** C2C12 cells were labelled with heavy or light arginine and lysine before being treated with salbutamol. Samples of both the treated and control cells were collected after 24 hr and 96 hr of treatment. These were analysed using an Orbitrap Velos and the spectra run through MaxQuant to identify proteins from the corresponding peptide peaks. MaxQuant only calculated the heavy:light ratio for those proteins that appeared in all four replicates with more than one peptide. The averages from the four replicates are displayed.

| Replicates | Total identifications | | Total proteins identified with heavy:light ratio | | Total number of proteins modulated | | No. up-regulated | |
|---|---|---|---|---|---|---|---|---|
| | 24 hr | 96 hr | 24 hr | 96 hr | 24 hr | 96 hr | 24 hr | 96 hr |
| A | 710 | 749 | 350 | 397 | 7 | 9 | 5 | 5 |
| B | 648 | 632 | 333 | 340 | 6 | 10 | 4 | 4 |
| C | 679 | 644 | 350 | 332 | 6 | 11 | 3 | 6 |
| D | 588 | 598 | 318 | 317 | 4 | 5 | 2 | 3 |
| Average | 656 | 656 | 338 | 347 | 6 | 9 | 4 | 5 |
| Standard Deviation | 52.1 | 65.1 | 15.4 | 35 | 1.3 | 2.6 | 1.3 | 1.3 |

Surprisingly, only an average of 6 proteins per replicate were identified as having a fold change $\geq 1.5$, with a total of 18 non-redundant proteins from all four replicates (Table 4.3). At 96 hr a similar result was seen with an average of 9 proteins being highlighted per replicate with the same criteria, but with 27 non-redundant proteins identified when the four replicates are combined (Table 4.4). In the data from both time points just under half of the identified proteins are down-regulated. This suggests that either there is very little change in the proteome after treatment or an error occurred during treatment which is discussed in Chapter 7. A new batch of salbutamol was purchased and tested in differentiating C2C12. As a difference in the differentiating treated and control cells could be seen by eye it was confirmed that the salbutamol was able to produce the desired effect. The experiment was then repeated under the same conditions with the new salbutamol.

**Table 4.3** ***Modulated proteins after 24hr of treatment in four repeats.*** MaxQuant was used to gain heavy:light ratios for every protein identified. This was achieved using the raw data files produced from the LTQ-Velos. Only those proteins with two or more unique peptides were included when determining ratios. All those with a fold change, either up or down, equal to or >1.5 have been listed in the table for each of the biological replicates.

| Sample | Protein Identified | Fold change | Modulation | Labelling |
|---|---|---|---|---|
| A24 | Cytokeratin-2e | 2.5198 | Down regulated | Light -treated |
| | Thrombospondin-1 | 1.7889 | Down regulated | Light -treated |
| | Histone H1.2 | 1.9815 | Up regulated | Light -treated |
| | Myristoylated alanine-rich C-kinase substrate | 2.166 | Up regulated | Light -treated |
| | Neuronal axonal membrane protein NAP-22 | 2.396 | Up regulated | Light -treated |
| | Cytokeratin-10 | 3.8354 | Up regulated | Light -treated |
| | Cytokeratin-1 | 4.5486 | Up regulated | Light -treated |
| | | | | |
| B24 | Thrombospondin-1 | 2.723 | Down regulated | Light -treated |
| | Annexin A5 | 1.6584 | Down regulated | Light -treated |
| | Heterogeneous nuclear ribonucleoproteins A2/B1 | 2.136 | Up regulated | Light -treated |
| | 35 kDa lectin | 2.2604 | Up regulated | Light -treated |
| | Golgi reassembly-stacking protein 2 | 2.2719 | Up regulated | Light -treated |
| | Eukaryotic translation initiation factor 3 subunit 9 | 2.2868 | Up regulated | Light -treated |
| | | | | |
| C24 | ADP-ribosylation factor 1 | 1.8245 | Up regulated | Heavy-treated |
| | Calcium-transporting ATPase sarcoplasmic reticulum type | 1.7527 | Up regulated | Heavy-treated |
| | Flotillin-1 | 1.507 | Up regulated | Heavy-treated |
| | Thrombospondin-1 | 1.8173 | Down regulated | Heavy-treated |
| | Rho-related BTB domain-containing protein 3 | 2.3098 | Down regulated | Heavy-treated |
| | Cytokeratin-1 | 3.7114 | Down regulated | Heavy-treated |
| | | | | |
| D24 | Pyrophosphate phospho-hydrolase | 2.2198 | Up regulated | Heavy-treated |
| | Chaperonin subunit 2 (Beta), isoform CRA_a | 1.6631 | Up regulated | Heavy-treated |
| | Thrombospondin-1 | 2.3061 | Down regulated | Heavy-treated |
| | Cytokeratin-1 | 6.3155 | Down regulated | Heavy-treated |

**Table 4.4** ***Modulated proteins after 96hr of treatment in four repeats.*** MaxQuant was used to gain heavy:light ratios for every protein identified. This was achieved using the raw data files produced from the LTQ-Velos. Only those proteins with two or more unique peptides were included when determining ratios. All those with a fold change, either up or down, equal to or > 1.5 have been listed in the table for each of the biological replicates.

| Sample | Protein Identified | Fold change | Modulation | Labelling |
|---|---|---|---|---|
| A96 | Endophilin-B1 | 2.9866 | Down regulated | Light -treated |
| | Histone H4 | 1.7713 | Down regulated | Light -treated |
| | Gamma-coat protein | 1.6141 | Down regulated | Light -treated |
| | Integrin alpha-7 | 1.5959 | Down regulated | Light -treated |
| | Ribosomal protein S5 | 1.8133 | Up regulated | Light -treated |
| | Myristoylated alanine-rich C-kinase substrate | 1.9376 | Up regulated | Light -treated |
| | Tetratricopeptide repeat, ankyrin repeat and coiled-coil domain-containing protein 2 | 2.0392 | Up regulated | Light -treated |
| | Cytokeratin-10 | 2.2775 | Up regulated | Light -treated |
| | Cytokeratin-1 | 3.7757 | Up regulated | Light -treated |
| | | | | |
| B96 | Thrombospondin-1 | 2.0935 | Down regulated | Light -treated |
| | Neuronal axonal membrane protein NAP-22 | 1.9485 | Down regulated | Light -treated |
| | Myosin heavy chain 2b | 1.7842 | Down regulated | Light -treated |
| | Myosin heavy chain 8 | 1.646 | Down regulated | Light -treated |
| | Capping protein (Actin filament), gelsolin-like | 1.5133 | Down regulated | Light -treated |
| | Erythrocyte band 7 integral membrane protein | 1.5073 | Down regulated | Light -treated |
| | Ras-related protein Rab-5C | 1.6785 | Up regulated | Light -treated |
| | Cytokeratin-10 | 3.8695 | Up regulated | Light -treated |
| | Cytokeratin-1 | 4.6677 | Up regulated | Light -treated |
| | Cytokeratin-2e | 6.0197 | Up regulated | Light -treated |
| | | | | |
| C96 | ADP-ribosylation factor 1 | 1.8624 | Up regulated | Heavy-treated |
| | Hsc70/Hsp90-organizing protein | 1.8498 | Up regulated | Heavy-treated |
| | Twinfilin-1 | 1.7191 | Up regulated | Heavy-treated |
| | Sulfide:quinone oxidoreductase, mitochondrial | 1.6338 | Up regulated | Heavy-treated |
| | ATP synthase subunit gamma, mitochondrial | 1.6091 | Up regulated | Heavy-treated |
| | Butyryl-CoA dehydrogenase | 1.5108 | Up regulated | Heavy-treated |
| | Eukaryotic initiation factor 5A isoform 1 | 1.7081 | Down regulated | Heavy-treated |
| | Neuronal axonal membrane protein NAP-22 | 2.234 | Down regulated | Heavy-treated |
| | Thrombospondin-1 | 2.6319 | Down regulated | Heavy-treated |
| | Cytokeratin-10 | 3.4917 | Down regulated | Heavy-treated |
| | Cytokeratin-1 | 6.8484 | Down regulated | Heavy-treated |
| | | | | |
| D96 | Calponin-2 | 1.9162 | Up regulated | Heavy-treated |
| | Annexin A4, isoform CRA_b | 1.5942 | Up regulated | Heavy-treated |
| | Cytosolic prostaglandin E2 synthase | 1.5524 | Up regulated | Heavy-treated |
| | Cytokeratin-10 | 2.0107 | Down regulated | Heavy-treated |
| | Cytokeratin-1 | 2.485 | Down regulated | Heavy-treated |

### 4.3.2.2 SILAC 2

The results for the second SILAC experiment were similar to those previously found. At 24 hr, an average of 481 ± 63 more protein identifications are made with a mean of 413 ± 12 more proteins with calculated heavy over light ratios. However, the number of proteins with a fold change >1.5 averaged at 13 ± 9 per repeat. The 96 hr data set has similar results with 482 ± 61 more identifications on average with 763 ± 5 of these identifications having heavy over light ratios calculated by MaxQuant (Table 4.5). When the data from the four replicates were combined a total of 43 and 32 non-redundant proteins were identified with a fold change >1.5 for 24 hr and 96 hr respectively (Table 4.6 and 4.7). Sample S, the fourth replicate, had over twice the number of proteins identified as being modulated compared to the other three replicates.

**Table 4.5** ***Table listing the average number of proteins identified and considered modulated using MaxQuant.*** A second SILAC experiment was performed under the same conditions. C2C12 cells were labelled with heavy or light arginine and lysine before being treated with salbutamol. Samples of both the treated and control cells were collected after 24 hr and 96 hr of treatment. These were analysed using an Orbitrap Velos and the spectra run through MaxQuant to identify proteins from the corresponding peptide peaks. MaxQuant only calculated the heavy:light ratio for those proteins that appeared in all four replicates with more than one peptide. The averages from the four replicates are displayed.

| Replicates | Total identifications | | Total proteins identified with heavy:light ratio | | Total number of proteins modulated | | No. up-regulated | |
|---|---|---|---|---|---|---|---|---|
| | 24 hr | 96 hr | 24 hr | 96 hr | 24 hr | 96 hr | 24 hr | 96 hr |
| P | 1126 | 1148 | 759 | 767 | 7 | 6 | 2 | 1 |
| Q | 1153 | 1125 | 751 | 758 | 10 | 10 | 5 | 5 |
| R | 1122 | 1132 | 750 | 752 | 7 | 7 | 3 | 5 |
| S | 1145 | 1145 | 746 | 774 | 26 | 20 | 22 | 11 |
| Average | 1137 | 1138 | 751 | 763 | 13 | 11 | 8 | 6 |
| Standard Deviation | 14.9 | 10.8 | 5.4 | 9.7 | 9.1 | 6.4 | 9.4 | 4.1 |

**Table 4.6** ***Modulated proteins after 24 hr of treatment in four repeats after a repeat of the SILAC experiment.*** An identical SILAC experiment was perform using heavy and light arginine and lysine to label C2C12 cells. Samples were collected after 24 hr of treatment and analysed via an Orbitrap. MaxQuant was used to gain heavy:light ratios for every protein identified. This was achieved using the raw data files produced from the LTQ-Velos. Only those proteins with two or more unique peptides were included when determining ratios. All those with a fold change, either up or down, equal to or > 1.5 have been listed in the table for each of the biological replicates.

| Sample | Protein Identified | Fold change | Modulation | Labelling |
|---|---|---|---|---|
| P24 | Citrate hydro-lyase | 2.149 | Up regulated | Heavy-treated |
| | Septin-11 | 1.5808 | Up regulated | Heavy-treated |
| | C-terminal LIM domain protein 1 | 1.8487 | Down regulated | Heavy-treated |
| | DNA-binding protein NEFA | 3.0854 | Down regulated | Heavy-treated |
| | Cytokeratin-1 | 4.1616 | Down regulated | Heavy-treated |
| | Cytokeratin-10 | 4.8548 | Down regulated | Heavy-treated |
| | Cytokeratin-9 | 13.4562 | Down regulated | Heavy-treated |
| | | | | |
| Q24 | Elongation of very long chain fatty acids protein 1 | 2.6445 | Up regulated | Heavy-treated |
| | Long-chain acyl-CoA synthetase 3 | 2.0859 | Up regulated | Heavy-treated |
| | 5'-AMP-activated protein kinase subunit gamma-1 | 1.6248 | Up regulated | Heavy-treated |
| | Naphthalene dehydrogenase | 1.5078 | Up regulated | Heavy-treated |
| | Rab-12 | 1.5075 | Up regulated | Heavy-treated |
| | Carboxypeptidase C | 1.6924 | Down regulated | Heavy-treated |
| | 17-beta-hydroxysteroid dehydrogenase 12 | 3.7135 | Down regulated | Heavy-treated |
| | Aldehyde dehydrogenase 18 family, member A1 | 5.9776 | Down regulated | Heavy-treated |
| | Cytokeratin-9 | 7.6313 | Down regulated | Heavy-treated |
| | Cytokeratin-1 | 29.4664 | Down regulated | Heavy-treated |
| | | | | |
| R24 | Long chain 3-hydroxyacyl-CoA dehydrogenase | 2.0223 | Down regulated | Light-treated |
| | 3-ketoacyl-CoA thiolase, mitochondrial | 1.8452 | Down regulated | Light-treated |
| | Formyltetrahydrofolate synthetase | 1.6869 | Down regulated | Light-treated |
| | Elongation factor Tu, mitochondrial | 1.5371 | Down regulated | Light-treated |
| | Tubulin beta-2A chain | 1.7018 | Up regulated | Light-treated |
| | UPF0568 protein C14orf166 homolog | 2.3383 | Up regulated | Light-treated |
| | Aldehyde dehydrogenase 18 family, member A1 | 2.3459 | Up regulated | Light-treated |
| | | | | |
| S24 | High mobility group nucleosomal binding domain 2 | 3.9858 | Down regulated | Light-treated |
| | Thymosin beta-10 | 1.9358 | Down regulated | Light-treated |
| | GTPase-interacting protein 2 | 1.6739 | Down regulated | Light-treated |
| | Calcium-transporting ATPase sarcoplasmic reticulum type, fast twitch skeletal muscle isoform | 1.5378 | Down regulated | Light-treated |
| | Ribosomal protein L23 | 1.9633 | Up regulated | Light-treated |
| | ATP synthase subunit gamma, mitochondrial | 2.0211 | Up regulated | Light-treated |
| | Neuroendocrine-specific protein-like 1 | 2.1969 | Up regulated | Light-treated |
| | Electron transfer flavoprotein subunit beta | 2.2982 | Up regulated | Light-treated |
| | 3-oxoacid-CoA transferase 1 | 2.4112 | Up regulated | Light-treated |
| | Electron transfer flavoprotein subunit alpha, mitochondrial | 2.516 | Up regulated | Light-treated |
| | Serine carboxypeptidase 1 | 2.5248 | Up regulated | Light-treated |
| | 2-methylbutyryl-coenzyme A dehydrogenase | 2.6699 | Up regulated | Light-treated |
| | Hydroxysteroid (17-beta) dehydrogenase 10 | 2.747 | Up regulated | Light-treated |
| | Phosphatidylinositide phosphatase SAC1 | 2.8013 | Up regulated | Light-treated |
| | Succinyl-CoA ligase [GDP-forming] subunit alpha, mitochondrial | 3.6049 | Up regulated | Light-treated |

| | | | | |
|---|---|---|---|---|
| | Glycine amidinotransferase, mitochondrial | 3.747 | Up regulated | Light-treated |
| | ATP synthase subunit beta, mitochondrial | 3.9518 | Up regulated | Light-treated |
| | Isocitrate dehydrogenase [NADP], mitochondrial | 3.9703 | Up regulated | Light-treated |
| | 3-ketoacyl-CoA thiolase | 4.2781 | Up regulated | Light-treated |
| | Cytokeratin-9 | 4.4877 | Up regulated | Light-treated |
| | ATP synthase subunit alpha, mitochondrial | 4.9085 | Up regulated | Light-treated |
| | ATP-specific succinyl-CoA synthetase subunit beta | 5.1414 | Up regulated | Light-treated |
| | Long chain 3-hydroxyacyl-CoA dehydrogenase | 6.3099 | Up regulated | Light-treated |
| | Cytokeratin-1 | 6.6212 | Up regulated | Light-treated |
| | Cytokeratin-10 | 17.0369 | Up regulated | Light-treated |
| | Cytokeratin-2e | 38.0387 | Up regulated | Light-treated |

**Table 4.7** ***Modulated proteins after 96 hr of treatment in four repeats after a repeat of the SILAC experiment.*** An identical SILAC experiment was perform using heavy and light arginine and lysine to label C2C12 cells. Samples were collected after 96 hr of treatment and analysed via an Orbitrap. MaxQuant was used to gain heavy:light ratios for every protein identified. This was achieved using the raw data files produced from the LTQ-Velos. Only those proteins with two or more unique peptides were included when determining ratios. All those with a fold change, either up or down, equal to or > 1.5 have been listed in the table for each of the biological replicates.

| Sample | Protein Identified | Fold change | Modulation | Labelling |
|---|---|---|---|---|
| P96 | N-ethylmaleimide-sensitive factor attachment protein alpha | 1.5322 | Up regulated | Heavy-treated |
| | Cytokeratin-10 | 3.376 | Down regulated | Heavy-treated |
| | Acetoacetyl-CoA thiolase | 3.399 | Down regulated | Heavy-treated |
| | Proteasome (Prosome, macropain) 26S subunit ATPase 3 | 5.3084 | Down regulated | Heavy-treated |
| | Cytokeratin-1 | 7.3975 | Down regulated | Heavy-treated |
| | Cytokeratin-9 | 13.2723 | Down regulated | Heavy-treated |
| | | | | |
| Q96 | UV excision repair protein RAD23 homolog B | 2.2698 | Up regulated | Heavy-treated |
| | Long chain 3-hydroxyacyl-CoA dehydrogenase | 1.8169 | Up regulated | Heavy-treated |
| | Actopaxin | 1.7604 | Up regulated | Heavy-treated |
| | Nuclear matrix protein 200 | 1.7266 | Up regulated | Heavy-treated |
| | Septin-11 | 1.5008 | Up regulated | Heavy-treated |
| | Cytokeratin-10 | 2.0511 | Down regulated | Heavy-treated |
| | FAS-associated factor 2 | 2.1625 | Down regulated | Heavy-treated |
| | C33 antigen | 2.6159 | Down regulated | Heavy-treated |
| | Utrophin | 5.3758 | Down regulated | Heavy-treated |
| | Cytokeratin-1 | 5.4324 | Down regulated | Heavy-treated |
| | | | | |
| R96 | IFN-response element binding factor 2 | 1.9266 | Down regulated | Light-treated |
| | SH3 domain-binding glutamic acid-rich-like protein 3 | 1.7369 | Down regulated | Light-treated |
| | APOBEC1-binding protein 2 | 2.41 | Up regulated | Light-treated |
| | Hepatoma-derived growth factor | 2.8184 | Up regulated | Light-treated |
| | Cytokeratin-10 | 5.139 | Up regulated | Light-treated |
| | Cytokeratin-9 | 5.6709 | Up regulated | Light-treated |
| | Cytokeratin-1 | 6.5686 | Up regulated | Light-treated |
| | | | | |
| S96 | Thymosin beta-10 | 2.6147 | Down regulated | Light-treated |
| | Calcyclin-binding protein | 2.5895 | Down regulated | Light-treated |

| | | | | |
|---|---|---|---|---|
| | Thymosin beta-4 | 2.3691 | Down regulated | Light-treated |
| | SH3 domain-binding glutamic acid-rich-like protein 3 | 2.0248 | Down regulated | Light-treated |
| | NADH dehydrogenase [ubiquinone] iron-sulfur protein 8, mitochondrial | 1.8572 | Down regulated | Light-treated |
| | Prothymosin alpha | 1.8339 | Down regulated | Light-treated |
| | Heterogeneous nuclear ribonucleoprotein A0 | 1.5308 | Down regulated | Light-treated |
| | Microtubule-associated protein 7 domain | 1.5285 | Down regulated | Light-treated |
| | Eukaryotic initiation factor 5A isoform | 1.5267 | Down regulated | Light-treated |
| | Proteasome chain 13 | 2.0434 | Up regulated | Light-treated |
| | Proteasome (Prosome, macropain) 26S subunit, non-ATPase, 12 | 2.0776 | Up regulated | Light-treated |
| | Protein kinase C and casein kinase II substrate protein 3 | 2.1725 | Up regulated | Light-treated |
| | Hepatoma-derived growth factor | 2.4305 | Up regulated | Light-treated |
| | Cytosol aminopeptidase | 2.7178 | Up regulated | Light-treated |
| | Aldehyde dehydrogenase 18 family, member A1 | 2.8961 | Up regulated | Light-treated |
| | Proteasome (Prosome, macropain) 26S subunit ATPase 3 | 4.3885 | Up regulated | Light-treated |
| | Cytokeratin-10 | 4.6352 | Up regulated | Light-treated |
| | Cytokeratin-9 | 6.8951 | Up regulated | Light-treated |
| | Cytokeratin-2e | 8.0444 | Up regulated | Light-treated |
| | Cytokeratin-1 | 12.4642 | Up regulated | Light-treated |

In the first experiment at 24 hr, thrombospondin-1 (Table 4.3) is identified in all four repeats as being down regulated. This protein is found in two of the repeats in the 96 hr sample set (Table 4.4) but not identified in the second SILAC experiment (Tables 4.6 and 4.7). Cytokeratin 1, 2e, 9 and 10 are seen throughout both experiments and the two time points. At least one of these proteins was identified in 14 of the 16 replicates taken over the two experiments. Interestingly, when comparing proteins identified at each time point separately it was noticed that the cytokeratins were sometimes identified as being up regulated in one replicate and down regulated in another. Very few proteins are identified in more than one replicate but as this may be due to the mass spectrometry analysis the total protein list of modulated proteins were analysed further to identify if the proteins had similar functions. Figure 4.6 visualises the different numbers of similar functions or biological processes of the 58 non-redundant proteins identified over the two experiments (8 replicates) at 24 hr. The biological process of each identified protein was determined using

UniProtKB (http://www.uniprot.org). The most modulated processes are regulation of transcription and translation, cytoskeletal and fatty acid regulation. The same analysis was performed for the data collected over the two experiments at 96 hr, resulting in 56 non-redundant proteins. From Figure 4.6 it can be observed that cytoskeletal proteins, fatty acid regulation and regulation of transcription and translation are processes containing a large amount of the modulated proteins. However, there are more proteins identified involved in cell morphology. A total of 9% are also proteins identified with degradation, apoptosis and DNA damage functions, suggesting the cells are undergoing different changes to those at 24 hr.

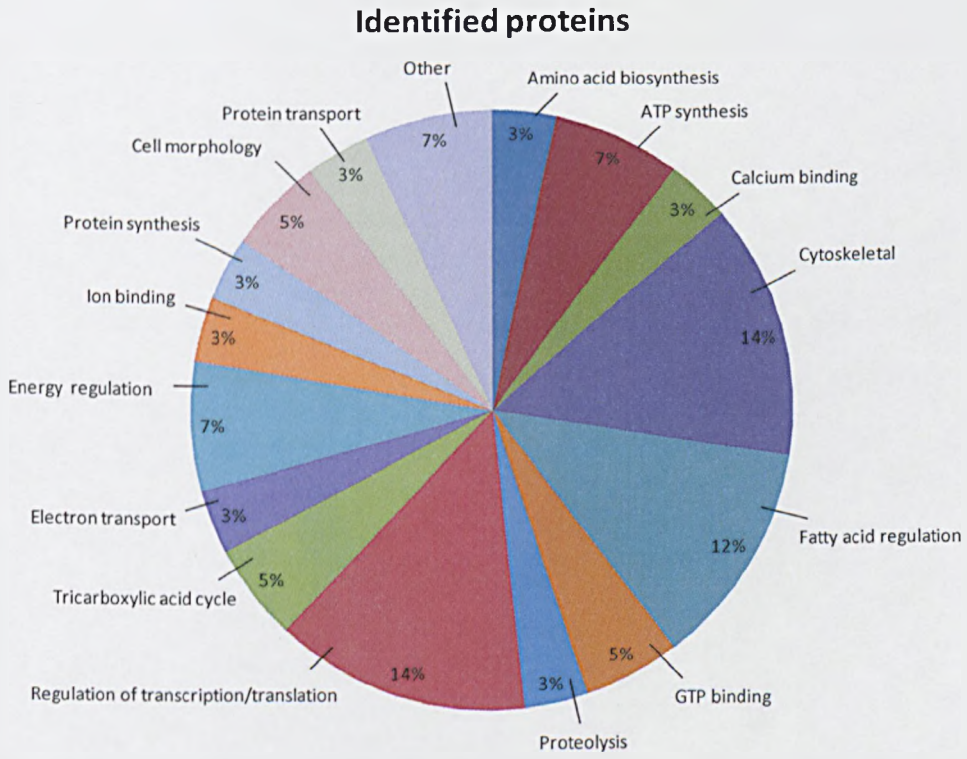



**Figure 4.5** ***Pie chart representing the protein functions of those identified as modulated over 8 replicates at 24 hr.*** A SILAC study using heavy or light arginine and lysine to label C2C12 mouse muscle cells was undertaken. Half the cells were treated with salbutamol. Control and treated samples were collect at 24 hr after treatment, mixed, digested and analysed via an Orbitrap. Proteins were identified with MaxQuant. This experiment was repeated equalling a total of 8 replicates. n=58. All proteins with a fold change >1.5 were identified and functions determined.

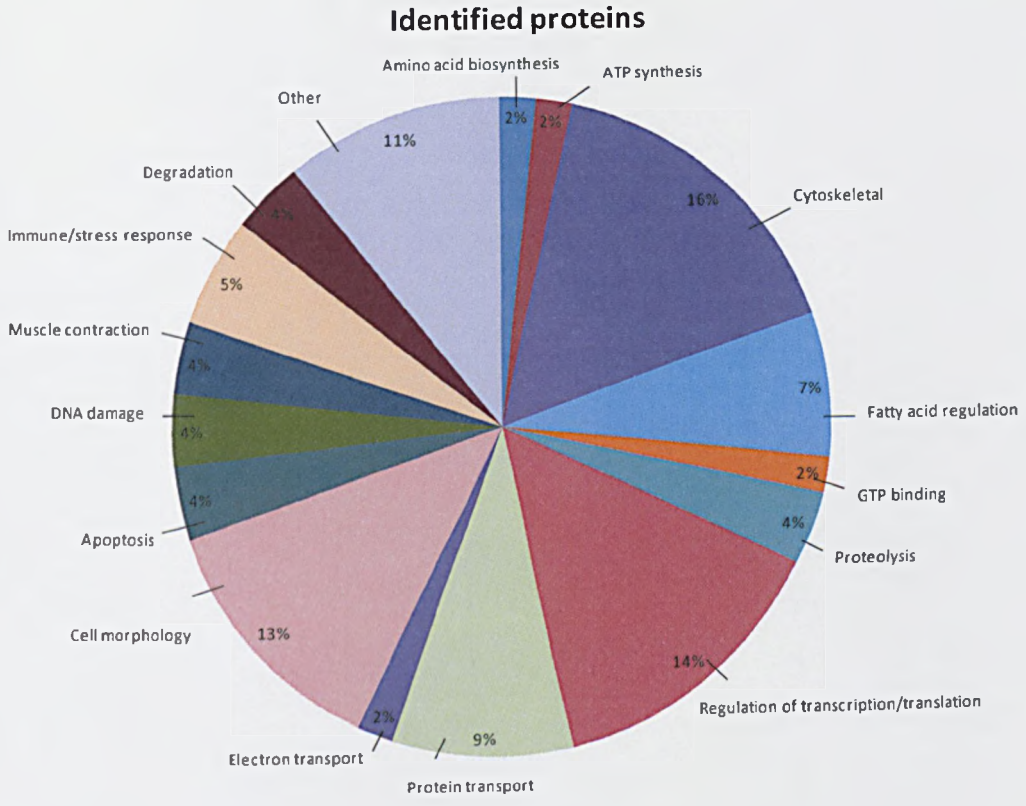



**Figure 4.6** ***Pie chart representing the protein functions of those identified as modulated over 8 replicates at 96 hr.*** A SILAC study using heavy or light arginine and lysine to label C2C12 mouse muscle cells was undertaken. Half the cells were treated with salbutamol. Control and treated samples were collect at 96 hr after treatment, mixed, digested and analysed via an Orbitrap. Proteins were identified with MaxQuant. This experiment was repeated equalling a total of 8 replicates, n=56. All proteins with a fold change >1.5 were identified and functions determined.

## 4.4 Discussion

SILAC was chosen for this study over other labelling techniques due to its added advantages; the labelling is not restricted to proteins containing cysteine residues as in ICAT and the proteins are uniformly labelled enabling several peptides from the same protein to be compared to identify if the changes are the same (Ong *et al.*, 2002). This technique has, however, presented an interesting phenomenon within this study. From the results gained it would seem as if the C2C12 cells undergo very little change at a protein level after treatment with salbutamol. However, Chapter 3 identified several hundred gene modulations suggesting a large protein change would be occurring. To understand the meaning behind the data collected each possible scenario and limitation needs to be analysed.

A small number of proteins were identified with a fold change > 1.5. For example; in the first experiment, thrombospondin-1 is identified in all four repeats as being down regulated. This protein is an adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions. When C2C12 cells undergo differentiation, the myoblasts fuse together forming myotubes. As this protein is down regulated it suggests that the growth induced by salbutamol is not due to continuous fusion and cell-cell interaction. However, this protein was not identified when the experiment was repeated. This may be due the differences occurring at a mass spectrometry run level. If peptides perform slightly differently in each run or the sample is very complex meaning not all peptides may be identified, this would cause varying results. A large number of cytoskeleton proteins in the form of cytokeratin were identified in both experiments in more than one repeat. Cytoskeleton proteins in the

form of keratin have often been identified as contaminates in mass spectrometry analysis, as these proteins were seen as up-regulated in one repeat but down-regulated in another it is difficult to determine if these are true changes. There are other cytoskeletal proteins identified such as tubulin, thymosin and twinfilin which suggest that there is an effect within the structure of the cells, which are involved with the organisation of the cytoskeleton as well as filament forming.

Although only a few proteins were identified in all the replicates, when comparing the biological processes of those found there was a similarity across the repeats but differ slightly across time points. There were a number of proteins indentified, at both 24 and 96 hr, which are involved in the acetyl-CoA conversion and fatty acid regulation as well as the regulation of transcription and translation. All the proteins with a role in fatty acid metabolism, such as long-chain acyl-CoA sythetase 3 and 2-methylbutyryl-coenzyme A dehydrogenase, were identified as being up-regulated. However, 3-ketoacyl-CoA thiolase was identified as being down-regulated in at 24 hr in one repeat but up-regulated in another. This may be due to an isoform or a post-transcriptional modification on the protein. Eukaryotic initiation factor 5A isoform was one of the translation and transcriptional proteins identified. This protein is observed at 96 hr as being down-regulated in both SILAC experiments with similar fold changes in both studies (1.5 and 1.7). Eukaryotic initiation factor 5A (eIF5A) is an mRNA-binding protein involved in translation elongation. Several different functions have been attributed to eIF5A, including its role in cell proliferation (Hanauske-Abel *et al.*, 1994; Parreiras-e-Silva *et al.*, 2010) and apoptosis (Jin *et al.*, 2003) as well as its interaction with specific sequences of mRNAs suggesting a possible role in mRNA decay (Zuk & Jacobson, 1998). This is

not the only protein identified as modulated with a role in apoptosis and autophagy. At 96 hr both endophilin-B1 and FAS-associated factor 2, known to have a role in apoptosis (Imai *et al.*, 2002; Takahashi *et al.*, 2005; Yang *et al.*, 2010), are identified as being down regulated. Energy production is another important process during and cell differentiation or growth. At 24 hr several proteins involved in energy regulation and ATP synthesis are identified as up-regulated, such as ATP synthase, glycine amininotransferase and 5'-AMP-activated protein kinase. Although there are several interesting proteins and biological process highlighted, due to the lack of identification of proteins across more than one repeat it is difficult to draw a statistically significant conclusion.

### 4.4.1 Identification of scenarios resulting in little differentiation of the proteome

This study has identified only a small number of differentially modulated proteins compared to the number of genes modulated in Chapter 3. This difference may be due to limitations which could prevent the analysis of the proteome occurring. Potential problems and limitations can arise in both the experimental procedure as well as the methodology itself. These limitations are discussed below in an attempt to further understand the results gained.

#### 4.4.1.1 Determining the effectiveness of the labelling process

To deter any error occurring from the labelling technique, systematic checks were undertaken during the experiment by sampling the cells and analysing randomly picked peptides to determine the amount of labelling. Previous studies of a similar

type have used a standard of greater than five doublings in the heavy media to allow for complete labelling (Cui *et al.*, 2009; Graumann *et al.*, 2008; Ong *et al.*, 2002; Pan *et al.*, 2008). For C2C12 cells one doubling is equal to one day. For this experiment six doublings were chosen as a suitable value for the labelling as this allowed for the heavy amino acids to be incorporated as well as leaving the cells in a position to be treated before reaching cell death. To confirm that the cells were fully labelled, a sample of each biological replicate was taken after six doublings, prior to treatment verify the absence of light peptides. A random selection of peptides identified as being labelled with the heavy amino acids were chosen and the spectra was manually checked to assess for any light peptide peaks for the corresponding peptide. As only the weight of the peptides should have changed when labelled with the heavy amino acids the corresponding peptide should be found at a shift of either 8 Da or 10 Da depending on which amino acid it is labelled with. These did not highlight any light peptides but when run through MASCOT a small number of peptides appeared without labelling, which means a 100% labelling has not occurred. This number was less than 1% of the total peptide count and so was assumed to be reasonable. Normalisation of the results would overcome this problem.

It is unlikely that the results gained were due to a lack of labelling as this was carefully examined. SILAC is a well used technique that has been proved to work efficiently in other studies so it is mostly likely due to some other limitation or biological phenomenon.

#### 4.4.1.2 Loss of sample could cause bias in results

As part of the sample preparation for mass spectrometry analysis the collected cells needed to be lysed. To achieve this, the cells were mixed with a lysis buffer and subjected to a freeze/thaw technique. To allow for solubilisation of the proteins a detergent called CHAPS was present. CHAPS, along with urea, thiourea and glycerol for the OFFGEL samples needed to be removed before samples could be run on the mass spectrometer. Although the precipitation step was able to remove the glycerol, urea, tris and thiourea the CHAPS was still present in an amount determined as high enough to cause damage to the instrument. Precipitation of the sample before focusing as well as after was then undertaken. Again CHAPS was present in the sample but at a much lower level. Unfortunately, there was a loss of peptides and MASCOT was unable to identify as many proteins as usual.

Using different precipitation methods produced similar results. The first precipitation was then removed and instead the peptides were cleaned using stage tips after digestion. Although this reduced the loss of sample the problem with the CHAPS reappears. A desalting column was also tested as an alternative clean up method. Again, this procedure was able to clean the sample but resulted in a huge loss of the sample meaning it was not a viable method. An attempt to reduce the amount of CHAPS in the lysis buffer was then tested. The concentration of CHAPS was reduced to its lowest amount (0.6 % (w/v)) before being ineffective and the solution buffered using HEPES buffer. This, along with removing the freeze/thaw step, was to reduce any CHAPS dropping out of solution during the precipitation. Eventually, this solution combined with a two day precipitation method and using non-binding tubes was able to produce a result where the CHAPS peak did not

overwhelm the peptide results. With all of these methods protein loss was seen, if the loss of protein in these steps is directed only at a specific group of proteins a bias in the results could be obtained.

Overall, using a lysis buffer containing a detergent, as well as urea and thiourea was causing large sample losses during the clean up stages performed after lysis. By using RapiGest instead of the lysis solution more of the protein could be recovered and taken forward. Although this decreased any bias in the result due to the reduced losses, not all the protein could be recovered, for example the insoluble proteins are lost at the lysising stage. For the second SILAC experiment, the insoluble proteins were collected and stored at -20°C. With more time it would have been highly informative to be able to study these proteins to determine if there was a change that was not identified using this method which only studied the soluble proteins. The results using the RapiGest instead of the lysis buffer are similar to the original suggesting the sample loss is not a cause of the limited change detected in the proteome.

#### 4.4.1.3 Biological explanation of results

When cells are grown in the presence of a stimulus, it is extremely unlikely that no change within the cell would be seen. The cell would be under a type of stress resulting in at least a response to protect itself. In this case, the lack of response may be due to the cell line used. The cells were left to become fully differentiated to allow the cells to model mammalian muscle cells without using live animals. However, due to this decision the problem may have arisen. It is possible that once

the cells become differentiated they are unable to grow any further and so do not respond to stimulus. Another reason behind the results gained may be due to the protein change being extremely small and not identified using this method. It is possible that a much larger change in gene expression is needed to produce a small change in the protein. A third reason could be due to the time lag between gene expression and protein modulation. The majority of gene modulations occurred at 24 hr and 96 hr meaning that a large protein change may not occur until after these time points in the proteome. This theory is supported by the fact there were a much smaller number of modulated genes at 2 hr and 6 hr (Chapter 3 section 3.3.4) which would correlate with the small protein change at 24 hr and 96 hr. However, although a smaller number of genes were identified as modulated compared to data collected at 24 hr and 96 hr a total of 161 and 299 genes at 24 hr and 96 hr, respectively, were identified. This is considerably higher than the number of proteins observed as modulated in this study. As the link between these two areas is still not fully understood, differences between these two experiments may not correlate due to limitations causing different identifications to be made. Other studies have identified similar differences between gene expression and protein expression. Griffin *et al.,* (2002) observed differences between the abundance ratio of the messenger RNA transcript and the corresponding protein product in *Saccharomyces cerevisiae*. Fu *et al.,* (2009) investigated the gene expression and protein expression of human brain samples and observed a moderate correlation between RNA-Seq expression measurements and protein expression data (Pearson correlation of 0.36).

#### 4.4.1.4 Analysis of pre-fractionation

Due to the minimal changes in protein expression it is difficult to analyse the benefits of separation before mass spectrometry analysis. Using fractionation a total of six samples from one cell lysate were obtained as the 12 wells are combined into 6. Each of these 6 samples was run on the LTQ-Velos and all the spectra combined to analyse the changes within the one cell lysate. However, without fractionation the one cell lysate was digested and run once on the LTQ-Velos. As the pre-fractionation method allowed one cell lysate to not only be simplified but also to have a total of 6 runs on the mass spectrometer the total data was much greater. To investigate further the benefits of separation, the same amount of runs for both samples should be undertaken to remove this variation. If more changes had been identified it would have also been possible to compare the results gained to see if the same group of proteins were identified by the two methods. Overall, a larger number of protein identifications were made using the OFFGEL fractionators. For one sample over 1800 identifications were made compared to the 656 made without separation.

### 4.4.2 Modulated proteins identified from the literature

From the results of the DIGE and transcriptomic experiments, a large number of proteins being modulated were expected in the SILAC experiment. The DIGE experiment identified a larger number of proteins being modulated in 3 out 4 of the repeats then the SILAC experiment. The transcriptomic highlight a very large number of genes being differentially expressed which would suggest a number of proteins would be modulated though the increase may not be large. As the DIGE experiment used rat primary muscle cells, and the correlation between gene

transcripts and protein products is still unclear, it was hypothesised that a number of proteins previously identified as not being modulated due to salbutamol may be highlighted. Previous studies that have examined the hypertrophic effects of salbutamol and other β-agonists have highlighted results which relate to our previous findings. For example, recent reports suggest that β-agonists increase the expression of several contractile proteins (Oishi *et al.*, 2002; Spurlock *et al.*, 2006) and myogenin (Delday & Maltin, 1997) both of which were expected to be found in this study. Due to data gained in Chapter 3 it was hypothesised that MAPK proteins would be identified due to the large amount of gene expression change in the transcriptome. Creatine is known as a key intermediate in the energy metabolism of muscle fibres (Young *et al.*, 2010). It was hoped that the proteins that it interacts with would also be identified to confirm its link with steroid use and hypertrophy. Supplementation of creatine is often used among athletes as a means for increasing muscle mass, strength and endurance (Casey *et al.*, 1996; Mujika & Padilla, 1997). For elderly people, creatine supplementation has been seen to enhance muscle strength (Gotshalk *et al.*, 2008). These studies along with creatine kinase identified in the DIGE experiment provoked interest in this metabolic pathway. Although little can be determined from the results gained, one protein identified at 24 hr was glycine amidinotransferase. This protein is known to catalyse the biosynthesis of guanidinoacetate, the immediate precursor of creatine. This result links in with previous studies (Warriss *et al.*, 1990) involving β-agonists but further work needs to be undertaken to determine if the result is a substantial finding. Both previous studies (Chapters 2 and 3) have highlighted a change in the calcium channels and calcium signalling. As β-agonists are thought to act through G-coupled receptors it is likely that calcium would be activated as a secondary messenger. Calcium

signalling and channels have been seen to have a wide range of functions on many different areas, within different cells, giving reasonable weight to these being identified in a proteomic experiment such as this. There are a great number of other possible pathways and proteins that were expected to be affected but as the results show none of these new questions have been brought forward.

### 4.4.3 Overall conclusions

In summary, there was a small change within the muscle proteome after treatment with salbutamol but much smaller to that seen at a genomic level. Some potentially interesting biological processes were highlighted at both time points, for example the change of fatty acid metabolism and cell morphology particularly cell adhesion and cytoskeletal structure. Efforts have been made to ensure all steps of the method work efficiently and large amounts of data have been gained. The proteins which have been identified were often only identified in one or two of the repeats which makes it difficult to ascertain if they are true changes. If this experiment was taken on its own it would seem that the β-agonist, salbutamol, has only a subtle effect on the proteome in mouse muscle cells. However, a physical morphological difference was seen under a microscope when the cells were treated during differentiation which would suggest a larger protein change would be occurring. This, along with the transcriptome data, allows a possible conclusion that a very small proteome change has occurred from a large genomic change. One possible scenario to why only a small change within the proteome has occurred may be due to the cell line not responding to stimulus once it has passed the differentiation stage. Another explanation may be due to the effects on a protein level being too small to identify.

Limitations in the SILAC technique or an unknown factor could also lead to a bias in the results suggesting a lack of differentiation, though this is unlikely to cause such a small number to be identified. However, the data gained was able to identify areas of potentially interesting processes as well allowing the potential benefits to pre-fractionation to be examined.

# 5 Chapter 5

# Further investigation into the effect of salbutamol on the mouse muscle proteome using label-free mass spectrometry analysis.

## 5.1 Introduction

Labelling techniques have great advantages in proteomics; however they still have some flaws. Errors in quantification can arise from incomplete labelling as well as difficulty in simultaneously quantifying the amount of proteins/peptides in multiple samples due to limited numbers of labelling reagents. Currently, only Tandem Mass Tags (TMT) and iTRAQ allow for the comparison of up to eight samples at the same time. Other labelling methods only compare the relative quantity of a protein between two or three different samples. Also, current labelling methods can typically only quantify a few hundred peptides. Furthermore, the high costs of labelling reagents, the increased time and complexity of the sample preparation and the requirement for higher sample concentrations make these quantification methods difficult to be commonly applied. Recently, techniques have been developed which remove labelling from the equation and use mass spectrometry to identify any differences between samples. Due to mass spectrometry technology increasing in accuracy and efficiency these new proteomic quantification methods are being used more frequently. Label free techniques have the advantage of low cost and simplicity. There are currently two very different protein quantification strategies used for label-free quantification: spectral counting and peptide chromatographic peak intensity measurements.

### 5.1.1 Spectral counting

Spectral counting is based on the observation that if there is a linear correlation between the MS/MS spectra acquired of peptides from a particular protein to the relative concentration of said protein (Zybailov *et al.*, 2005). Relative quantification

can therefore be calculated by comparing the number of spectra between experiments. An advantage of this technique is that both protein identification and protein quantification can be achieved with extensive MS/MS data collected across a chromatographic time scale simultaneously. This technique is controversial due to it not measuring any direct physical property of a peptide. It also relies on the spectra quality and any error would lead to inaccurate quantification. Another disadvantage to this technique is that it only provides relative quantification and assumes that all peptides will respond identically during analysis as well as being detected equally well during MS. To achieve the high accuracy needed, spectral counting requires a high-resolution and fast mass spectrometer.

### 5.1.2 Peptide chromatographic peak intensity measurements

Peptide ion intensity measurement uses peak areas from extended ion chromatograms to track peptide and protein abundance across samples. This is achieved by comparing the intensity values for each peptide, between one or more experiments, to gain the relative quantification. Chelius *et al.*, (2002) demonstrated that there is a linear correlation between the peak areas for a peptide and its concentration. An advantage to this strategy is the possibility of distinguishing between distinct peptide sequences with interfering signals of similar masses due to the zoom scan and MS/MS data. However, the limiting factor to this method is the requirement for obtaining assigned MS/MS spectra prior to peak matching. If the acquisition of MS/MS spectra between experiments is low, it can reduce the number of proteins that are able to be compared. The chromatographic profile needs be optimised to allow for reproducibility and for corresponding peptides between experiments to be found without too much difficulty (Bantscheff *et al.*, 2007). An

alternative method has been proposed which involves the mass spectrometer no longer cycling between MS and MS/MS mode, but instead to detect and fragment all peptides simultaneously by rapidly alternating between high- and low-energy conditions in the mass spectrometer known as $LCMS^E$ (Bantscheff *et al.*, 2007). Peptide ion intensity measurements will be utilised for this investigation and an outline of the method to be used is provided in Figure 5.1.

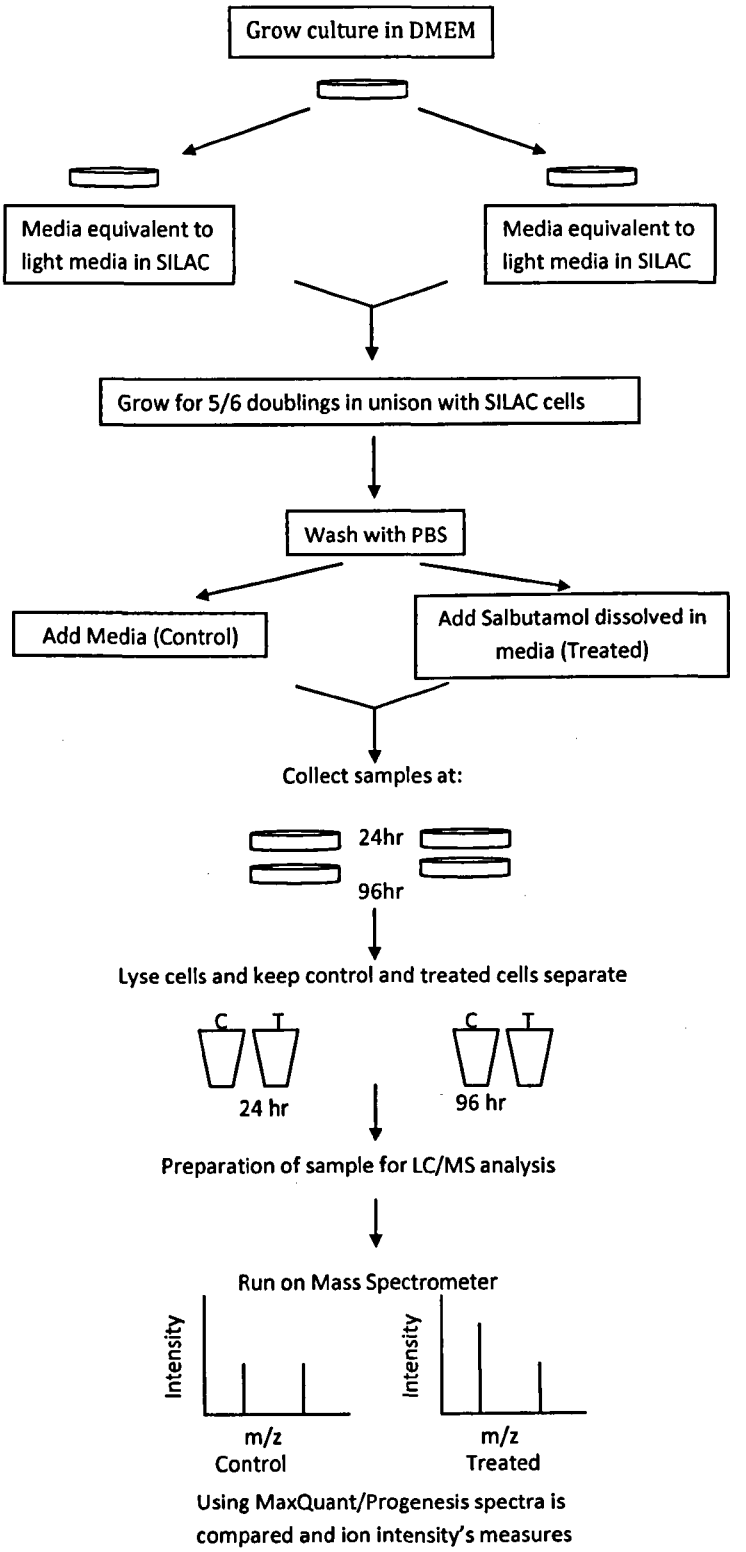



**Figure 5.1 Overview of Label Free workflow**

During the first half of the Label Free method the C2C12 cells will be grown in unison with the SILAC and RNA-Seq cells. Cells will be grown in the equivalent of the SILAC light media to allow direct comparison between these techniques. Once half the cells have been treated with salbutamol the samples can be harvested but kept separate. Once digested, they can be run on a mass spectrometer to gain spectra for both the treated and control samples. Identifications and quantification can be made through either MaxQuant or Progenesis which will compare the intensities for the same ions in both samples.

### 5.1.3 Protein identification and quantification

From any mass spectrometry run, hundreds to tens of thousands of fragment ion spectra can be generated. To assign all of the fragment ion spectra to peptide sequences, as well as the identification of proteins represented by each peptide and the estimation of their abundances, a complex computational system is required and is still a high statistical challenge. Protein identification from the spectra is the first task undertaken. Usually, MASCOT, SEQUEST and X! Tandem database search algorithms are used to assign peptide sequences to protein identifications. The confidence in protein identification is increased with the number of unique sequences identified. Therefore, it is often practice to categorize proteins into different priority groups depending on the number of distinct sequences matched.

To determine protein quantification for spectral counting, the numbers of MS/MS spectra from the same protein are analysed. There are no specific tools or algorithms developed for spectral counting due to its ease of implementation. However, normalization and statistical analysis of spectral counting data is important for accurate and reliable detection of protein modulation. Dong *et al.,* (2007) reported on a simple normalization method to account for the variation from run to run. It is also important to take into consideration the length (Powell *et al.*, 2004) or sequence (Ishihama *et al.*, 2005) of a protein when determining the abundances of proteins using spectrum counting. This is due to smaller proteins tending to have fewer peptides identified per protein compared to larger proteins. Therefore, a normalized spectral abundance factor (NSAF) was created to account for the effect of protein length on spectral count (Florens *et al.*, 2006). The NSAF is determined by the number of spectral counts (SpC) identified for a protein, divided by the protein's

length ($L$) and then divided by the sum of SpC/$L$ for all proteins in the experiment. NSAF allows the comparison of individual proteins in several independent samples and has been applied to help quantify the expression changes (Florens *et al.*, 2006).

$$(\mathrm{NSAF})_K = \frac{\left(\mathrm{SpC}/\mathrm{L}\right)_k}{\sum_{i=1}^{N}\left(\mathrm{SpC}/\mathrm{L}\right)_i}$$

However, for peak intensity measurements it is the chromatographic peak that is measured. Due to the large amount of data collected during the LC-MS/MS runs, the data analysis of these spectra requires an automated process. Computer algorithms were developed to compare the peak intensities between LC-MS samples automatically. Peptide peaks are first distinguished from any background noise or neighbouring peaks. LC-MS retention times are adjusted to be able to correctly match the corresponding mass peaks between the different LC-MS runs. Either peak area or peak height is calculated for the chromatographic peak intensity and then normalized to enable accurate matching and quantitation. Lastly, a statistical analysis is performed to determine the significance of changes between multiple samples (Higgs *et al.*, 2005; Wiener *et al.*, 2004). For both of these methods controlling peak alignment is critical in order to allow the most accurate comparative data. To qualify for being used in protein quantification each aligned peak must match precursor ions, charge states, fragment ions and retention times.

Normalization of data collected by both methods needs to be performed to reduce the variance observed between samples and replicates. For both label free techniques variability can be caused by different factors such as; sample preparation, gel-to-gel

differences (when gel based separation is used pre-mass spectrometry analysis) and changes in chromatography. Carvalho *et al.,* (2008) amongst others have shown the importance of normalizing data when quantifying proteins. Some of the current techniques available include; central tendency (Yang *et al.*, 2002), linear regression (Park *et al.*, 2003), local regression (Cleveland & Devlin, 1988), and quantile (Bolstad *et al.*, 2003) normalisation. However, the most effective normalization method for label free techniques has yet to be determined.

### 5.1.4 Research using label free techniques

To date there has been no research into the effects of β-agonists on muscle cells using label free techniques. However, there have been many studies showing the successful use of both label free methods for identifying and quantifying proteins in a range of samples. These techniques are often utilized over labeling techniques due its ability to compare a number of samples as it is not reliant on available labelling. Also, this technique is less expensive as isotope labels do not need to be purchased. Friso, *et al.,* (2011) were able to effectively use spectral counting for a comparative proteome analysis of *Arabidopsis* mutants with reduced levels of chloroplast (Clp) protease subunits as well as other comparative analyses. Silva *et al.,* (2006) showed successful use of a label free method for a qualitative and quantitative analysis of proteins within *E.coli.* The label free methodology has also been used to investigate the proteome of muscle of both human and animals. Fraterman *et al.,* (2007) were able to use a combination of OFFGEL fractionation and label free ion intensity measurements to identify 61 proteins that were differently expressed between two extraocular muscle types. In a different study Fraterman *et al.,* (2007) utilized a quantitative label-free shotgun proteomics approach to differentially quantitate

sarcomeric proteins; in total 40 sarcomeric proteins were identified. Cutillas *et al.*, (2007) used peak intensity measurements to quantify the most abundant proteins in five mouse tissues; brain, heart, kidney, liver, and lung. Hammer *et al.*, (2011) demonstrate the potential use for label-free approaches for identifying disease dependent modulations in the proteome within human heart biopsy samples.

All of these examples help justify using label free techniques as an alternative to stable isotope labelling, not only due to it being cheaper than labelling peptides but because of its ability to not be restricted by the amount of experiments that can be compared. They also advocate the use for this experiment as a way of not only examining the advantages to label free but to also validate previous findings and potentially highlighting new ones.

### 5.1.5 Aims

It is currently unclear whether using only one technology can gain complete and quantitative protein coverage of all proteins in a given tissue. It is therefore the aim of this chapter to use a different proteomic technology to confirm and increase the data already acquired. Label free techniques have proven to be a useful method to quantify proteins without the use of labeling. The first aim was to undergo optimization of the method to ensure samples were compatible on the LTQ-Orbitrap mass spectrometer as well as allowing comparisons to the previous SILAC (chapter 4) and transcriptomic (chapter 3) experiments. The second aim was to perform data analysis of the spectra to gain relative quantification of proteins being modulated by salbutamol by using the method of peak intensity measurements. The last aim is to

identify the biological relevance of the proteins identified and whether new information had been gained. This would then allow a conclusion on the necessity, if any, to using multiply techniques for quantification analyses.

## 5.2 Method

### 5.2.1 Cell Culture

#### 5.2.1.1 Growing C2C12 cells

The C2C12 cell line used for the SILAC and transcriptomic experiment (Chapter 3 and 4) was again the choice of model system. These cells were grown at the same time as the previous experiments as myoblasts in 25 $cm^2$ bottom (T25) vented cell culture flasks (BD Falcon$^{TM}$) with incubation at 37°C in a 5% $CO_2$ humidified incubator. As previously described, the cells were initially grown in filter sterilised DMEM medium (Sigma-Aldrich) supplemented with 10% (v/v) FCS and 1% (v/v) L-glutamine (Sigma-Aldrich) and 1% (v/v) penicillin-streptomycin (Sigma-Aldrich). A stock medium (DMEM low glucose minus arginine and lysine media (Sigma-Aldrich) including: 0.0159 g/L phenol red, 0.0001g/L proline, 0.105g/L L-leucine, 0.11g/L sodium pyruvate, 4.5g/L D-glucose, 1 % (v/v) penicillin-streptomyocin, 10% (v/v) fetal calf serum) which was used in all three experiments was initially assembled. L-arginine (0.08% (v/v)) and L-lysine (0.15% (v/v)) were added to complete the medium before transferring to flasks once cells had been split (5.2.1.2). These amino acids were made up as stock solutions of 0.1g/ml in DMEM low glucose media.

#### 5.2.1.2 Splitting Cells

As stated previously these cells need to be split before they become confluent. To split a 25 $cm^3$ flask of cells, the old medium was removed, the flask washed with 5 ml of HEPES Buffer (Sigma-Aldrich), buffer disposed of and 5 ml of trypsin

(Sigma-Aldrich) added before incubating for 2-7 min. Cells were then removed from the flasks and transferred to a 50 ml tube to be spun at 1500 g for 5 min. Before centrifugation, 10 µl of cells were used on a haemocytometer for counting. Once spun the trypsin was carefully discarded and fresh medium, in the volume of 5 ml, was used to resuspend the cells. Another 5 ml of fresh medium was added to new flasks before the correct volume of cells were added to equal a concentration of 1 x $10^4$ cells per flask.

A total of 32 flasks were prepared consisting of four biological replicates with four flasks for the two conditions (treated and untreated). These four flasks equalled two flasks for each time point; 24 hr and 96 hr. Once the cells were confluent the medium was changed for lower serum medium (DMEM low glucose minus arginine and lysine media (Sigma-Aldrich) including: 0.0159 g/L phenol red, 0.0001g/L proline, 0.105g/L L-leucine, 0.084g/L arginine, 0.146g/L lysine, 0.11g/L sodium pyruvate, 4.5g/L D-glucose, 1 % (v/v) penicillin-streptomyocin, 2% (v/v) horse serum). By lowering the serum the cells were able to differentiate into myotubes. Cells were left for six doublings to become fully differentiated. Medium was changed every two days.

Fresh low serum medium was prepared and either salbutamol at a concentration of 1 x $10^{-4}$, made in DMEM low glucose medium, or the same volume of just DMEM low glucose medium was added to the correct flasks.

#### 5.2.1.3 Harvesting cells

After 24 hr and 96 hr of treatment, cells were harvested from all four biological replicates. This process involves first removing the old medium and washing with 5 ml of HEPES buffer (Sigma-Aldrich). The buffer was then disposed of and 5 ml of trypsin (Sigma-Aldrich) was added and incubated for 4-7 min. Cells were then transferred in to a 50 ml tube and spun at 1500 g for 5 min before carefully discarding the trypsin. To the cell pellet 10 ml of PBS was added and mixed to resuspended the pellet. This was then spun for 5 min at 1500 g at 4°C. The PBS could then be removed and a fresh 1 ml added to transfer the cells to an ependorf tube. This time the cells were spun at 16000 g for 4 min before removing the PBS.

### 5.2.2 Sample preparation

#### 5.2.2.1 Lysis of cells

A 10 mg vial of RapiGest (Waters) was reconstituted in 1ml of 50mM ammonium bicarbonate to give a 1% (w/v) stock solution. A sample of this was taken and diluted down to 0.05% (v/v) in 25mM ambic before 500 µl was added to each sample pellet. Samples were mixed thoroughly before being subjected to five rounds of freeze/thawing with intense vortexing between each round. The samples were then heated for 10 min at 80°C before being spun for 10 min at 16000 g. The supernatant was carefully removed and an aliquot was taken for a protein assay.

#### 5.2.2.2 Protein assay

The Thermo Scientific Pierce Coomassie Plus Protein Assay was used to measure the protein concentration of the 16 samples. A set of protein standards, ranging from 2.5 – 25 µg/ml BSA, were prepared in ultra pure $H_20$. An aliquot of each sample was diluted 1:100 before completion of the assay. Next, 150 µl of each standard and the samples were transferred to different wells on a 96 well plate. An equal volume of Coomassie Plus reagent was added to all the standards and samples before being left to incubate at room temperature for 10 min. To determine the concentration of the samples a Multiskan Ascent microplate photometer (Thermo Scientific) was used to read the absorbance at 595 nm.

#### 5.2.2.3 Digestion of samples

An aliquot of 100µg from each sample was taken for digestion. Extra 25mM ammonium bicarbonate was used to reach a total volume of 160µl. To this 10µl of a 9.2mg/ml DTT solution was added, votexed and incubated at 60°C for 10 min. Next, 10µl of a 33mg/ml IAA solution was added before votexing and incubating at room temperature in the dark for 30 min. Sequence grade trypsin was diluted in 25 mM ammonium bicarbonate to give a concentration of 1 µg/µl before being mixed with the sample in a 50:1 protein:trypsin ratio. This was then incubated overnight at 37 °C. A 1D gel was used to check for complete digestion before the samples were analysed on a mass spectrometer.

## 5.2.3 Protein identification and quantification

### 5.2.3.1 Mass spectrometry analysis

As in chapter 4 (4.2.7.1.2), the LTQ Orbitrap Velos was used to analyse the label free samples. The samples were run by Dr Dong Xia, University of Liverpool.

### 5.2.3.2 MaxQuant

Raw files were processed using version 1.1.1.36 of MaxQuant (http://www.maxquant.org/). For protein identification the ipi.MOUSE protein database v3.87 (http://www.ebi.ac.uk/IPI/IPImouse.html) was uploaded to MaxQuant. Carbamidomethylation was set as fixed modification and variable modifications as oxidation (M). Initial peptide mass tolerance was set to 20 parts per million (ppm) and fragment mass tolerance was set to 0.8 Da. Two unique peptides were required for high-confidence protein identifications. The peptide and protein false discovery rates (FDR) were set to 0.01. The maximal posterior error probability (PEP), which is the probability of each peptide to be a false hit considering identification score and peptide length, was set to 1. Raw files were uploaded separately but all data was retrieved in one file to compare the intensities for proteins identified. All proteins identified with only 1 unique peptide were removed. The intensities for each protein were averaged across the four repeats before the fold change was determined. A test was applied to all four repeats at both 24 hr and 96 hr. Normalisation was performed by MaxQuant but this was manual checked and each treated result normalised to the average.

## 5.3 Results

### 5.3.1 Identifying proteins of interest

MaxQuant produced a table of results including the intensities for each peptide detected from all samples at either 24 hr or 96 hr. By taking the sum of all intensities across the four replicates for each condition at 96 hr for a particular peptide the overall fold change could be determined. To determine if any normalisation needed to occur prior to the MaxQuant statistics the log ratios of each peptide were plotted as a scatter graph (Figure 5.2). From the graph it can be seen that the original intensities require little normalisation due to the mean peptide log ratios lying across 0.

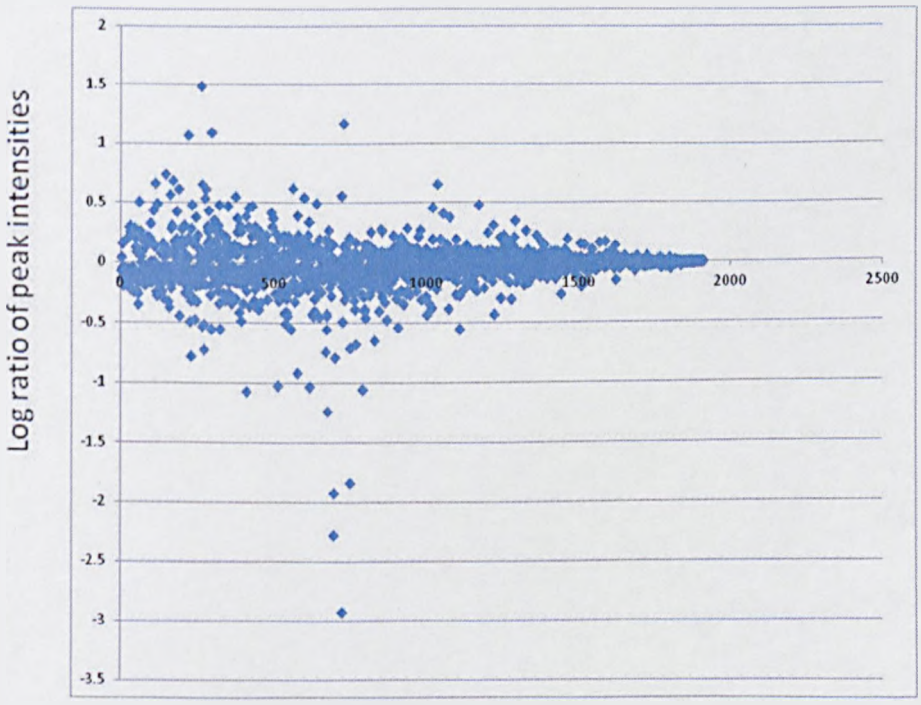



**Figure 5.2** ***Scatter graph visualising the log ratios of each peptide identified in the control and treated samples.*** MaxQuant was utilised to identify the intensity of peak height for each peptide identified in the LTQ-Velos spectra. The total change in intensity for each peptide across four biological replicates was determined and the log of this ratio plotted. The log ratio is plot along the y axis with the peptide number in order of identification along the x axis.

Using the peak intensity measurements MaxQuant produced a list of identified proteins. From this list the mean intensity for each protein was determined and the corresponding fold change between treated and control samples was determined. At 24 hr a total of 841 proteins were identified with a fold change $\geq 1.5$ but at 96 hr a total of 316 proteins had a fold change $\geq 1.5$. However, these identified proteins were often only seen in one of the four repeats. To determine how statistically significant these changes are a t-test was performed using the data from all four replicates. This reduced the number of identifications greatly to only 14 differentially expressed proteins at 24 hr and 19 proteins at 96 hr with a T value $\leq 0.01$ (Table 5.1 and 5.2). At 24 hr, six of the proteins identified were seen to be up-regulated whereas at 96 hr only three were up-regulated. Although none of the proteins identified at 24 hr were identified as modulated at 96 hr several of the proteins identified at both time points have similar biological processes. All of the process identified would have a role in cells which are growing or changing. For example, there were a number of proteins involved in transcription and translation such as histone deacetylase 2 and C-terminal-binding protein 1. There were also proteins associated with cytoskeletal organisation and cell-cell adhesion. When these cells differentiate they fuse together changing the cytoskeletal structure meaning proteins like inverted formin-2, vinculin and catenin delta-1 could have a role. Interestingly, guanine nucleotide-binding protein G(s) subunit alpha isoforms Xlas was identified at 96 hr as being marginally down regulated. This protein is involved in the regulation of adenylate cyclase which is activated in response to beta-adrenergic stimuli (Bastepe *et al.*, 2002). Although, these proteins have high t-test scores the fold changes were very small for a number of them with only seven from both time points having fold changes $\geq 1.5$.

**Table 5.1** ***Identified proteins with differential expression between control and treated mouse muscle cells at 24 hr after treatment.*** Spectra from C2C12 cells either untreated or treated with salbutamol were used to perform label free peak intensity analysis using MaxQuant. The intensity of peptide peaks was determined allowing the fold change and T-Test to be determined. Proteins with peptide intensities with a T value $\leq 0.01$ were listed as differentially modulated.

| Protein Identification | Max fold change | Modulation after treatment | T-test |
|---|---|---|---|
| Adenylate kinase isoenzyme 1 | 1.20 | Down regulated | 0.0013 |
| Peroxin-14 | 1.31 | Up regulated | 0.0030 |
| Collagen alpha-1(III) chain | 1.53 | Up regulated | 0.0033 |
| Proteasome subunit alpha type-5 | 1.10 | Up regulated | 0.0045 |
| Procollagen galactosyltransferase 1 | 1.11 | Up regulated | 0.0052 |
| Uncharacterized protein C1orf198 homolog | 1.26 | Down regulated | 0.0065 |
| C33 antigen | 1.16 | Up regulated | 0.0066 |
| Citrate synthase, mitochondrial | 1.25 | Up regulated | 0.0076 |
| C-terminal-binding protein 1 | 1.20 | Down regulated | 0.0081 |
| Vacuolar protein sorting-associated protein 18 homolog | 6.61 | Down regulated | 0.0085 |
| COP9 signalosome complex subunit 4 | 1.52 | Down regulated | 0.0086 |
| Serine-threonine kinase receptor-associated protein | 1.89 | Down regulated | 0.0086 |
| Endoplasmic reticulum-Golgi intermediate compartment protein 1 | 1.09 | Down regulated | 0.0098 |
| DNA methyltransferase 1-associated protein 1 | 1.35 | Down regulated | 0.010 |

**Table 5.2** ***Identified proteins with differential expression between control and treated mouse muscle cells after 96 hr of treatment.*** Spectra from control and salbutamol treated C2C12 cells were used to perform label free peak intensity analysis using MaxQuant. The intensity of peptide peaks was determined allowing the fold change and T-Test to be determined. Proteins with peptide intensities with a T value $\leq 0.01$ were listed as differentially modulated.

| Protein Identification | Max fold change | Modulation after treatment | T-test |
|---|---|---|---|
| rRNA 2'-O-methyltransferase fibrillarin | 1.10 | Up regulated | 0.0003 |
| Talin 1 | 1.16 | Down regulated | 0.0009 |
| Eukaryotic translation elongation factor 1 epsilon-1 | 1.35 | Down regulated | 0.0018 |
| NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial | 1.15 | Down regulated | 0.0019 |
| Epidermal growth factor receptor substrate 15 | 1.45 | Up regulated | 0.0021 |
| Histone H1.2 | 1.10 | Down regulated | 0.0024 |
| Guanine nucleotide-binding protein G(s) subunit alpha isoforms Xlas | 1.09 | Down regulated | 0.0032 |
| Catenin delta-1 | 1.37 | Down regulated | 0.0038 |
| AP-2 complex subunit alpha-1 | 1.57 | Down regulated | 0.0043 |
| Tripeptidyl-peptidase 2 | 1.21 | Down regulated | 0.0047 |
| Fatty acid synthase | 1.14 | Down regulated | 0.0047 |
| Transcriptional activator protein Pur-beta | 1.12 | Down regulated | 0.0053 |
| Calumenin | 1.27 | Down regulated | 0.0055 |
| Vinculin | 1.10 | Down regulated | 0.0057 |
| Prefoldin subunit 6 | 1.27 | Down regulated | 0.0067 |
| Neddylin | 1.42 | Down regulated | 0.0082 |
| Histone deacetylase 2 | 1.31 | Down regulated | 0.0083 |
| Sorting nexin-2 | 1.52 | Down regulated | 0.0090 |
| Inverted formin-2 | 1.60 | Up regulated | 0.0111 |

When the t-test value was lowered to $\leq 0.05$ a much larger number of proteins are identified (64 at 24 hr and 52 at 96 hr) as well as having higher fold changes. A total of 26 proteins from both time points were identified with a fold change $\geq 1.5$. Tables 5.3 and 5.4 list the proteins identified as modulated with the less stringent t-test cut off. Once again a number of proteins identified as cytoskeletal or membrane proteins were identified, such as cytokeratin and troponin. There were also several proteins identified to have a role in protein transport, for example, sorting nexin-2 and monocarboxylate transporter 1. Monocarboxylate transporter 1 (MCT1) had the highest fold change within these proteins identified with a 7.9 fold change increase after 24 hr treatment. This protein is thought to have a role in important metabolic processes including energy metabolism in the brain, skeletal muscle, heart and tumor cells as well as T-lymphocyte activation, gluconeogenesis in the liver and kidney, spermatogenesis, bowel metabolism of short-chain fatty acids, and drug transport. This protein has also been previously identified as being modulated by treatment with another β-agonist, clenbutermol (Hoshino *et al.*, 2012). However, in this study the rat skeletal muscle showed a reduction of MCT1 after treatment. The highest fold change at 96 hr was a 2 fold increase after treatment by motile sperm domain-containing protein 2; this is a relatively small protein with an unknown physiological function. Thaler *et al.*, (2011) recently identified the gene for this protein was strongly expressed in mesenchymal cells with advanced differentiation status when examining mouse tissues. When investigating other cell lines they also identified that *Mospd1* was up-regulated during differentiation in osteoblastic, myoblastic, and adipocytic cell lines. Although, increasing the t-test cut off point identified more proteins with higher fold changes there is still no overlap between proteins from the 24 hr and 96 hr data sets.

**Table 5.3** ***Identified proteins with differential expression between control and treated mouse muscle cells after 24 hr of treatment.*** Spectra from control and salbutamol treated C2C12 cells were used to perform label free peak intensity analysis using MaxQuant. The intensity of peptide peaks was determined allowing the fold change and T-Test to be determined. Proteins with peptide intensities with a T value $\leq 0.05$ and a fold change $\geq 1.5$ were listed as differentially modulated.

| Protein Identifications | Max fold change | Modulation | T-test |
|---|---|---|---|
| Monocarboxylate transporter 1 | 7.93 | Up regulated | 0.0388 |
| Vacuolar protein sorting-associated protein 18 homolog | 6.61 | Down regulated | 0.0085 |
| Structural maintenance of chromosomes protein 3 | 2.86 | Down regulated | 0.0102 |
| Cytochrome P450 2F2 | 2.50 | Up regulated | 0.0255 |
| VPS10 domain-containing receptor SorCS2 | 2.49 | Down regulated | 0.0445 |
| N-acylneuraminate cytidylyltransferase | 2.26 | Up regulated | 0.0188 |
| Myogenin | 2.19 | Up regulated | 0.0443 |
| Oligoribonuclease, mitochondrial | 2.14 | Down regulated | 0.0342 |
| Serine/threonine-protein phosphatase 4 regulatory subunit 3B | 2.13 | Down regulated | 0.0253 |
| Serine-threonine kinase receptor-associated protein | 1.89 | Down regulated | 0.0086 |
| UPF0505 protein C16orf62 homolog | 1.80 | Down regulated | 0.0332 |
| COP9 signalosome complex subunit 3 | 1.68 | Up regulated | 0.0336 |
| Collagen alpha-1(III) chain | 1.53 | Up regulated | 0.0033 |
| COP9 signalosome complex subunit 4 | 1.52 | Down regulated | 0.0086 |
| Interferon induced transmembrane protein 3 | 1.52 | Up regulated | 0.0393 |
| CD63 antigen | 1.52 | Down regulated | 0.0333 |

**Table 5.4** ***Identified proteins with differential expression between control and treated mouse muscle cells after 96 hr of treatment.*** Spectra from control and salbutamol treated C2C12 cells were used to perform label free peak intensity analysis using MaxQuant. The intensity of peptide peaks was determined allowing the fold change and T-Test to be determined. Proteins with peptide intensities with a T value $\leq 0.05$ and a fold change $\geq 1.5$ were listed as differentially modulated.

| Protein Identification | Max fold change | Modulation | T-test |
|---|---|---|---|
| Motile sperm domain-containing protein 2 | 2.01 | Up regulated | 0.0288 |
| Short/branched chain specific acyl-CoA dehydrogenase | 1.96 | Up regulated | 0.0251 |
| Tight junction protein ZO-1 | 1.93 | Up regulated | 0.0446 |
| AMP deaminase 3 | 1.80 | Up regulated | 0.0331 |
| Cytokeratin-1 | 1.73 | Down regulated | 0.0406 |
| Troponin T, slow skeletal muscle | 1.69 | Down regulated | 0.0346 |
| Translationally-controlled tumor protein | 1.68 | Down regulated | 0.0476 |
| Carbonyl reductase [NADPH] 2 | 1.63 | Up regulated | 0.0227 |
| Inverted formin-2 | 1.60 | Up regulated | 0.0111 |
| Sorting and assembly machinery component 50 homolog | 1.58 | Up regulated | 0.0452 |
| AP-2 complex subunit alpha-1 | 1.57 | Down regulated | 0.0043 |
| Sorting nexin-2 | 1.52 | Down regulated | 0.0090 |

### 5.3.2 Comparison to SILAC data

To determine the significance of the overlapping proteins between SILAC and Label free the cover and agreement of the data between the two experiments was investigated. The total number of proteins identified from each technique with a peptide count ≥ 2 was determined. As seen in Figure 5.2 the label free approach was able to identify 1058 more proteins at 24 hr and 937 more proteins at 96 hr than SILAC. A total of 95% of the proteins identified in the SILAC experiement at both time points are also identified in the label free data.

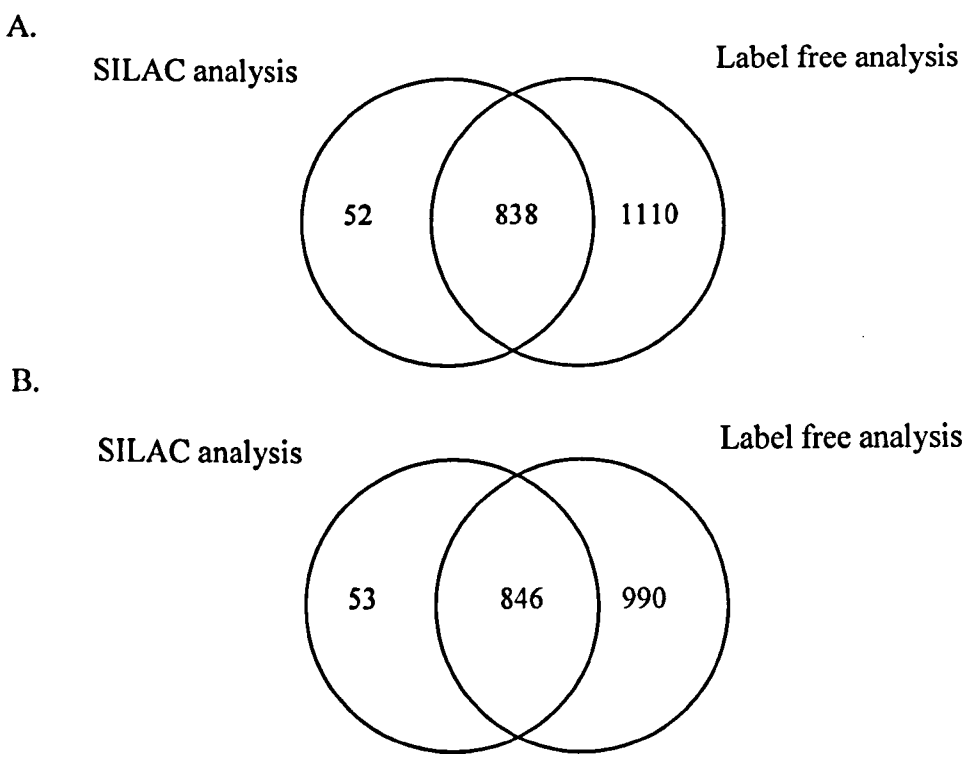



**Figure 5.3** ***Venn diagram of protein identifications made using either SILAC or label free analysis.*** All protein identifications with a unique peptide count of ≥ 2 were identified for both the SILAC (24 hr n=890, 96 hr n=899) and label free (24 hr n= 1948, 96 hr n= 1836) experiments investigating the mode of action of salbutamol on mouse skeletal muscle cells. The identifications from each technique were compared with each other to identify the total number of proteins which overlap. A. Represents the data from the 24 hr data set and B. the 96 hr data set.

## 5.4 Discussion

From this study a small number of proteins were identified as being significantly modulated between control and salbutamol treated mouse muscle cells. When examining those proteins identified with a fold change > 1.2 and a t-test value <0.01 it was noticed that several of the proteins belonged to similar processes at both time points. For example, transcription and protein transport processes both have proteins identified. However, although the processes are similar across the two time points no exact protein was identified as being modulated at both time points. This suggests that different areas within the same process may be activated over time.

Although a small number of proteins have been determined as modulated it is very possible that these proteins have an important role in the growth and change of the muscle cells after treatment with a β-agonist. For example, inverted formin-2 which was identified as up-regulated at 96 hr and may have a role in the changes occurring in the cytoskeleton of the muscle cells. Formins are actin-binding proteins that have multiple effects on actin dynamics and microtubules (Chhabra & Higgs, 2006; Paul & Pollard, 2009). Actin filaments and microtubules have important roles within a number of different cellular processes, including cell migration, membrane transport, and cell division (Rodriguez *et al.*, 2003). Young *et al.*, (2008) suggested that inverted formin-2 (INF1) is the only formin family identified which is likely to act primarily through direct effects on the microtubule cytoskeleton. INF1 has been identified as being expressed in ventricular muscle of the heart and is important in modulating the progression of cardiac hypertrophy (Cooper, 2006). This could suggest that it is may have an important role in skeletal muscle hypertrophy.

Another protein identified as up-regulated at 96 hr was epidermal growth factor receptor substrate 15 (Eps15). Eps15 is an endocytic adaptor protein that is known to support clathrin-mediated internalization of epidermal growth factor receptor (EGFR) at the plasma membrane. It is also known to have a function in endocytosis at the plasma membrane and in downstream trafficking (Carbone *et al.*, 1997). During the development and growth of skeletal muscle fusion of myoblasts occur. During fusion a protein, myoferlin, is highly expressed and is necessary for efficient myoblast fusion to myotubes. Doherty *et al.*, (2008) found that myoferlin binds directly to the eps15 homology domain protein. This suggests that this protein along with myofelin, indirectly regulates the re-organization of the cytoskeleton that accompanies myoblast and myotube fusion. Although myoferlin was not identified as being significantly different it was identified at both time points with a slight increase in expression after treatment.

Catenin delta-1 was also identified at 96 hr but was noted as down-regulated. This protein is also thought to have a role within myogenesis. Catenin delta-1, also known as p120 catenin, is responsible for cadherin stability in the plasma membrane. Cadherin mediates homophilic cell-cell adhesion through their extracellular domain, wheras their intracellular domain associates with catenins that produce attachment sites for the F-actin cytoskeleton (Drees *et al.*, 2005; Yamada *et al.*, 2005). When catenin delta-1 is absence most cadherins are internalised and often degraded, which suggests that they have a role in cadherin turnover at the cell surface (Davis *et al.*, 2003; Davis *et al.*, 1990; Xiao *et al.*, 2003). As this protein was identified as being down-regulated it suggests that cell-cell adhesion may be being reduced or stopped. As myogenesis occurs in the C2C12 cells during differentiation the reduction of cell

fusion may indicate that other process are involved in the increase in growth of the cells rather than further fusion.

One other interesting protein identified at 96 hr was histone deacetylase 2, a key protein in gene transcription and protein function through its ability to alter the acetylation status of histones or other proteins. The removal of acetyl groups from histone tails by histone deacetylase 2 promotes transcriptional repression by allowing chromatin compaction (Haberland *et al.*, 2009). As this protein was seen to be down-regulated in this study it suggests that an increase in transcription may be occurring. Moresi *et al.*, (2012) also showed that histone deacetylase 2 controls muscle homeostasis and autophagy flux in mouse skeletal muscle.

At 24 hr there are also proteins involved in transcription which are identified as being modulated. One protein which has a t-test value $\leq 0.01$ and a fold change $> 1.2$ is DNA methyltransferase 1-associated protein 1 (Dnmt1). DNA methylation is a post-replicative, covalent modification of DNA that is implicated in growth homeostasis and differentiation (Li *et al.*, 1992). Dnmt1 has been identified as associating with replication sites by directly binding to proliferating cell nuclear antigen which helps maintain DNA methylation patterns in the newly synthesized strand after DNA replication (Chuang *et al.*, 1997; Leonhardt *et al.*, 1992). Mortusewicz *et al.*, (2005) has also suggested that Dnmt1 has a direct role in the restoration of epigenetic information during DNA repair. It is interesting that this protein was identified as being down-regulated within this study. This suggests that methylation is not occurring which is surprising in a cell which is growing and changing.

At 24 hr the protein with the highest fold change and a t-test value < 0.01 is Vacuolar protein sorting-associated protein 18 homolog. This protein is thought to have a role in vacuole transportation. Robinson *et al.*, (1991) proved that the *VPS18* gene is necessary for the sorting and processing of soluble and membrane-associated vacuolar hydrolases. However, Yogosawa *et al.*, (2005) identified human vacuolar protein sorting-associated protein 18 as an ubiquitin ligase which can induce the ubiquitylation and degradation of Serum-inducible kinase (SNK). SNK is a member of the polo-like kinases which act as regulators for several events during cell division (Barr *et al.*, 2004). This protein is seen to have a six fold decrease within mouse muscle cells after treatment with salbutamol. This suggests that cell division may be occurring due to the lack of degradation on SNK which would occur with limited vacuolar protein sorting-associated protein 18.

In 1998, Datta *et al.*, identified a novel WD domain-containing protein, called serine-threonine kinase receptor-associated protein (STRAP), which was identified as interacting with transforming growth factor-β1 (TGF-β1) in a yeast two-hybrid system. In this study STRAP was identified as being down-regulated at 24 hr. Although this protein has yet to be identified with β-agonist treatment, TGF-β1 has been determined to have role in hypotrophy within mouse heart muscle after a transgenic mouse model that over expresses a mature form of TGF-β1 was treated with a β-adrenoceptor blocker (Huntgeburth *et al.*, 2011). Although TGF-β1 was not identified as being modulated the down-regulation of STRAP would eventually have an effect on the workings of this protein which may be responsible for the hypertrophy occurring with the C2C12 cells.

### 5.4.1 Comparison to SILAC data

When the proteins identified as modulated were compared to the proteins modulated in the SILAC experiment, only histone H1.2 and C33 antigen were identified as being present in both experiments. Histone H1.2 was identified as up-regulated at 24 hr in the SILAC experiment but down-regulated at 96 hr in the label free experiment. Whereas, C33 antigen was observed as being down-regulated at 96 hr in the SILAC experiment and up-regulated at 24 hr in the label free experiment. This suggests that not only are different mechanisms at work at different times but that the two proteomic techniques are able to identify different proteins with each repeat of the experiment. Histones are known to be involved in DNA structure (Littau *et al.*, 1965) and are vital to chromatin structure in mammals (Fan *et al.*, 2005); histone H1 has been identified as a fundamental part of chromatin and located at or near the entry/exit sites of DNA on the nucleosome (Huntgeburth *et al.*, 2011). During the cell cycle, chromatin undergoes major structural changes, and the participation and location of H1 is thought to be highly important. As this protein was identified in two experiments it is most likely it has an important role in the growth of muscle cells after salbutamol treatment. As histone deacetylase 2 was also identified as down-regulate in this experiment it could indicate that transcription is taking place after treatment. Histone deacetylase has been witnessed as being induced by heat shock protein 70 in cardiac hypertrophy in mice (Kee *et al.*, 2008). In the SILAC study, Hsp70/Hsp90-organizing protein was considered to be up-regulated at 96 hr. This protein reversibly links together the protein chaperones Hsp70 and Hsp90. This suggested that although Hsp 70 was not identified as modulated in either experiment, it may have been present but not identified and be involved in the process of hypertrophy. When analysing overlapping proteins the correlation between total data

muat be taken into account. Label free analysis was able to identify over twice the amount of protein as SILAC with 95% of the proteins from SILAC being found in the label free data. This suggests the the correlation of identifications is high but label free has a greater capacity for identifications. However, even though nearly all the proteins in the SILAC experiment are identified in the label free investigation only two modulated genes were obtained from both. This highlights a potential limitation in these techniques when identify peptide pairs and therefore modulation.

Whereas histone H1.2 is involved in transcription, C33 antigen, also known as CD82 antigen, has roles within cell-cell adhesion (Abe *et al.*, 2008), cell motility (Ono *et al.*, 1999) and apoptosis (Schoenfeld *et al.*, 2004). It has also been reported that CD82 has a role in attenuating epidermal growth factor receptor signalling by encouraging the internalisation of the activated receptor. Although the epidermal growth factor receptor protein was not identified in either study, the epidermal growth factor receptor substrate 15 protein was identified as up-regulated at 96 hr in this study. It is possible that these proteins are working within the same pathway to allow cell-cell adhesion or fusion.

Although there were limited numbers of exact corresponding proteins, there were many proteins identified in both experiments with similar functions. For example, there were several proteins involved in the metabolic processes surrounding acetyl-CoA and fatty acid metabolism such as citrate synthase and fatty acid synthase identified using label free as well as 3-ketoacyl-CoA thiolase and long chain 3-hydroxyacyl-CoA dehydrogenase identified using SILAC. There were also a number of proteins which are involved in the transportation of proteins. Proteins such as

vacuolar protein sorting-associated protein 18 homolog are associated with endocytosis, whereas ADP ribosylation factor is a regulator of vesicle biogenesis in intracellular traffic. Both experimental procedures observed a number of different proteins involved in the regulation of transcription and translation. For example, heterogeneous nuclear ribonucleoprotein A0 was determined to be down-regulated at 96 hr in the SILAC study and DNA methyltransferase 1-associated protein 1 was identified as down-regulated at 24 hr in the label free analysis.

There are several different proteins but with similar functions highlighted over the two experiments which are involved in the mechanisms of a cell undergoing a change in formation. As both experiments only identify a small number of modulating proteins it could be concluded that the treatment of salbutamol only effected a small change on the proteome. However, both experiments have high criteria to be determined as statically differentiating. In the SILAC experiment the protein must be identified in all four of the replicates and have a fold change $\geq 1.5$ whereas the label free study used a t-test value $\leq 0.01$ and the protein must appear in three of the four replicates with intensity values $> 0$ for both control and treated samples. It is possible that the protein changes were marginal which means they were not detected by these techniques as when the criteria was loosened a larger number of proteins were identified.

### 5.4.2 Limitations of label free analysis

Like any other proteomic method there are limitations to this experimental design. As with the SILAC experiment only the soluble proteins have been analyzed leaving a portion of the proteome undetected. The use of RapiGest as a lysis buffer lowers

the potential for loss of other proteins due to the removal of any clean up steps. The majority of limitations for this technique occur during the mass spectrometry and data analysis steps.

One presumption made when using label free techniques to relatively quantify proteins is that known peptides in multiple different runs will be located in chromatographic and MS spectra with narrow fraction tolerances. If peptide signals are spread over a large retention time range it can cause an overlap with co-eluting peptides resulting in a difficulty to ascertain the peak intensities. It also assumes there are no significant differences in sample preparation and content, which could result in multiple signals for the same peptide, as well as assuming there are no technical variations in retention time and MS intensity. There must also be a lack of different sample background noise from chemical interference in each sample. In this study a small amount of background noise was detected which may have had an effect on the results. Old *et al.*, (2005) proved that often peptides will not behave as anticipated by showing that in one experiment 90% of the total peptides eluted in a single strong cation-exchange (SCX) chromatography fraction. Any experimental drifts in retention time and m/z will affect accurate comparisons of the multiple LC-MS datasets. Unaligned peaks during comparison steps can also result in large variability and inaccuracy in quantification (Zhu *et al.*, 2010). As MaxQuant performs the peak alignment as part of the processing of the data it is not known without manually checking if the peaks aligned well. It is highly important that reproducible LC-MS and careful chromatographic peak alignment are undertaken when using label free techniques. There is then the challenge of normalizing data from multiple samples across multiple runs, which is discussed in Chapter 7.

Another limiting factor for quantitation by peak area intensity measurements is the requirement for obtaining assigned MS/MS spectra prior to peak matching. This could reduce the number of comparable proteins if the reacquisition of MS/MS spectra between experiments is low. This would be most problematic for ions with weak intensities. Label free analysis depends greatly on the depth of MS/MS data due to ratios from peak area intensity being most significant for proteins with a larger number of overlapping peptide ions (Porteus *et al.*, 2011). Protein ratios estimated from peak area intensities can sometimes be limited to abundant proteins with high sequence coverage which can bias the results.

There is also a need for reducing potential interferences that could lead to the suppression of protein identifications (Coombs, 2011). This can often lead to an increased need for high-resolution or multidimensional chromatography which would help reduce suppression effects, as well as allowing the detection of low-abundance proteins. However, a concern is that the suppression effects may vary between the samples being compared, which could adversely affect the results.

One last problem involves the decision of which peptides to use for identification. When analyzing the spectra the list of peptides is separated into either shared peptides or unique peptides. Shared peptides are peptide sequences that can be matched in a database to more than one protein. Unique peptides are only matched to one exclusive protein, which allows no ambiguity to the assignment of the protein in the database (Nesvizhskii *et al.*, 2003). However, a single gene can result in hundreds of different proteins through splicing variants and post-translational modifications. This can then create indistinguishable protein identifications when dealing with

incomplete sequence coverage (Black, 2000). This means that when using shared peptides for identification it can become unclear which proteins they are actually derived from. In this method, using MaxQuant, it can be seen that there are a number of different proteins identified per result. There is no clear cut way to deal with this problem. Suggested solutions include removing all shared peptides and focus only on unique peptides, or distribute peptides across the homologous proteins (Neilson *et al.*, 2011). Although analysing unique peptides results in unambiguous identifications it reduces the probability of obtaining data for those proteins with only a few unique peptides.

### 5.4.3 Overall conclusions

Overall, a small number of proteins have been identified with changing abundances between control and treated mouse muscle cells. The functions of the proteins identified correlate with the SILAC data suggesting these proteins or proteins involved in these processes may be changing. However, for this cell type, when treated with a steroid that can often cause a great hypertrophic response, it was expected that there would be a much greater change within the proteome. This, along with the transcriptomic data (Chapter 3) suggests that a full understanding of the mechanisms of salbutamol has not been determined through this experiment. However, as both proteomic studies only identify a small number of proteins it is possible that the effects on the proteome at these time points are very minimal but still enough to start the processes of hypertrophy.

Label free quantification is becoming more frequently used as a technique for examining the proteome but it still has disadvantages compared to labelling proteomic methods. It is, however, able to provide identifications for a large number of proteins as well as being used as collaborating data. This technique is very much dependent on stable, reproducible and accurate chromatography platforms, as well as highly sophisticated analyses software for normalisation and peak to peak analysis. Further development in protein extraction methods is essential in order to improve broad spectrum detection of proteins at any one time and to capture those proteins which often have low expression levels or are hard to solubilise.

# 6 Chapter 6

# A technical comparison of the transcriptomic and proteomic methodologies, data and data analysis software

## 6.1 Introduction

A current problem for biologists, when attempting to study the cellular biology of many cell types, is which technique to employ to allow the most in-depth analysis of the proteome and/or transcriptome. In this study four different techniques; DIGE, SILAC, label free peak intensities and SOLiD sequencing, were utilized to identify the effects of salbutamol, a beta agonist, on the cellular biology within muscle cells. The decision to use a variety of techniques was based on the idea that this would enable a comprehensive study of the workings of salbutamol. If only one technique is used it may limit findings, resulting in an incomplete picture of how tissues and cells truly function. As it can be seen from Table 6.1 there are advantages and disadvantages to each of the technique used. Eukaryotic cells, for example skeletal muscle cells, contain tens of thousands of genes and their protein products which can be modified in a number of different ways including, but not limited to, phosphorylation, glycosylation, acetylation and methylation. This makes investigating the cellular biology of any eukaryotic cell a daunting task. Not only are there several different levels within the proteome to consider, understanding the link between the transcriptome and the proteome is fundamental to fully grasping the complex processes occurring within a cell. Both transcriptomics and proteomics are essential for systems biology, and without techniques that can perform analysis that span the entire genome and proteome it becomes difficult to understand the complete system.

**Table 6.1** ***Advantages and disadvantages of all techniques used.***

| Technique | Advantages | Disadvantages |
|---|---|---|
| DIGE | Identification of protein modulation;<br>No clean up prior to separation needed;<br>Can compare more than one condition in one experiment;<br>Well established technique; | Expensive Dyes;<br>Protein identification can be difficult;<br>Labour intensive;<br>Inability to analyse hydrophobic and less abundant proteins; |
| SILAC Labelling | Can compare several samples in the same mass spectrometry run;<br>MaxQuant/Progenesis can undertake analysis;<br>Quantitative protein analysis;<br>Uniformed labelling to allow for several peptides of the same protein to be compared; | Isotopic labels are expensive;<br>Depending on lyses method extensive clean up needed;<br>Bioinformatical knowledge needed to understand normalisation methods;<br>Specialised media needed;<br>Sample is more complex when analysing through mass spectrometer;<br>Any incomplete labelling can effect results; |
| Label Free | No specialised media needed;<br>No expensive labels needed;<br>No complexity issues when analysing using a mass spectrometer;<br>Protein quantification;<br>No limit on number of conditions compared;<br>Less time consuming; | Results affected by experimental and instrumental reducibility;<br>Bioinformatical knowledge needed for comprehensive analysis;<br>Reliant on software for analysis; |
| SOLiD sequencing | RNA relative quantification;<br>Large amount of software for function and pathway analysis; | With no bar-coding statistical analysis is difficult;<br>Expensive;<br>Cannot identify any post-translational modifications; |

However, deciding which technique to use is not the only problem facing biologists but also how to analysis the large quantities of data produced by the proteomic and transcriptomic processes currently available. There are a number of freely available as well as commercial software packages that will analyse the data and perform statistical analyse, but it is unclear if using different packages can cause bias within the experimental data. Once the data has been processed biologists usually have a list of modulated proteins or genes. There are several packages available for further analysing these lists, for example, pathway analysis software to visualise how the proteins/genes are related to each other. Again a number of options are available, each allowing different processes to be undertaken.

This chapter aims to compare how adequately different proteomic and transcriptomic techniques compare to each other by analysing the quantity of data produced as well as identifying any differences in proteins or their respective genes listed across the different technologies. Only the proteomic SILAC and label free (Chapters 4 and 5) and the transcriptomic SOLiD data (Chapter 3) will be compared as the same cell line and biological method were used, which involved collecting control or salbutamol treated mouse skeletal muscle cell samples. A second aim is to compare how different software packages correlate when identifying the intensities of different peptides and which proteins are identified as being modulated.

## 6.2 Material and Methods

### 6.2.1 Acquisition of RNA-Seq data and proteomic data

C2C12 muscle cells were treated with salbutamol and collected over different time points before being prepared for transcriptomic or proteomic analysis. The method for the collection, depletion and concentration of RNA is outlined in Chapter 3. The method for collection, lysis, digestion and clean up of proteins for the SILAC investigations is outlined in Chapter 4, whereas the method for label free analysis is in Chapter 5. Once data had been obtained either through sequencing for transcriptomic analysis or mass spectrometry for the proteomic analysis it needed to be processed to identify genes and proteins of interest.

### 6.2.2 EdgeR

EdgeR was used to determine differently expressed genes in the transcriptomic analysis. The control count data for the four time points were used as replicates, allowing each treated sample set to be compared to the four controls. Once the files containing the counts had been uploaded into the software the low counts were removed. Next, the normalisation factors were calculated, followed by estimating the common dispersion. An exact test was performed on this data and a list of differentially expressed genes with a p value $<0.01$ was compiled.

### 6.2.3 MaxQuant

Raw files of Orbitrap data were processed using version 1.1.1.36 of MaxQuant (http://www.maxquant.org/). For protein identification the ipi.MOUSE protein

database v3.87 (http://www.ebi.ac.uk/IPI/IPImouse.html) was uploaded to MaxQuant. Carbamidomethylation was set as fixed modification and variable modifications as oxidation (M). Initial peptide mass tolerance was set to 20ppm and fragment mass tolerance was set to 0.8 Daltons (Da). Two unique peptides per protein were required for high-confidence protein identifications. The peptide and protein false discovery rates (FDR) were set to 0.01. The maximal posterior error probability (PEP), which is the probability of each peptide to be a false hit considering identification score and peptide length, was set to 1. Raw files were uploaded separately but all data was retrieved in one file to compare the fold changes between experimental runs and time points.

### 6.2.4 Progenesis

Raw files from the LTQ-Orbitrap data were processed using Progenesis (version 4 Nonlinear Dynamics, Newcastle upon Tyne, UK). Samples were aligned using an automatic alignment feature and the resulting aggregate spectrum filtered to include +1, +2 and +3 charge states only. The samples were grouped according to experimental conditions (treated or control). Features without MS2 spectra were deleted from the analysis as this information is necessary for peptide/protein identification. An mgf file representing the aggregate spectrum was exported and searched using Mascot. Settings were set at: 1 missed cleavage, fixed modification of Carbamidomethyl (C), variable modifications of Oxidation (M), peptide tolerance of +/-20 ppm and MS/MS tolerance of +/-0.8 Da. The resulting .xml file was re-imported into Progenesis to assign peptides to features, using the following thresholds: Mascot score >40 and hits >2 to ensure only high quality identifications were included.

### 6.2.5 Progenesis Post Processor

Data from the Progenesis analyse was imported for the Post Processor analysis. This was performed by Da, Qi at the University of Liverpool. The following settings were used: a mass difference of 6 Da (with a 0.5 Da window) for lysine and arginine, one miscleavage allowed and a 0.8 min retention time window.

### 6.2.6 Pearson correlation

A Pearson correlation coefficient was performed in the R package to determine the correlation between methods and software. For these comparisons all the intensity data for every protein collected from the two proteomic techniques analysed by the three software packages were compiled into one table. The proteins were linked with their corresponding genes for previous analysis.

## 6.3 Results and Discussion

### 6.3.1 Transcriptomics vs Proteomics

The response of a cell or organism to an environment alteration or challenge is mediated by changes in both gene and protein expression and modification. Therefore, the understanding of how gene and protein expression and modification are quantitatively and qualitatively affected by these challenges is of high importance to biologists. However, fully understanding the link between gene expression and protein regulation has still not been achieved. Correlating the two areas of cell biology becomes difficult due to post-transcriptional gene modifications, protein turnover, specific proteolyic processing and the different stabilities and biological half-lives of RNA and proteins.

There have been several studies (Arvas *et al.*, 2011; Barker *et al.*, 2012; Fu *et al.*, 2009; Griffin *et al.*, 2002; Gygi *et al.*, 1999) which have attempted to compare mRNA levels and protein expression over the last decade. Until recently the process most often used to explore the genome was microarrays. Several of these studies show that mRNA levels cannot be consistently relied upon to predict protein abundance, though the correlation seems to be higher when using more recent technologies, such as RNA-seq. For example, Fu *et al.,* (2009) identified a Pearson correlation of $r = 0.24$ between protein and mRNA expression levels when using microarray measurements, however when using RNA-Seq expression data for the same set of genes the correlation is significantly higher (Pearson correlation, $r = 0.36$).

Gygi *et al.,* (1999) examined the relationship between mRNA and protein expression levels for selected genes expressed in the yeast *Saccharomyces cerevisiae. They used* two-dimensional gel electrophoresis followed by LC-MS/MS to identify proteins and quantified the proteins using metabolic labelling and scintillation counting. The corresponding mRNA levels were determined from serial analysis of gene expression (SAGE) frequency tables (Velculescu *et al.*, 1995; Velculescu *et al.*, 1997). They were unable to confidently correlate between mRNA and protein levels, resulting in the inability to predict protein expression levels from quantitative mRNA data. Also, when Griffin *et al.,* (2002) investigated the effects of carbon source perturbation on *Saccharomyces cerevisiae* they concluded that the insights gained into the perturbative effects could not have been identified using only messenger RNA measurements or protein levels, confirming the need to integrate different types of data. However, Schmidt *et al.,* (2007) profiled fission yeast *Schizosaccharomyces pombe* and was able to achieve a relatively good correlation between proteomic and transcriptomic data. When comparing 1367 protein–mRNA pairs they achieved a Pearson correlation coefficient of 0.58. When functional pathway analysis was applied a strong correlation was seen for those proteins involved in signalling and metabolic processes, but for components of protein complexes there was large variance. Kislinger *et al.,* (2006) also showed concordance between tissue expression patterns of gene products and proteomic profiling. A total of 1758 proteins were mapped with only 503 pairs of gene products identified as being statistically significantly different.

There are several reasons explaining the differences seen between mRNA data and protein expression. On a biological front these differences could result from RNA

splicing that is not detectable by the microarray platform often previously used, differential RNA and protein turnover, post-translational modifications and proteolytic processing events. However, they may also arise from the experimental method, for example, challenges can occur in data interpretation, as well as technological limitations.

With transcriptomic techniques constantly improving, providing a robust, high-throughput technology capable of quantifying tens of thousands of mRNA species, it is hoped that the depth of coverage will increase helping to bridge the gap across genomics and proteomics. It is clear that one area of research is only allowing part the picture to be gained. In this study, two proteomic techniques were compared to transcriptomic data to gain an insight into the biological effect salbutamol produces on mouse muscle cells.

#### 6.3.1.1 Comparison of differentially modulated genes and proteins across data sets

A total of 2600 and 1873 genes were identified using edgeR (Chapter 3) as being differentially expressed at 24 hr and 96 hr respectively after treatment with salbutamol. Whereas, in the SILAC experiment a total of 43 and 33 proteins were determined to be significantly modulated at 24 hr and 96 hr (Chapter 4). In the label free study a total of 14 and 19 proteins were identified as being significantly modulated between control and treated cells after 24 hr and 96 hr respectively (Chapter 5). From these data sets it can already be seen there is a large discrepancy between the three protocols. Table 6.2 illustrates the differences in numbers of proteins and genes identified across all the techniques used as well as the differences

between number of proteins/genes identified and the number modulated. As the DIGE experiment is not directly comparable with the other techniques the focus of the discussion will be on the remaining three techniques.

**Table 6.2** ***Comparison of four techniques relating to the number of proteins/genes identified and number identified as modulated.*** For a protein to be classed as identified for SILAC and Label free the protein must have >2 unique peptides. From SOLiD sequencing the gene must have a count >2 in both the control and treated samples. For the protein to be classed as modulated in DIGE, SILAC and Label free it must have a fold change >1.5. For SOLiD sequencing the genes must have a read count >50 with a fold change >2.

| | Technique | | | |
|---|---|---|---|---|
| | **DIGE** | **SILAC** | **Label Free** | **SOLiD Sequencing** |
| No. of identifications made at 24 hr in two or more replicates | N/A | 890 | 1948 | 9062 |
| No. of modulated proteins/genes identified at 24 hr in two or more replicates | 31 (17 spots) | 43 | 14 | 2600 |
| No. of identifications made at 96 hr in two or more replicates | N/A | 899 | 1836 | 10000 |
| No. of modulated proteins/genes identified at 96 hr in two or more replicates | 110 (37 spots) | 33 | 19 | 1873 |

From the identified proteins found using the label free technique only two of the proteins had corresponding genes identified as being differentially expressed in the transcriptomic experiment, whereas the SILAC experiment had ten protein-gene pairs (Table 6.3). For several of the genes, the time points at which the protein products were identified in the proteomic techniques did not correlate with the time points the genes were expressed. This is not unusual as these gene may have only just been expressed, meaning the protein product may not be detected yet as the protein had not had time to be fully translated, transcribed and possibly modified.

**Table 6.3** ***Genes identified as being modulated in either the SILAC experiment or the label free experiment.*** The corresponding genes for each protein which were determined as significantly modulated between control and treated cells after 24 hr or 96 hr in the SILAC and label free experiments were compared to the list of genes differentially expressed determined using RNA-seq. The time point the genes were identified in are shown, if a gene was not identified in either time point then a – is present.

| Genes | Time points (hr) | | |
|---|---|---|---|
| | Transcriptomics | SILAC | Label free |
| Tubb2a | 24 | 24 | - |
| Hmgn2 | 24, 96 | 24 | - |
| Rragd | 24 | 24 | - |
| Hadhb | 96 | 24 | - |
| Prkag2 | 24,96 | 24 | - |
| Cyp2f2 | 96 | 24 | - |
| Tufm | 24,96 | 24 | - |
| Cacybp | 24 | 96 | - |
| Psmb3 | 24 | 96 | - |
| Psmc3 | 24 | 96 | - |
| Pex14 | 24 | - | 24 |
| Eps15 | 24 | - | 96 |

As the amount of expression of each gene does not necessarily correlate with the amount of expression of protein product, further analysis of the data were undertaken. The highly abundant genes from the transcriptome data were compared with the SILAC and label free data to assess if they were first identified and, secondly, if there was any minimal changes which were too small to be identified in the proteomics. All genes identified with a fold change >2 and read counts over 1000 for both conditions were gathered for both 24 and 96 hr. This list of genes from the transcriptomic data was compared with the genes associated with the proteins identified in the two proteomic studies; the top 15 genes with values from one or both proteomic methodologies, was compiled for both time points. The fold change was compared for each method, seen in Table 6.4 A and B.

**Table 6.4** ***Comparison of gene fold change differences between treated and control samples from three different methodologies.*** Genes identified with read counts over 1000 in both conditions and a fold change >2 are matched with the corresponding proteins identified in the SILAC and label free experiments under the same conditions. Fold changes between treated and control samples plus standard deviations are shown. Table A is a list compiled from genes identified as being modulated at 24 hr and Table B those from 96 hr after treatment. A – represents those genes that were not identified in an experiment. If a 0 is shown, this depicts those genes that are identified but with only 1 peptide and therefore MaxQuant did not perform any further analysis. Arrows highlighted in red represent genes which have been up-regulated and those in blue down-regulated. Genes highlighted are those which are found to be modulated in the same direction in all three experiements.

**A**

| Genes | Fold change (treated/control) | | | | | |
|---|---|---|---|---|---|---|
| | RNA-Seq | | SILAC | | Label free | |
| Lrp1 | 3.11 | ↑ | 0.97 ± 0.05 | ↓ | 1.13 ± 0.09 | ↑ |
| Tm9sf3 | 0.45 | ↓ | - | | 1.03 ± 0.23 | ↑ |
| Hsp90b1 | 0.40 | ↓ | 0.94 ± 0.07 | ↓ | 0.99 ± 0.08 | ↓ |
| Sept02 | 0.44 | ↓ | 1.01 ± 0.02 | ↑ | 0.98 ± 0.07 | ↓ |
| Pdia3 | 0.36 | ↓ | 0.96 ± 0.09 | ↓ | 0.96 ± 0.09 | ↓ |
| Csde1 | 0.47 | ↓ | 0.85 ± 0.00 | ↓ | 0.95 ± 0.40 | ↓ |
| Eno3 | 0.35 | ↓ | 1.13 ± 0.16 | ↑ | 0.94 ± 0.10 | ↓ |
| Anxa2 | 0.43 | ↓ | 1.00 ± 0.04 | ↑ | 0.93 ± 0.10 | ↓ |
| S100a6 | 0.47 | ↓ | 0.99 ± 0.03 | ↓ | 0.89 ± 0.14 | ↓ |
| Kif5b | 0.50 | ↓ | 1.03 ± 0.22 | ↑ | 0.89 ± 0.11 | ↓ |
| Tnnc1 | 0.44 | ↓ | 0.98 ± 0.09 | ↓ | 0.89 ± 0.21 | ↓ |
| Pls3 | 0.39 | ↓ | 0.99 ± 0.07 | ↓ | 0.87 ± 0.18 | ↓ |
| Tpm1 | 0.43 | ↓ | 0.98 ± 0.04 | ↓ | 0.84 ± 0.31 | ↓ |
| Cnn3 | 0.35 | ↓ | 0.90 ± 0.05 | ↓ | 0.84 ± 0.58 | ↓ |
| Eef1a1 | 0.37 | ↓ | 0.98 ± 0.17 | ↓ | 0.84 ± 0.21 | ↓ |

**B**

| Genes | Fold change (treated/control) | | | | | |
|---|---|---|---|---|---|---|
| | RNA-Seq | | SILAC | | Label free | |
| Nt5dc3 | 2.14 | ↑ | - | | 5.58 ± 0.80 | ↑ |
| Eif2c2 | 2.35 | ↑ | - | | 1.31 ± 1.37 | ↑ |
| Bgn | 3.08 | ↑ | 1.02 ± 0.12 | ↑ | 1.18 ± 1.25 | ↑ |
| Ctnnb1 | 2.33 | ↑ | 1.08 ± 0.19 | ↑ | 1.06 ± 0.05 | ↑ |
| Fech | 2.15 | ↑ | 0 | | 1.04 ± 0.26 | ↑ |
| Copa | 2.49 | ↑ | 1.06 ± 0.06 | ↑ | 1.03 ± 0.11 | ↑ |
| Dlgap4 | 2.16 | ↑ | 0 | | 0.97 ± 0.66 | ↓ |
| Sgta | 2.29 | ↑ | 0 | | 0.93 ± 0.37 | ↓ |
| Synpo2 | 2.76 | ↑ | 0 | | 0.93 ± 0.58 | ↓ |
| Ywhag | 2.11 | ↑ | 1.03 ± 0.02 | ↑ | 0.93 ± 0.10 | ↓ |
| Actr1a | 2.41 | ↑ | 0.10 ± 0.02 | ↓ | 0.91 ± 0.09 | ↓ |
| Surf4 | 2.59 | ↑ | 1.09 ± 0.07 | ↑ | 0.89 ± 0.98 | ↓ |
| Gna13 | 2.10 | ↑ | - | | 0.85 ± 0.25 | ↓ |
| Ptgs1 | 2.77 | ↑ | - | | 0.84 ± 0.29 | ↓ |
| Asns | 0.47 | ↓ | - | | 0.50 ± 0.43 | ↓ |

From this data it can be seen there is a limited correlation between the fold changes across all three technologies. However, the majority of the genes, although the amount may differ, have correlating regulation. Those that are seen to be down-regulated in the treated cells in the transcriptomic experiment often have fold changes which would represent down-regulation in the proteomics data, though the fold change is so slight that it is undetected as differentially expressed in these experiments. It would seem that the label free data correlates best with the transcriptomic data as five of the genes identified in Table 6.4 A in the SILAC experiment are the reverse of that found in the other techniques. However, in Table 6.4 B for the gene *Nt5dc3* a value for fold change of 5.57 is listed for the label free analysis. This protein product was not presented in Chapter 5 as being significant due to this protein product being removed as no peptides were identified for the control sample. It is assumed that all proteins with no peptides identified, therefore no intensity value, are not a true result rather than this protein has not been expressed and so is removed. However, due to the removal of these proteins part of the biological picture may again be lost as, in this case, the gene was identified in a different experiment. For SILAC there are also assumptions made which effect the data produced. MaxQuant does not calculate ratios for those proteins which do not appear in all the repeats. As only one run is completed on the Orbitrap-Velos per repeat it is unlikely that all the proteins will be identified, especially with a sample mixture which is quite complex. Due to time constraints further runs were not able to be performed. By allowing those proteins which are observed in three out of the four replicates to also be included as differentially expressed a larger number of identifications may have been made which, in turn, may have allowed more correlation between techniques.

Further analysis is undertaken in section 6.2.2 on the data correlation but, on a biological front, it seems that those genes that are being highly regulated may be having a very minimal effect on the proteome. It is possible that gene expression may need to be considerably higher before any changes in the proteome are seen. This may be a further reason for why many studies struggle to correlate the two sets of data.

These findings differ from several other studies undertaken on mouse cells. For example, Drexler *et al.,* (2011) have recently undertaken a study looking at fast and slow muscle fibres in mice on a proteomic and transcriptomic level. The global protein levels of slow soleus and fast extensor digitorum longus (EDL) muscles were obtained using *in vivo* SILAC based on fully lysine-6 labelled mice. These results are compared to the mRNA expression profiles obtained from the same samples. For this study a Pearson correlation coefficient of 0.81 was gained indicating a significant overlap of all detected transcripts and proteins. However, in this study when they identified genes of interest a much less stringent cut off value was used of fold change >1.5 and p value <0.05. This would have allowed for many more identifications to be made and so increasing the chance of correlation. All transcript intensities were compared with mRNA intensities from their protein dataset which suggested that most of the identified proteins were detected from transcripts with higher abundance. Again this shows that several orders of magnitude in sensitivity and dynamic range are needed to detect lower abundant transcripts from skeletal muscle tissue.

### 6.3.2 Pearson correlation between transcriptomic and proteomic data

To determine how well the RNA-Seq SOLiD sequencing data compared to both the label free peak intensity and SILAC proteomic data, a Pearson correlation coefficient performed in the R package, was performed on log ratios of all identified genes and their respective protein products. Log ratios are the log of the total fold change of treated and control intensities for a protein. In regards to two variables, Pearson's correlation coefficient is defined as the covariance of the two variables divided by the product of their standard deviation (Pearson, 1909). For these comparisons all the data collected was used as the identified modulated proteins were limited, making this comparison identification based rather than modulation based. To perform this test all the data from the three experiments and different software results (section 6.3.1.1) were compiled into one table by matching genes with their related protein products (Appendix III, Table 1). Figure 6.1 visualises the correlation between the transcriptomic total log ratios of treated/control for all genes at both 24 and 96 hr compared to the mean log ratios from the four replicates of each identified protein using the SILAC technique. Figure 6.2 displays the correlation between the transcriptomic data and label free data.

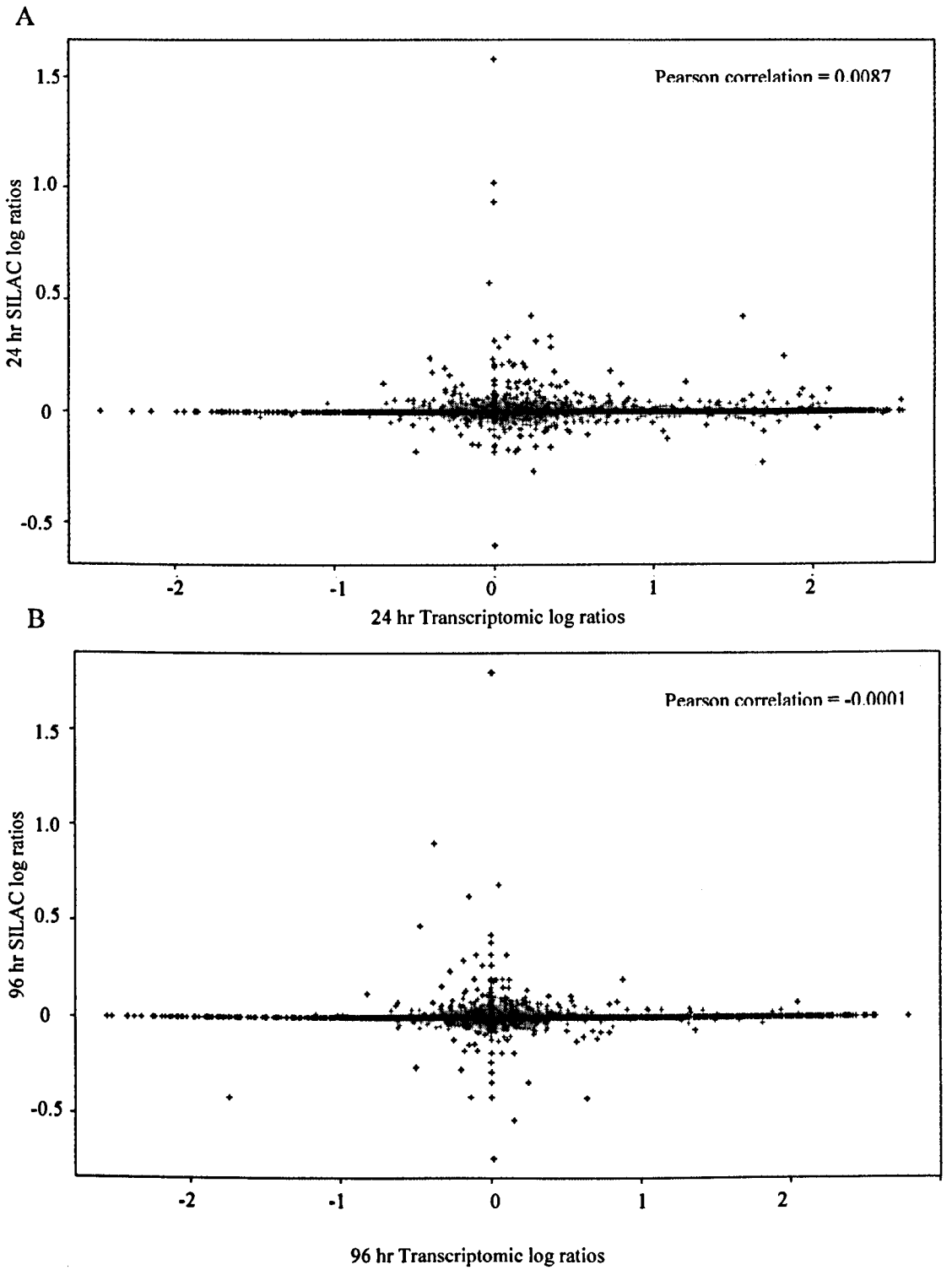



**Figure 6.1** ***Comparison of data from RNA-Seq and SILAC analysis of the effect of salbutamol treatment on mouse muscle cells.*** The genes identified by RNA-Seq SOLiD sequencing were compared with the proteins identified using SILAC, which uses stable isotope labelling to identify peptide pairs and determine modulation, and mass spectrometry. Samples were taken at a) 24 hours and b) 96 hours post treatment with salbutamol. A Pearson correlation was performed on the log ratios to compare the datasets. Log ratios were ascertained by taking the log of the fold change between treated and control intensities or counts for all proteins or genes.

When comparing the transcriptomic data to the SILAC data the Pearson correlation was 0.0087 at 24hr and -0.0001at 96hr. From these values and the graph it can be seen that the correlation between these two methods was almost none (Figure 6.1). It is interesting to note, however, that a large proportion of the genes identified did not have any corresponding protein identifications. This may be due to the complexity of the sample and suggests that even with four replicates, and therefore, four runs on the Orbi-trap, only a small portion of the proteome was detected. A very similar result can be seen when comparing the genomic data with the label free proteome data (Figure 6.2). A correlation of -0.0036 at 24 hr and an identical score for 96 hr was determined for the label free comparison. This indicates that, again, there was very little correlation but any that was seen is negative.

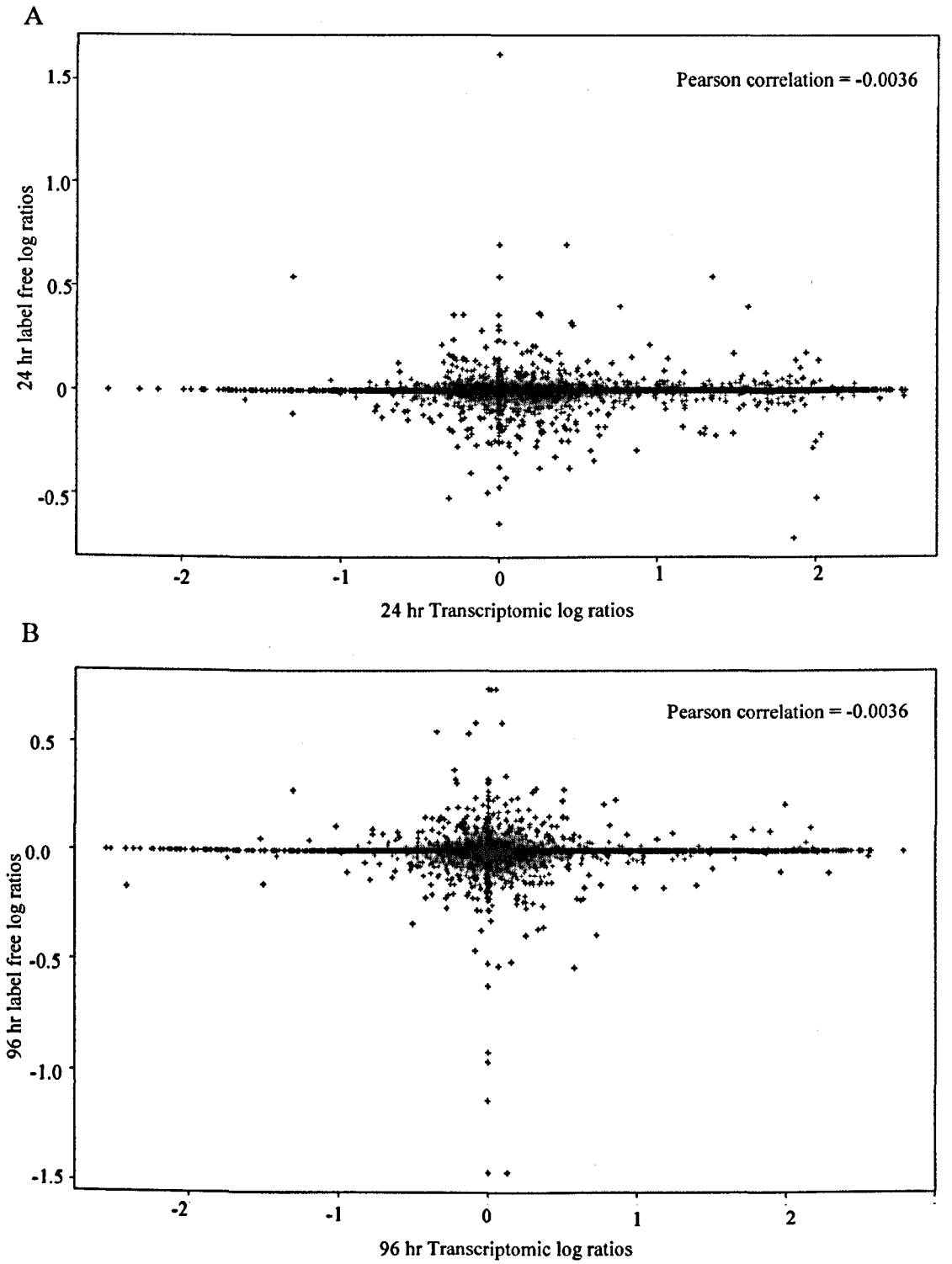



**Figure 6.2** ***Comparison of data from RNA-seq and label free analysis of the effect of salbutamol treatment on mouse muscle cells.*** The genes identified by RNA-Seq SOLiD sequencing were compared with the proteins identified using label free, which uses peak intensity measurements to determine modulation, and mass spectrometry. Samples were taken at a) 24 hours and b) 96 hours post treatment with salbutamol. A Pearson correlation was performed on the log ratios to compare the datasets. Log ratios were ascertained by taking the log of the fold change between treated and control intensities or counts for all proteins or genes.

A large number of the genes or proteins identified in one study were not found in another. This may be due to technical limitations, for example, some genes identified may code for insoluble proteins which were removed before mass spectrometry analysis. Pearson correlations have also been performed on the data sets with only products identified in both the experimental procedures. This produces data sets of several thousand identifications rather than several hundreds of thousands (Appendix III, Tables 2-5). If the techniques are able to identify correlating biological changes between genes and proteins the Pearson correlation should be higher. This would also suggest that a reason for the poor correlations between the techniques is due to their abilities to identify different ranges of proteins and genes. However, when examining Figures 6.3 and 6.4 it is clear that even the data collected does not correlate well. With further work investigating the absolute abundance of any modulated gene and its corresponding protein would give further insight into this hypothesis.

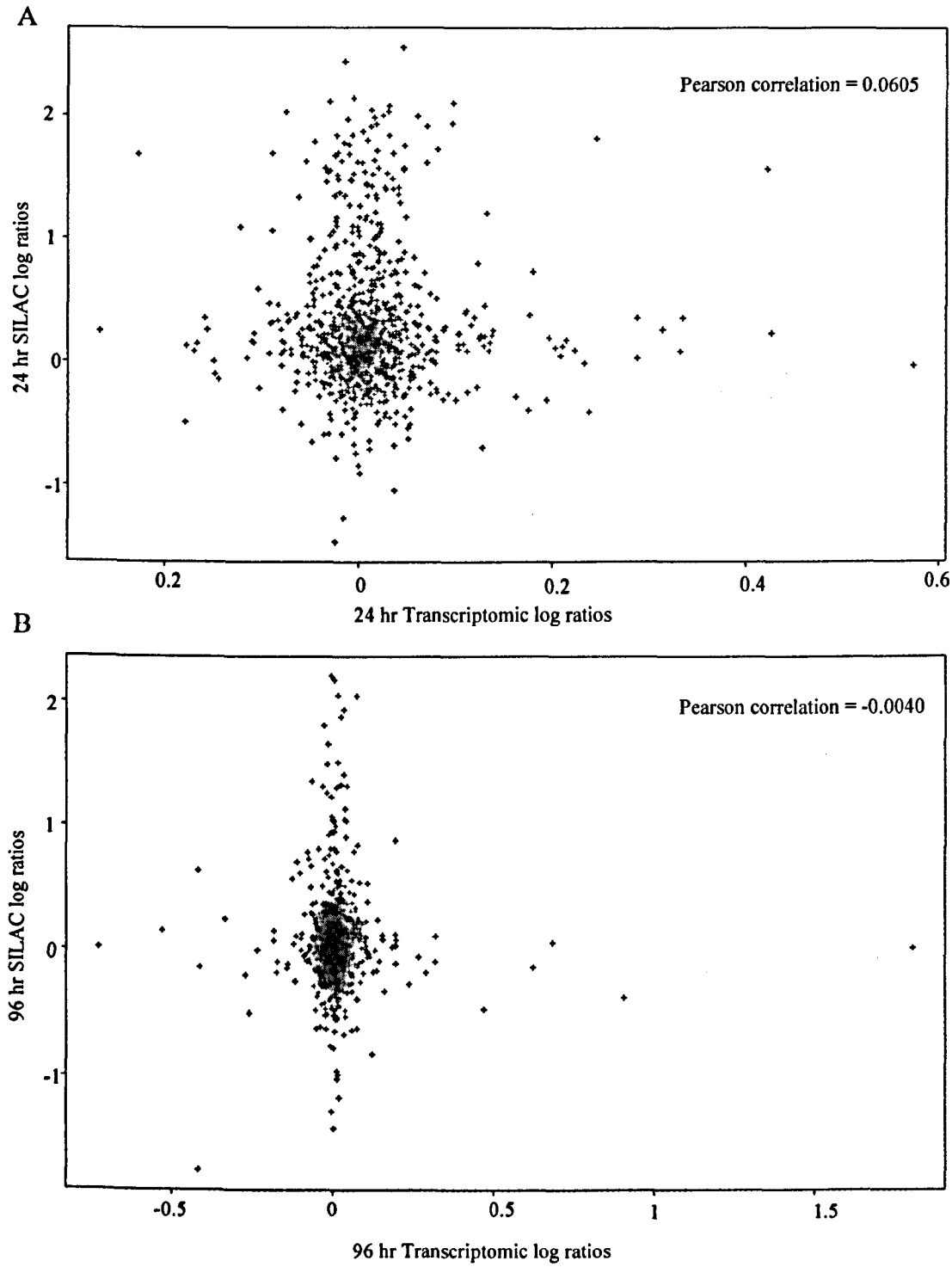



**Figure 6.3** ***Comparison of data from RNA-seq and SILAC analysis of the effect of salbutamol treatment on mouse muscle cells.*** The genes identified by RNA-Seq SOLiD sequencing were compared with the proteins identified using SILAC, which uses stable isotope labelling to identify peptide pairs and determine modulation, and mass spectrometry. Samples were taken at a) 24 hours and b) 96 hours post treatment with salbutamol. A Pearson correlation was performed on the log ratios to compare the datasets. Log ratios were ascertained by taking the log of the fold change between treated and control intensities or counts for proteins or genes which had a value for both the control and treated samples.

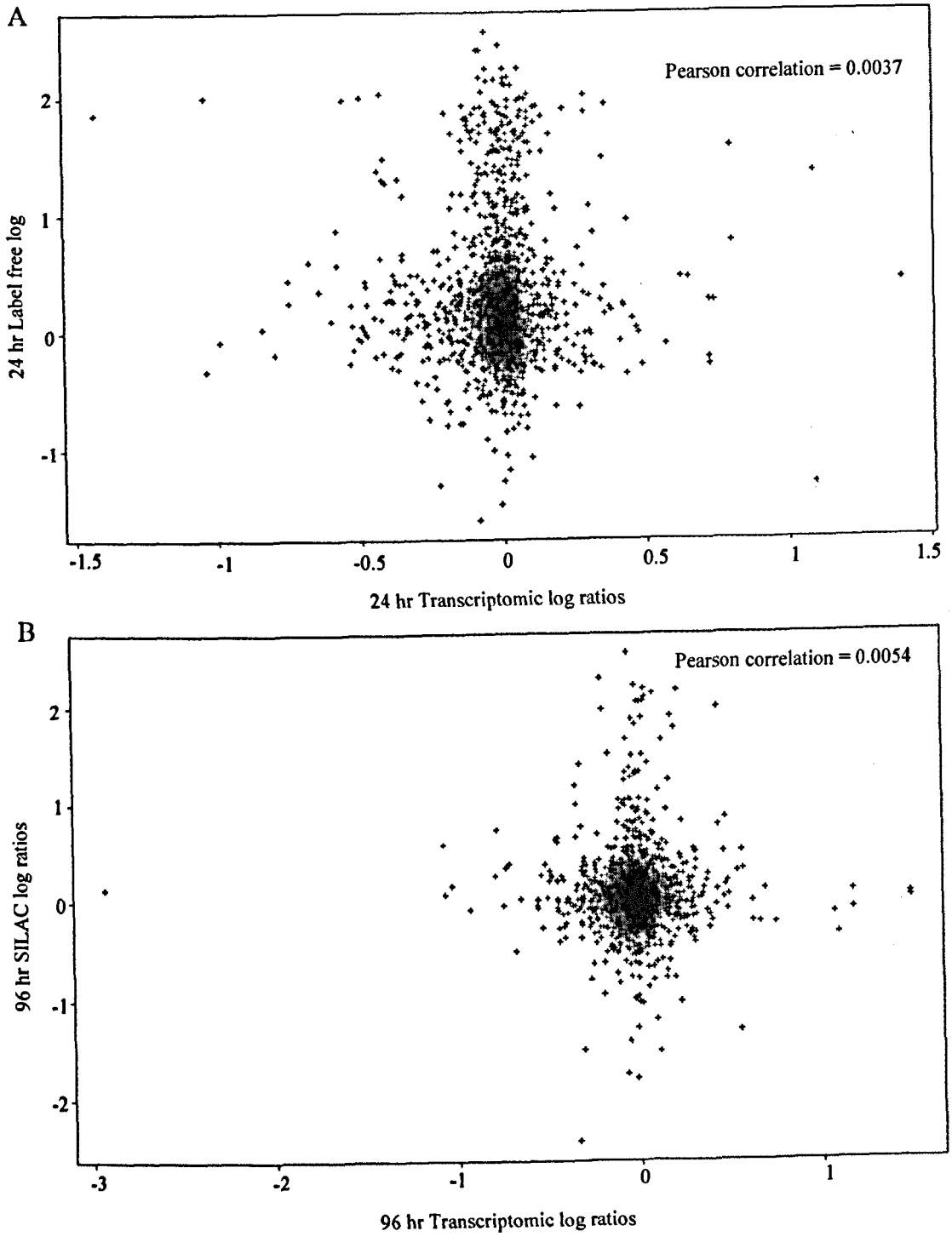



**Figure 6.4** ***Comparison of data from RNA-seq and label free analysis of the effect of salbutamol treatment on mouse muscle cells.*** The genes identified by RNA-Seq SOLiD sequencing were compared with the proteins identified using label free, which uses peak intensity measurements to determine modulation, and mass spectrometry. Samples were taken at a) 24 hours and b) 96 hours post treatment with salbutamol. A Pearson correlation was performed on the log ratios to compare the datasets. Log ratios were ascertained by taking the log of the fold change between treated and control intensities or counts for each protein or gene which had a value for both the control and treated samples.

Using the log ratios, which are the logs of the treated:untreated ratios, from all three methods as well as three computer software programmes (Progenises, MaxQuant and R) a heat map was produced using a freely available programme called Gene-E (http://www.broadinstitute.org/cancer/software/GENE-E/). This programme also allows clustering to be included within the heat map, providing information as to which data sets are similar (Figure 6.5). It was expected that each technique would cluster together, with the proteomic techniques being more similar to each other than to the transcriptomics. Although the transcriptomics, excluding the data from the 2 hr and 6 hr sample set, have clustered together it could be concluded that the data was clustered closer together with the SILAC data than that of the label free. Another observation implies that the MaxQuant label free data does not correlate with the Progenesis label free data, though the label free data from Progenesis at 24 hr does correlate with the SILAC data. Overall, this figure gives further proof to the underlying problem of the lack of correlation between different techniques as well as the worry that different methodologies can result in different conclusions being made about a biological system.

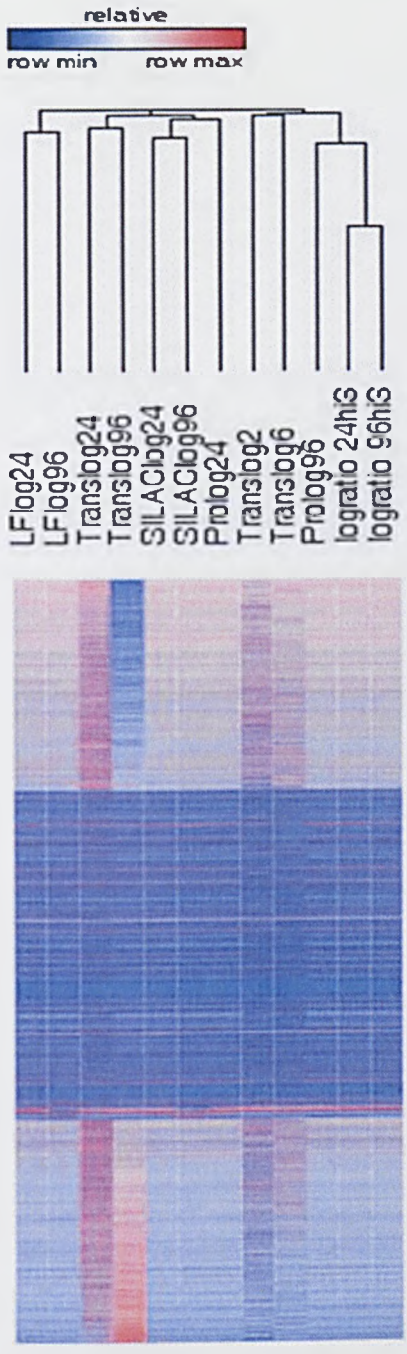



**Figure 6.5** ***Heat map displaying the log ratios from three different experimental procedures and three different computational data analysis processes.*** The log ratios from the RNA-Seq SOLiD sequencing data as well as the SILAC proteomic data, label free data produced by MaxQuant, Progenesis and Progenesis Post Processor. Log ratios, which are the logs of the fold change of treated and control samples, from both the 24 hr and 96 hr sample sets for the proteomic data and 2 hr, 6hr, 24 hr and 96hr are shown. Translog represents the data from the transcriptomic experiment with the number relating to the time point, SILAClog is the data from the SILAC experiment analysed by MaxQuant, LF from the Label free experiment analysed by MaxQuant, Prolog is the data from the label free experiment analysed by progenesis and logratio hi3 is the label free data analysed by Progenesis Post Processor.

Both transcriptomic and proteomic methods produce large quantities of data which need to be processed to allow the relevant and statically viable results to be gained. Each technique produces a different data output which needs to be analysed in a different way to correctly determine the important information. There are many software packages available to analysis these types of data, though few are compatible with more than one type of data output. This may result in software related bias, and therefore in variation in results. On top of this potential problem is the issue of normalisation. Normalisation is extremely important when analysing large amounts of data with many repeats but again there are several methods available, such as central tendency (Yang *et al.*, 2002), linear regression (Park *et al.*, 2003), local regression (Cleveland & Devlin, 1988), and quantile (Bolstad *et al.*, 2003) normalisation, which all have their own benefits. Once lists of important products have been identified it is often necessary to determine the relationship between these products. This can be achieved using functional or pathway analysis but as like all other bioinformatic processes there are a number of packages available all producing slightly different outputs.

### 6.3.3 Correlation between different software packages for proteomic studies

One of the most challenging aspects of proteomic analysis is being able to quantify proteins of interest. Currently there are a number of techniques available to undertake this task. Differential analysis of two-dimensional gel electrophoresis is still widely used for protein separation and quantification. (Burniston *et al.*, 2011; Choi *et al.*, 2010; Kosako & Nagano, 2011; Tafelmeyer *et al.*, 2008; Thomas *et al.*, 2011) However, despite the continuous improvement of this technique, several challenges

still exist when analysing and quantifying the proteins indentified. These challenges lie at the limits of the gel's analytical capability, causing issues with proteins with very high or low molecular weights, those with extremes in isoelectric points or of low abundance. This has encouraged researchers to use mass spectrometery based techniques in quantification. The use of stable isotopes has become frequently used as a process for quantification in recent years (Deeb *et al.*, 2012; Drexler *et al.*, 2011; Munday *et al.*, 2012; Swa *et al.*, 2012) but this technique comes with its own flaws, such as incomplete labelling and difficulty in simultaneously quantifying proteins/peptides in multiple samples. This led to the development of label free methods however, these techniques also have problems, such as to accurately quantify proteins each run must be identical to the last which is not always the case. Due to both of these techniques being fundamental to current quantitative studies it thoroughly important to understand of how label and label-free techniques compare.

A few studies have compared different proteomic techniques to each other in an attempt to decipher if one technique has an advantage over the other (Hendrickson *et al.*, 2006; Patel *et al.*, 2009; Ryu *et al.*, 2008). For example, Collier *et al.*, (2010) compared SILAC to spectra counting in quantifying the proteome of human embryonic stem cells. They determined, with respect to protein quantification, that the spectral counting methodology was able to quantify more proteins (885) than the SILAC procedure (450), although less accurately unless a five spectral count limit was established which then reduced the number of proteins quantified (340). Collier *et al.*, went on to further compare these two techniques by relatively quantifying cytoplasmic proteins expressed during the differentiation of human embryonic stem cells toward the trophectoderm at 0, 6 and 12 day time points (Collier *et al.*, 2011).

Again they determined that the label free spectral counting technique identified more unique peptides; 27004 peptides compared with 15072 peptides identified by SILAC. However, they also concluded that the magnitude of change identified by spectral counting is consistently less than that determined by SILAC. Although, through further analysis of the data they found that the sensitivity of spectral counting relative to SILAC increases with increasing spectral counts.

However, the techniques themselves are not the only thing that needs to be considered when a comparison is drawn. The software available to identify the large amount of mass spectrometry data can often differ between each package, possibly giving different results depending on which is used. Correlations between the software packages are vital, here two packages are looked at which are able to analyse both label free and SILAC data as well as software which can ascertain the absolute abundance for each protein. As well as this correlation, the techniques themselves are looked at within each software package.

#### 6.3.3.1 Comparison of the three software packages

With the wide availability of mass spectrometry methods to analyse a variety of complex biological samples the necessity for computational tools to analyse and statistically evaluate data generated from LC-MS experiments has become apparent. There have been many developments of different software packages to accommodate the different proteomic techniques. More recently software packages have started being able to handle more than one data set for example MaxQuant and Progenesis

(both of which have been used in this study), which are able to relatively quantify proteins from SILAC and label free data.

MaxQuant was originally designed for the analysis of SILAC spectra but can now also be used for label free data analysis of peak intensities. It produces peak lists, quantitation, false positive results determined from search engine results, peptide to protein group assembly and data filtration and presentation (Graumann *et al.*, 2008). It is specifically aimed at high resolution mass spectrometry data using its own search engine; Andomeda. Progenesis LC-MS (NonLinear Dynamics) was designed for label free data analysis, also allowing quantification and identification. Progenesis has the ability to combine and compare results from different runs, allowing for between-run variation in the LC separation technique. The software then produces an aggregated data set from the aligned runs to increase consistency in peak matching. Again, Mascot can be used as the search engine and statistical analysis applied to the peptides identified to produce a list of modulating proteins. Andy Jones' group from the University of Liverpool have recently developed a piece of post-processing software allowing Progenesis to analyse stable-isotope labelling data and top3 pseudo-absolute quantification, called the *Progenesis Post-Processor*.

To determine whether both MaxQuant and Progenesis produce similar data sets a Pearson correlation was performed on all the data generated by these software packages. The MaxQuant label free proteomic data from both 24 hr and 96 hr sample sets were compared to results produced using the same samples but run through Progenesis. From Figure 6.6 it can be seen that there is a slight correlation which is reflected in the Pearson correlation values of 0.3438 and 0.3358 for 24 hr and 96 hr

respectively. The untreated samples show almost identical results (Appendix III, Figures 1 and 2). Although a correlation can be seen, it is not as great as expected providing evidence that, although both use peak intensities, the results do not correlate well across packages when comparing total identifications. However, when using the only data from proteins that are detected in both software packages the correlation is 0.8989 and 0.8777 for the 24 hr and 96 hr sample sets respectively (Figure 6.7). This suggests that when data is identified for a protein the correlation is good but there are a number of identifications which are only made in one or the other software package. If the user is not aware of the different criteria the software packages use to identify proteins, differences in results will been identified.

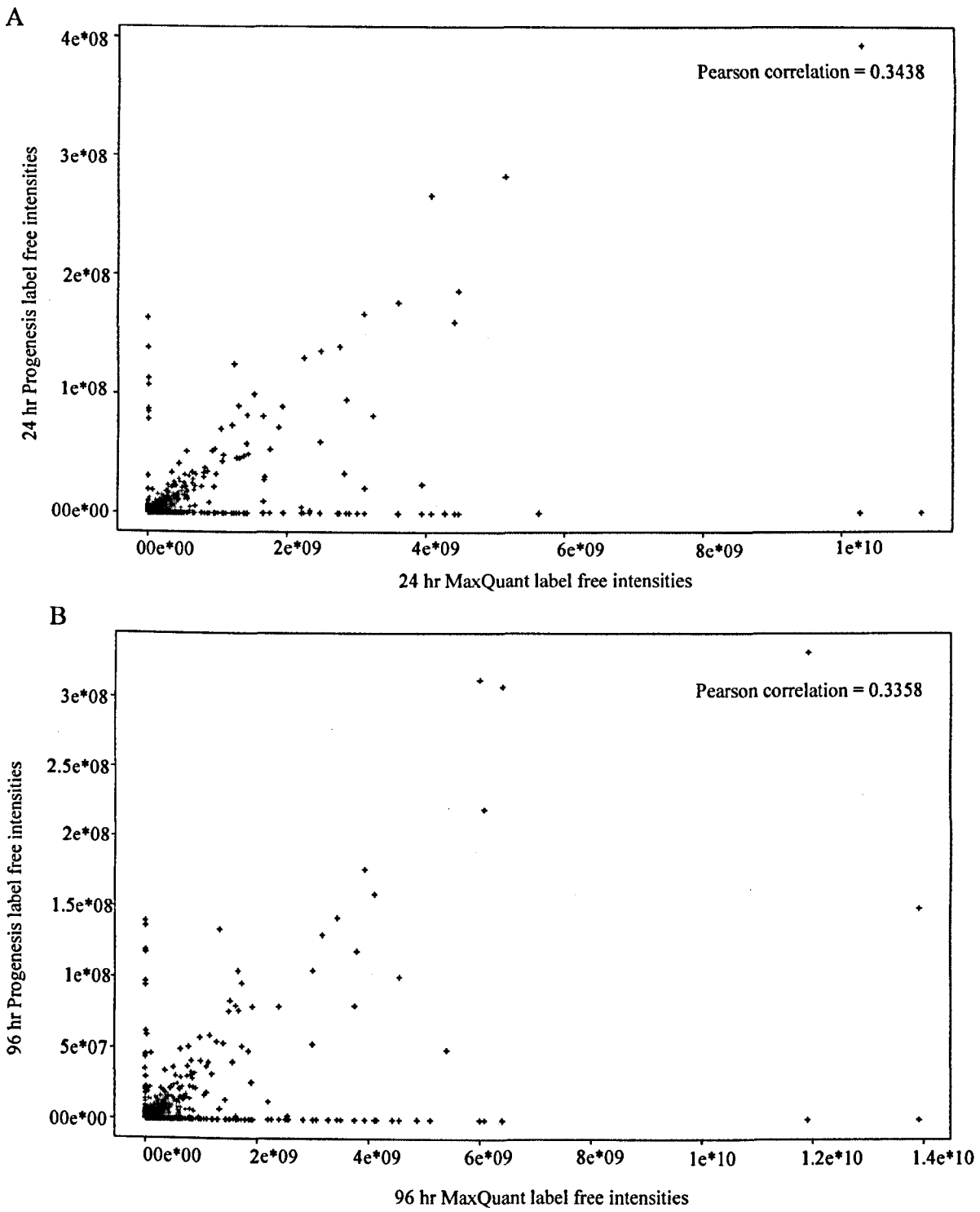



**Figure 6.6** ***Comparison of data from the label free analysis when analysed through either MaxQuant or Progenesis.*** The spectra obtained from the label free technique were analysed using either MaxQuant or Progenesis. These software packages allowed protein identifications and the relative abundance for each protein to be determined. A Pearson correlation was performed within R using the intensity values for each protein identified from the two bioinformatic analyses. Data from treated samples taken at a) 24 hours and b) 96 hours post treatment with salbutamol are shown.

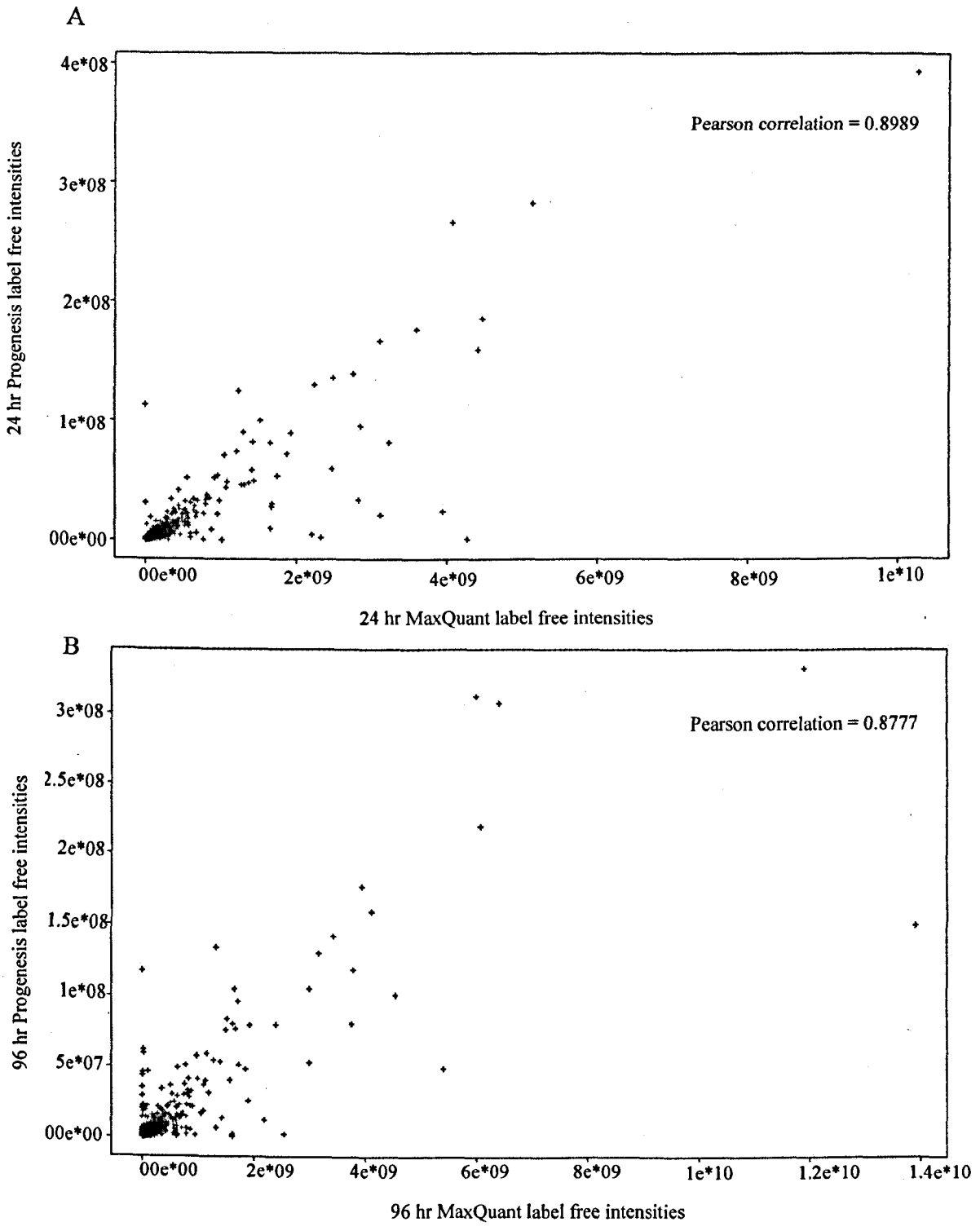



**Figure 6.7** ***Comparison of data from the label free analysis when analysed through either MaxQuant or Progenesis.*** The spectra obtained from the label free technique were analysed using either MaxQuant or Progenesis. These software packages allowed protein identifications and the relative abundance for each protein to be determined. A Pearson correlation was performed within R using the intensity values for only proteins identified in both sets of analyses. Data from treated samples taken at a) 24 hours and b) 96 hours post treatment with salbutamol are shown.

Qi, *et al.*, (submitted) have developed an interface to allow the use of the feature-level quantitation performed by Progenesis for other techniques. As well as allowing SILAC data to be analysed this new software allows a similar method to label-free analysis known as top3 or average intensity protocols to be performed. This method was originally developed by Waters for the SYNAPT range of instruments (Silva *et al.*, 2006). A correlation has been observed between the average of the three most intense peptide ions and the absolute abundance of that protein. The data previously collected from the Progenesis and MaxQuant analysis is then compared to the Progenesis Post-Processor data to see if there is any correlation between relative abundance and absolute abundance. The intensities and peptide identifications from the Progenesis analysis are used to produce the absolute abundance data and should correlate better than the MaxQuant data. When Progenesis vs Post Processor for the label free data was compared, Pearson correlations of 0.2308 (24 hr) and 0.1745 (96 hr) were obtained for data collected on proteins identified in both processes (Figure 6.8). However, using the most up to date version, colleagues at the University of Liverpool have repeated this analysis and found extremely good correlations between the two. When the Post Processor label free data was compared to the MaxQuant label free data scores of 0.3234 (24 hr) and 0.2828 (96 hr) were obtained (Figure 6.9) suggesting again the relative and absolute abundances have little correlation. However, when the SILAC data from MaxQuant was correlated to that produced via the Post Processor a correlation coefficients of 0.5874 (24 hr) and 0.6412 (96 hr) were obtained with a positive correlation (Figure 6.10). The correlation between relative and absolute is nearly double that of the label free the data though this may be due to the SILAC method producing more identifications. It is also interesting to

note that with the SILAC data the correlation is greater using the 96 hr sample set however using the label free data it is the 24 hr sample set that is higher.

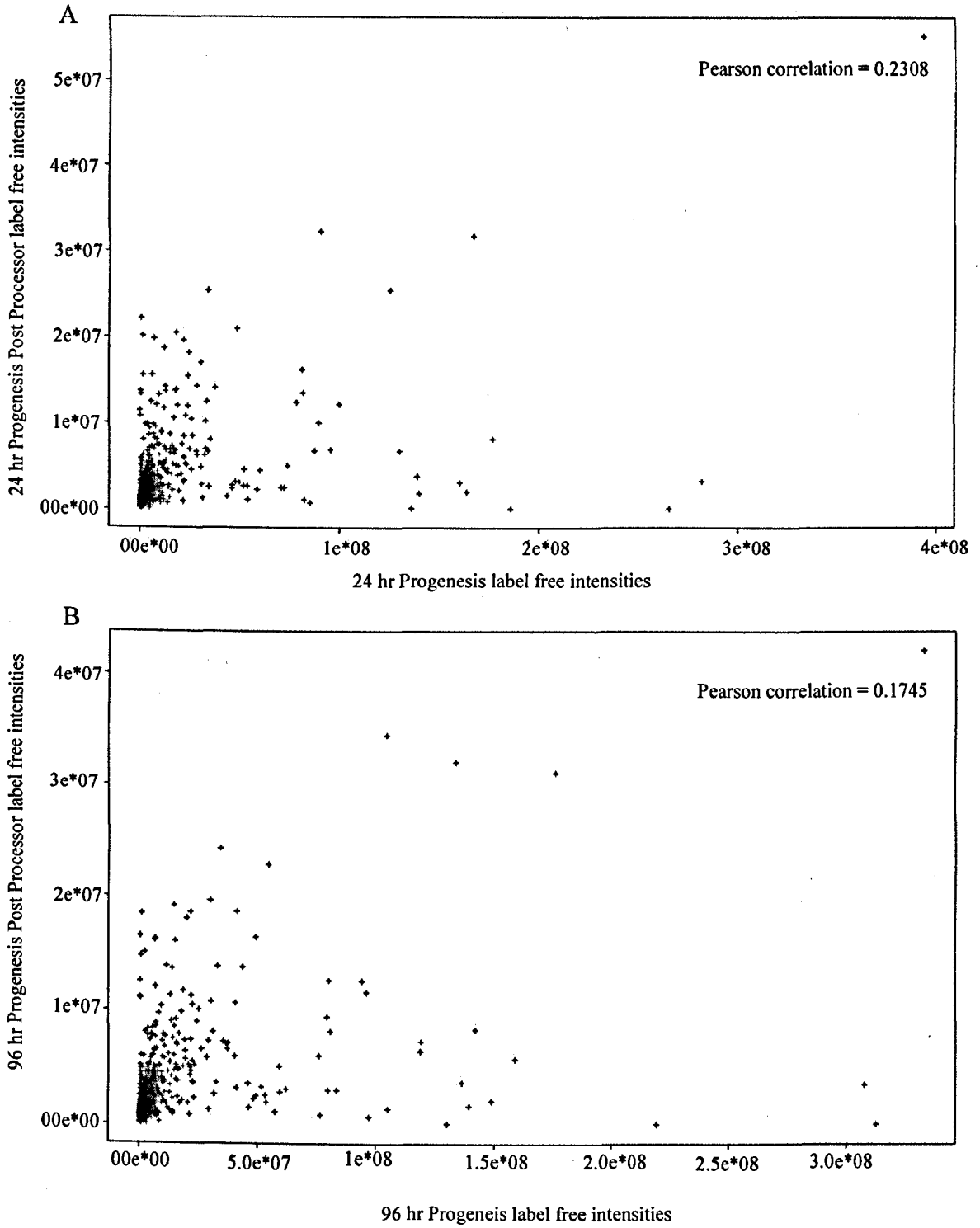



**Figure 6.8** ***Comparison of data from the label free analysis when analysed through either Progenesis or Progenesis Post Processor.*** The spectra obtained from the label free technique were analysed using either Progenesis, which provides relative abundance for each protein to be determined, or Progenesis Post Processor, which provides absolute abundance for each protein identified by analysing the top three most abundant peptide ions. A Pearson correlation was performed within R using the intensity values for only proteins identified in both sets of analyses. Data from treated samples taken at a) 24 hours and b) 96 hours post treatment with salbutamol are shown to identify any correlation between relative and absolute abundance.

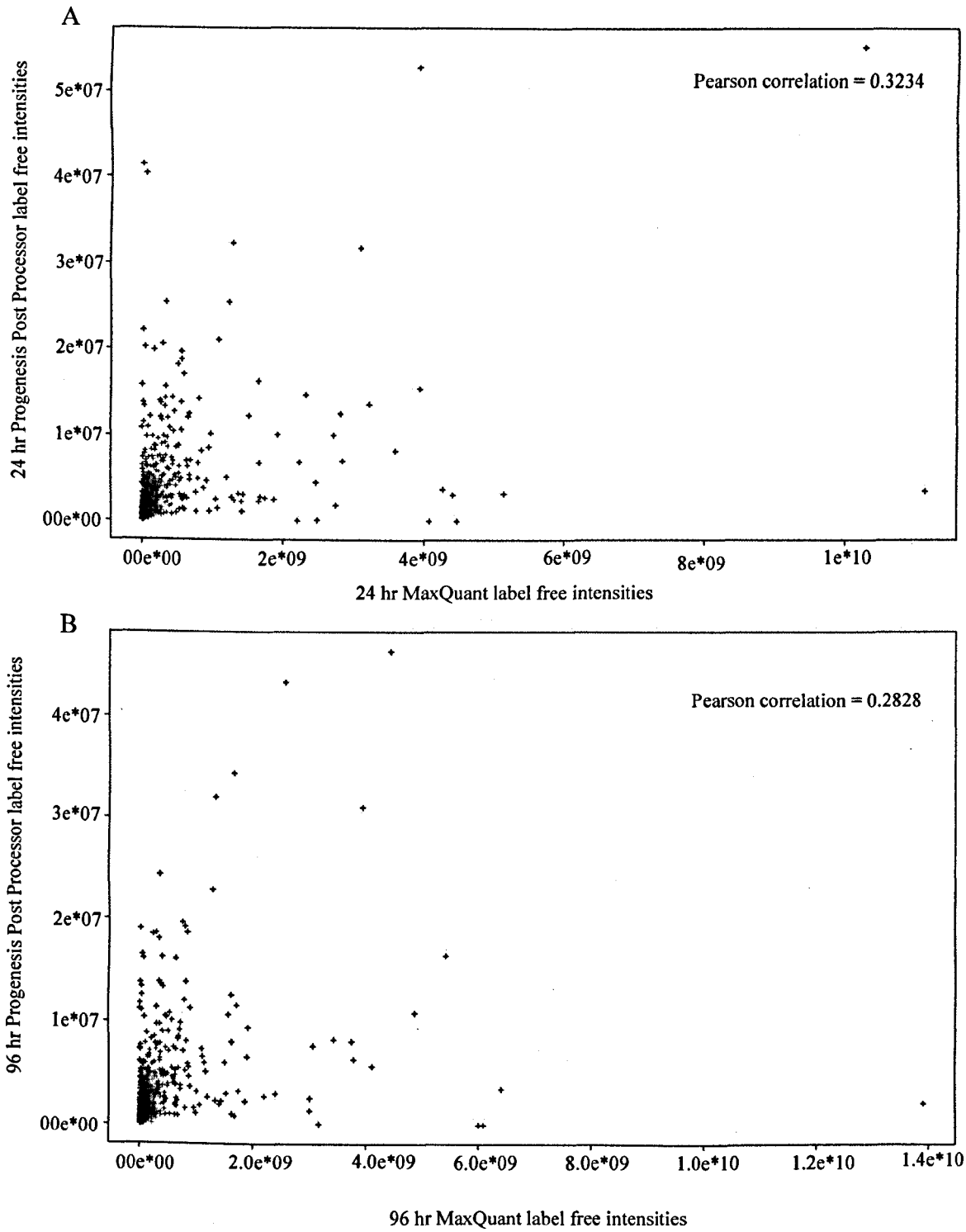



**Figure 6.9** ***Comparison of data from the label free analysis when analysed through either MaxQuant or Progenesis Post Processor.*** The spectra obtained from the label free technique were analysed using either MaxQuant, which provides relative abundance for each protein identified, or Progenesis Post Processor, which provides absolute abundance for each protein identified by analysing the top three most abundant peptide ions. A Pearson correlation was performed within R using the intensity values for only proteins identified in both sets of analyses. Data from treated samples taken at a) 24 hours and b) 96 hours post treatment with salbutamol are shown to identify any correlation between relative and absolute abundance.

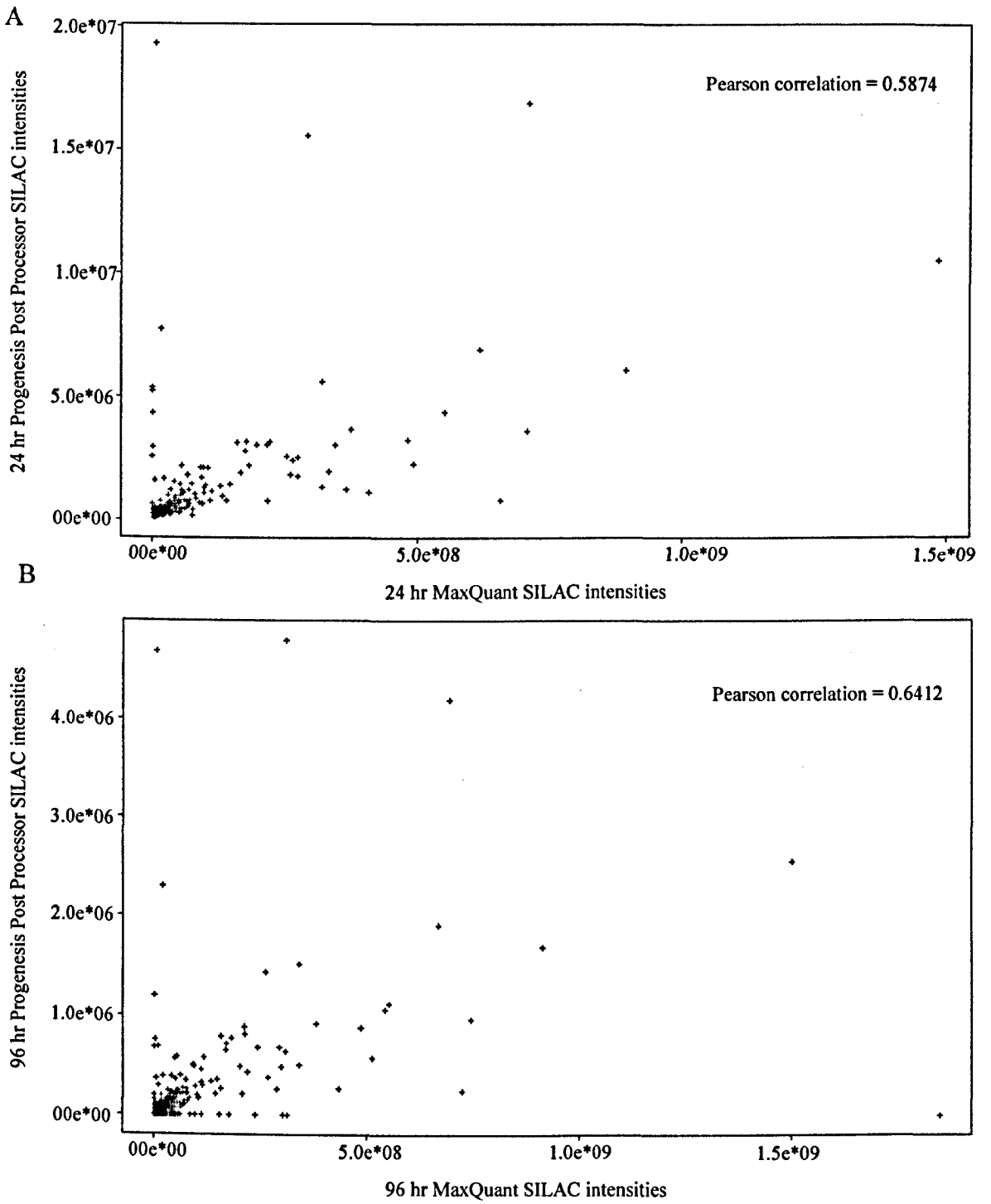



**Figure 6.10** ***Comparison of data from the SILAC analysis when analysed through either MaxQuant or Progenesis Post Processor.*** The spectra obtained from the SILAC technique were analysed using either MaxQuant, which provides relative abundance for each protein identified, or Progenesis Post Processor, which provides absolute abundance for each protein identified by analysing the top three most abundant peptide ions. A Pearson correlation was performed within R using the intensity values for only proteins identified in both sets of analyses. Data from treated samples taken at a) 24 hours and b) 96 hours post treatment with salbutamol are shown to identify any correlation between relative and absolute abundance.

### 6.3.4 Correlation between different techniques and their respective software packages

As previously discussed, identifying how well different techniques correlate to each other is important. Deciding upon a technique to use can often be reliant on the type of sample, the complexity of said sample, the length of time a procedure will take, the cost and the equipment available. Therefore, identifying how reliable and comparable the chosen technique is from the number of options available is highly important. By correlating the label free peak intensity data with the SILAC data, which has been produced under the same conditions with the same cells and reagents, it is hoped that an insight in any differences between the techniques can be highlighted. Different computational methods where undertaken to analyse the data as previously discussed. A comparison of the two methodologies has been completed using the data from MaxQuant as well as data from Post Processor to determine if the correlations are affected when absolute data is compared.

From this analysis it can be seen that the two proteomic techniques have a positive correlation with each other expressed by their Pearson correlations when MaxQuant data is compared, which are 0.6745 for 24 hr and 0.7712 for 96 hr. It can also be determined that both techniques identify a number of proteins not seen within the other (Figure 6.11). When these data points are removed leaving only values of proteins identified in both experiments the correlations stay very similar (0.6891 for 24 hr, 0.7649 for 96hr) suggesting the difference is in the data collected and not due to proteins not being identified (Figure 6.12). However, when Progenesis Post Processor was utilised to produce absolute data the correlation drops significantly with a value of 0.1801 for 24 hr and 0.1459 for 96 hr (Figure 6.13).

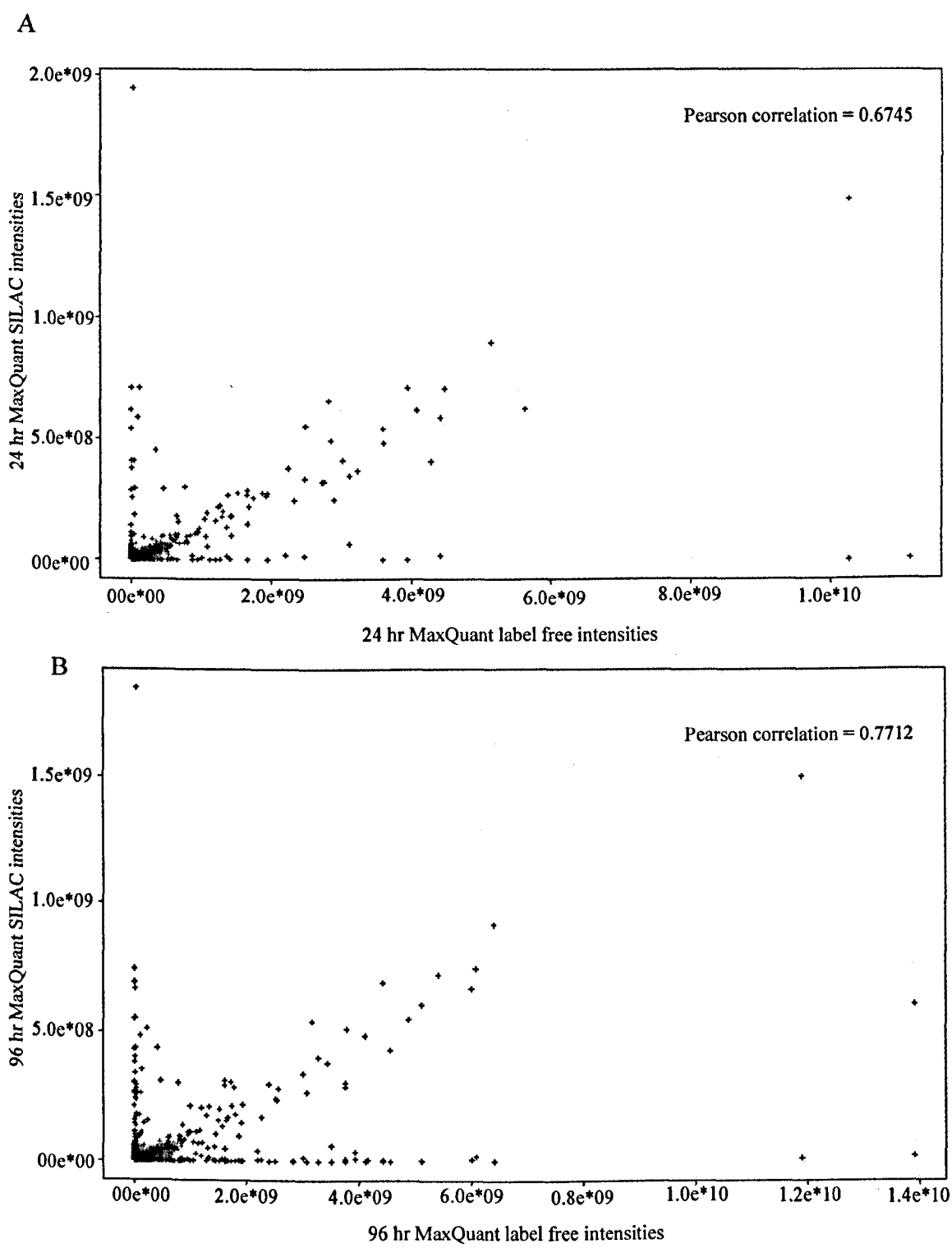



**Figure 6.11** ***Comparison of data from the SILAC and label free analysis when analysed through MaxQuant.*** The spectra obtained from the SILAC and label free experiments were analysed using MaxQuant, which provided relative abundance for each protein identified. A Pearson correlation was performed within R using the intensity values for all proteins identified from both sets of analyses. Data from treated samples taken at a) 24 hours and b) 96 hours post treatment with salbutamol are shown. This allows any correlation between the abundance data, from the different techniques, to be observed.

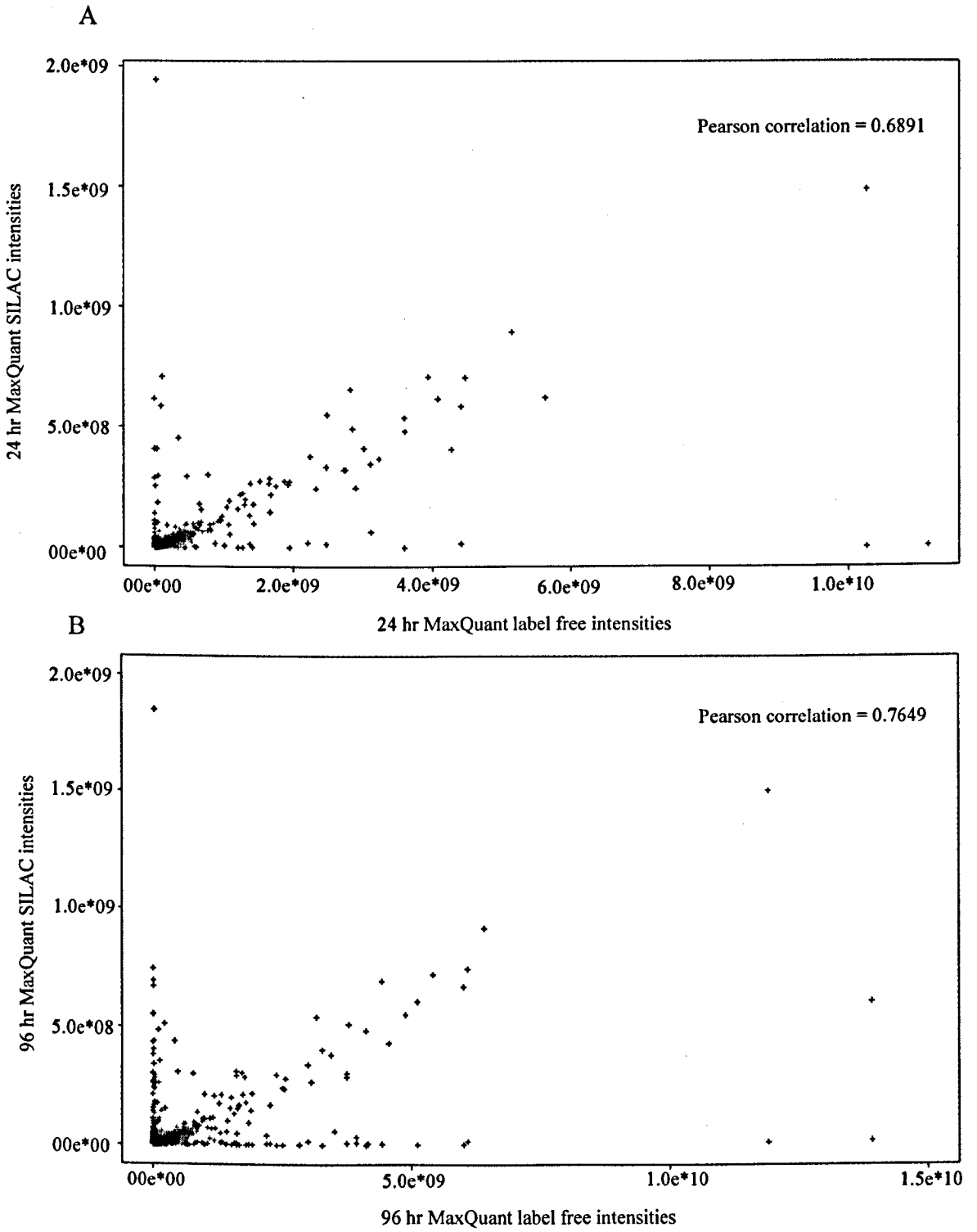



**Figure 6.12** ***Comparison of data from the SILAC and label free analysis when analysed through MaxQuant.*** The spectra obtained from the SILAC and label free experiments were analysed using MaxQuant, which provided relative abundance for each protein identified. A Pearson correlation was performed within R using the intensity values for only proteins identified in both sets of analyses. Data from treated samples taken at a) 24 hours and b) 96 hours post treatment with salbutamol are shown. This allows any correlation between the abundance data, from the different techniques, to be observed.

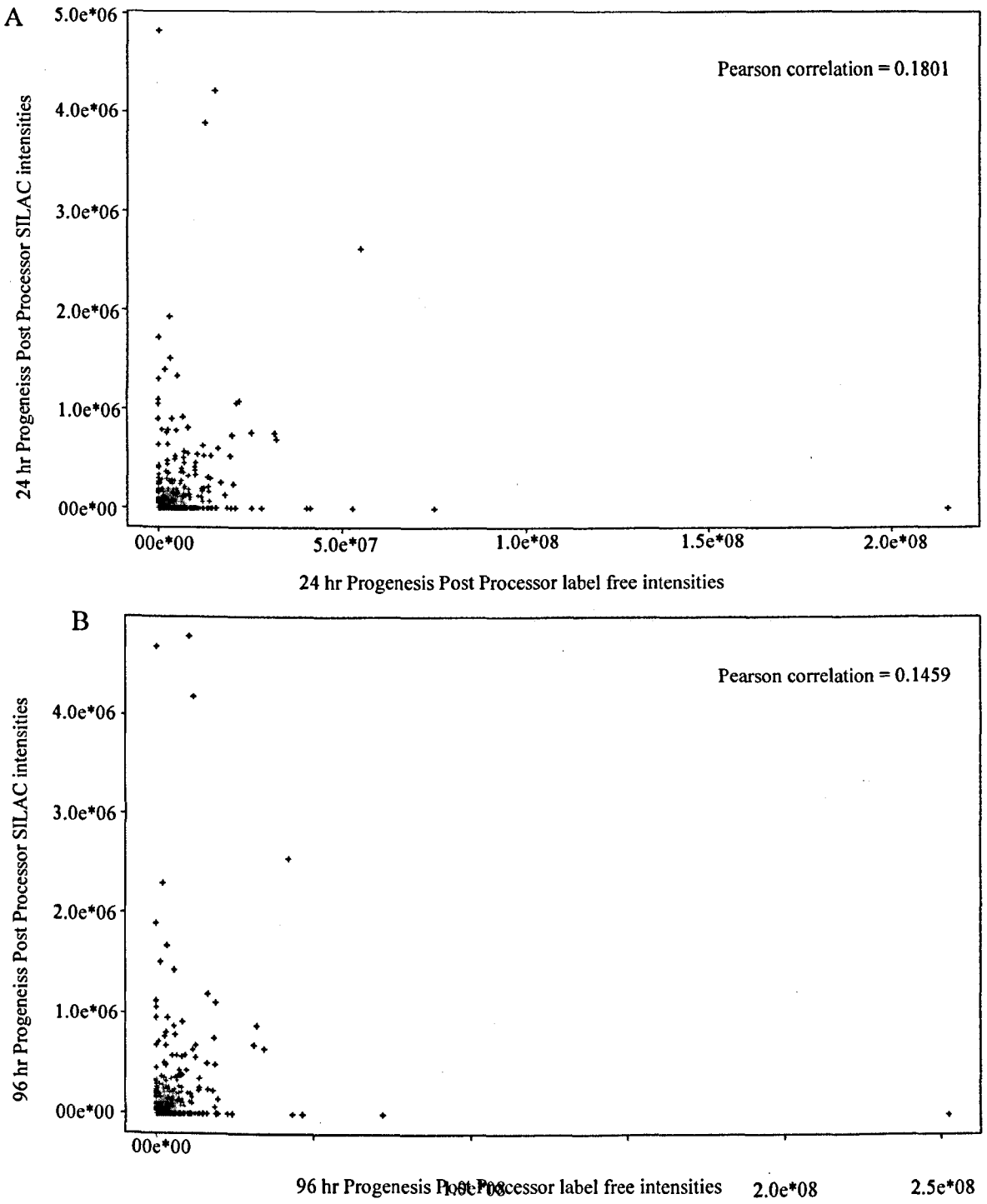



**Figure 6.13** ***Comparison of data from the SILAC and label free analysis when analysed through Progenesis Post Processor.*** The spectra obtained from the SILAC and label free experiments were analysed using Progensis Post Processor, which provided absolute abundance for each protein identified by analysing the top three most abundant peptide ions. A Pearson correlation was performed within R using the intensity values for only proteins identified in both sets of analyses. Data from treated samples taken at a) 24 hours and b) 96 hours post treatment with salbutamol are shown. This allows any correlation between the absolute abundance data, from the different techniques, to be observed.

## 6.4 Overall conclusion

Accurately correlating data from gene transcription to protein regulation is a difficult task. Due to factors such as post-transcriptional modifications and specific proteolyic processing, the cellular proteome will be at least one order of magnitude more complex than that its transcriptome (Hegde *et al.*, 2003). Furthermore, since current mass spectrometry-based techniques measure peptides which, depending upon the protease used and the number of possible cleavage sites on each protein, a complete digestion of an entire eukaryotic cell may contain up to approximately one million different peptides. This, along with factors such as protein turnover and different stabilities as well as the different biological half-lives of RNA and proteins, cause problems and discrepancies when investigating the modulation of both within a cell.

In this study, a transcriptomic investigation using RNA-Seq, to determine which genes have been differentially expressed after treatment with salbutamol, was compared to proteomic data from two different techniques; SILAC and label free peak intensity. A limited number of protein identifications were made using either SILAC or label free compared to the large numbers of identified genes, suggesting the link between genes and proteins is still not able to be fully investigated. However, when further investigating the data gained it can be seen that those genes with a high abundance have their corresponding proteins identified in the proteomic experiments with much lower fold changes but the same direction in expression. In contrast to the study by Schmidt *et al.*, (2007), the data collected across the three experiments does not confirm the expectation that mass spectrometry-based

proteomics have a bias towards detecting modulated proteins encoded by highly expressed mRNAs, as a wide range of different protein products were identified.

When data was compiled for each protein or gene identification made from the three experiments and compared to each other, the predicted correlations were not seen. The transcriptomic data did not correlate significantly with any of the proteome data, with a significant amount of protein products undetected for the genes. The proteomic data correlated with a higher degree but still with a number of identifications not made in one or the other technique. From the cluster analysis the SILAC data correlates slightly better than that of the label free to the RNA-Seq data. However, due to the limited statistical analysis that could be done with the RNA data due to pooling of the samples it is difficult to make any firm conclusions about corresponding proteins. Although there seems to be a number of identifications not detected across the datasets, this may be due to bias within the results. As the total data was analysed this included data with counts or intensities with only one measurement in either the control or treated sample. It was assumed that if no peptides were identified, therefore resulting in an intensity value of 0 was recorded for a protein in one condition but in the other condition peptides were identified, resulting in an intensity value, this cannot be taken as a true value of expression differences. It is thought this is most likely due to the technical limitations preventing this protein's peptides from being identified in the other condition. It is unclear how true this assumption is but if correct this would mean that a large proportion of the data gathered produced incorrect ratios which influenced the correlations. The differences between the transcriptomic and proteomic data may also be due to the complexity of the samples used. No pre-fractionation was performed for these

experiments with only one run per repeat being performed: although the data from all four runs were combined there could still be a number of identifications not detected. This could indicate the need for multiple mass spectrometry runs per sample or better fractionation methods to increase the range of proteins identified.

When examining the bioinformatic software applications available for proteomic analysis it became apparent that there are limited options available for using the same programme for different techniques. When correlations were made between Progenesis and MaxQuant using the label free data the correlation was not as strong as expected. A stronger correlation is needed to be able to have confidence that both of these packages are giving a true representation of the results. The majority of differences will probably arise from the different settings and criteria for analysis each programme use. However, as the two data sets being analysed are not identical some of the differences may lie within the techniques and not just the software. This theory was tested by comparing the analysis of the label free data by MaxQuant and Progenesis. As there was again a low correlation it could be concluded that the software does have a role to play in the discrepancies observed. Dr Andy Jones and his group at the University of Liverpool have recently submitted a paper comparing the performance of different packages for label free data. They have identified a number of errors which can be encountered when using different software packages and highlighted the need for a set of systematic guidelines for software users to follow, enabling increased confidence in results from quantitative studies. This study also provided the opportunity to utilise a new programme called Progenesis Post Processor. This allowed the absolute abundance of the data produced by both the label free and SILAC experiments to be determined. It would seem that the relative

abundance values calculated do not correlate well with the absolute abundance values, though those produced for the SILAC data have a much higher correlation than the label free. Overall, this chapter highlights the necessity for continuing development of methodologies to allow for better correlation of data across a range of areas.

# 7 Chapter 7
# Discussion and further work

Systems biology aims to understand biological processes at the level of the interactions, dynamics and complexity of multiple molecular elements. Proteins and genes are essential components of any biological process, and the analysis of their molecular function and regulation is crucial to the understanding of biological systems. It is known that, in eukaryotic cells, the transcriptome, proteome and metabolome are greatly influenced by the growth that cell (Adams & Haddad, 1996; Gutteridge *et al.*, 2010) which is the basis of this thesis. In 1967 Larsen, *et al.*, synthesized a compound commonly known today as salbutamol, a $\beta_2$-adrenergic agonist. These β-agonists were found to have an effect on muscle mass in the 1980's (Beermann *et al.*, 1987; Emery *et al.*, 1984; McElligott *et al.*, 1987; Reeds *et al.*, 1986), but like most steroids had many side effects. The continuing efforts to fully understand the mechanisms of muscle hypertrophy by these compounds endeavour to help refine the drugs to reduce the side effects. This would enable the use of β-agonists for combating muscle wastage in those patients suffering with this condition. This study also allowed for the opportunity to investigate how efficiently different areas of research are able to correlate their data in a system wide approach. As both the genes regulated and their corresponding protein products are essential to the understanding of any biological process, it is important that the data collected from both areas are able to link together. Both the molecular mechanisms of salbutamol and the importance in the technology available has been thoroughly investigated and discussed.

In Chapter 2, a proteomic technique known as DIGE was implemented to study the effects on the proteome of salbutamol on a primary cell line of rat skeletal muscle after different lengths of treatment. Gel based separation is widely used within

proteomics to identify proteins on a quantitative level, though DIGE has many advantages over the standard 2D gels. The computational, mass spectrometry-based and bioinformatic analysis of the DIGE gels and their resulting proteins were able to highlight several potentially interesting modulations occurring after treatment. A total of 31 non-redundant proteins were identified after 24 hr and 110 non-redundant proteins were identified after 96 hr of treatment, providing evidence that there is a greater response to salbutamol the longer the treatment was given. From these identified proteins 12 were seen in both time points, suggesting that different pathways are being utilized during different stages of the increased growth over time. Around a quarter of the identified proteins were cytoskeletal with the rest ranging from those involved in transcription and translation to calcium binding, energy production and cell proliferation. The use of DAVID and Ingenuity's Pathway Analysis software further increased the knowledge on how the modulated proteins interact with each other.

This study was not only able to increase the knowledge on the proteomic effect of β-agonists but also able to confirm previous findings. For example, Hong *et al.* (2011) used a 2-DE method to identify the proteomic changes within hypertrophic cardiomyocytes after treatment with carndothelin-1 (ET-1) and isoproterenol (ISO). They identified a number of functional classes which either ET-1 or ISO had an effect on including; signalling, hydrolase, cell organization, transcription, translation, chaperone, $Ca^{2+}$-binding protein, energy metabolism, and immune response. Many of these functions were also identified within this study using a similar agonist which produces the same hypertrophic response in muscle. Stella *et al.* (2011) investigated the protein changes within cattle skeletal muscle after the treatment of clenbuterol.

Using DIGE they were able to identify 19 non-redundant, differentially expressed proteins including several proteins that were identified in this study, for example, actin, desmin, myosin, heat shock protein beta-1 and beta-enolase. Burniston *et al.*, (2007) found that β-enolase was also increased, which correlates with our data. Several of the proteins identified within this study, for example ATP synthase, triosephophosphate isomerise and phosphoglycerate kinase 1, found at both 24 hr and 96 hr to be differentially expressed all have a role in either glycolysis or ATP synthesis. These findings suggest that an increase of energy may have an important role within the muscle undergoing forced hypertrophy. Downie *et al.*, (2008) concludes that it is likely that the increase in protein caused by clenbuterol is due to a drug-induced increase in transcription as opposed to an effect on translation, although in this study there were a number of transcriptional proteins identified there were also several translation initiation factors observed as up-regulated. This initial study has enabled a first insight into the mechanisms of salbutamol on rat muscle cells.

In Chapter 3, a study into the effect of salbutamol treatment on mouse muscle cells was undertaken. The transcriptome of a cell forms the template for protein synthesis, resulting in the corresponding proteome. Understanding the transcriptome is essential for understanding the mechanisms taking place within an organism on a cellular level. Furthermore, being able to compare transcriptomic and proteomic data is beneficial for full comprehension on the cells mechanics. However, this feat is yet to be fully achieved, as discussed in Chapter 6. In recent years, the use of "next generation sequencing" to monitor gene expression via sequencing of the whole genome has often been used for this kind of study. In this investigation this technique

is used on the Applied BioSystem's SOLiD platform to identify any modulation of gene transcription. Using EdgeR to perform a statistical analysis of the data collected from RNA-Seq of mouse cells after treatment a total of 161, 299, 2756 and 1346 differentially expressed genes were identified for sample sets taken at 2 hr, 6 hr, 24 hr and 96 hr after treatment, respectively. With further examination a total of 696 genes were found in two or more time points. The lists of modulated genes were mapped to known pathways using KEGG mapper, enabling visualisation of the genes and how they relate to each other. However, as genes can often transcribe more than one protein the same gene can appear in a number of pathways. The most prominent pathway identified was the MAPK pathway which is involved in cell signalling via cell surface receptors. It is thought that G-protein-coupled receptors may activate the MAPK cascade (Malarkey *et al.*, 1995) which supports the finding that β-agonists, a G-protein-coupled receptor agonist, encourage growth via the MAPK signalling pathways. In the study by Markou *et al.* (2008), the hypertrophy of cardiac muscle induced by phenylephrine, an $\alpha_1$-adrenergic agonists, was related to signalling mechanisms activating c-jun. They also identified that ERK1/2 and JNKs were the principal kinases responsible for phosphorylation of c-Jun. Although c-Jun is not identified in this experiment, several other genes with a role in the MAPK pathway which can regulate the expression of c-Jun were identified as being modulated. In Markou's study the effects of phenylephrine were only studied between 0-60 min, whereas, in this study the earliest time point taken was at 2 hr, which may provide an explanation to why this protein was not identified. However, from both this study and Markou's study it can be suggested that the MAPK pathway is affected from an early stage but is continuously modulated over time during the treatment. The time points studied play an important role in any investigation; as a lot of studies

involving β-agonists are performed *in vitro* (Bhavsar et al., 2010; Choo et al., 1992) the time points range from 24 hr to a couple of days or weeks. In this study, the time points ranged from 2 hr to 96 hr as salbutamol is a short acting β-agonist. A large increase in gene expression was seen between 2 hr and 24 hr suggesting that there was a lot of proteomic activity occurring during this time frame. Investigating this time frame in more detail may increase the knowledge on the mode of action of this drug.

Bogoyevitch *et al.* (1996) also showed that simulation of the adrenergic receptor activates the MAPK cascade in cardiac myocytes. Vaniotis *et al.* (2011) discovered that the MAPK signalling pathways, although not directly activated by the β-adrenergic receptors in the nucleus, modulate the ability of the receptors to mediate transcriptional events. In the transcriptomic analysis, NFκB was also identified in the pathway analysis as being modulated within the MAPK pathways. Previous work by Vaniotis *et al.* (2011) supports this as a target for β-agonists as it was seen to be down-regulated in response to treatment with isoproterenol. In this study, *Nfkb1* was identified as being down-regulated at 96 hr whereas *Nfkb2* was down-regulated at 24 hr. Overall, there is much evidence to suggest a role of MAPK in the hypertrophic response from β-agonists.

Although not directly comparable due to different cell models used, the DIGE data produced similar functional categories for the proteomic changes to those observed within the transcriptomic study, such as, signalling, cell organization and energy metabolism. Overall, an acquisition of numerous modulated genes resulting in a number of modulated pathways was achieved. This data set could then be compared

to the proteomic data to obtain further insight into mechanisms involved in the salbutamol induced hypertrophic response in muscle cells.

In Chapter 4 the first of the non-gel based proteomic investigations was performed. There are two prominent methods currently being commonly used in biological research, the first being stable isotope labelling and the more recent label free technique. In this study both of these techniques were utilised to, not only gain further knowledge on the workings of salbutamol, but also to investigate how efficiently different methodologies within the same research area correlate to each other. The first method to be performed was SILAC, which involved labelling mouse muscle cells with heavy arginine and lysine. With the use of an LTQ-Orbitrap and MaxQuant, a bioinformatic tool capable of ascertaining peptide pairs, a comprehensive investigation on the effects of salbutamol after 24 hr and 96 hr was hoped to be achieved. The full experiment was repeated twice with four biological replicates each time. However, across both experiments for each replicate a limited number of proteins, ranging between 4 and 27, were identified as being modulated. From these identifications only a handful were found in all replicates. When considering the replicates as different mass spectrometry runs and combining the data the number of non-redundant protein identifications increased. The number of modulated proteins was smaller than expected considering the comparatively greater amount of gene expression change, suggesting genomic data does not always correlate significantly with that of the proteome.

One consideration when analysing the results of this experiment was the varying protein identifications over the different replicates. However, this phenomenon is

most likely due to the complex nature of the samples and the inability of peptides to perform identically in each mass spectrometry run. Despite this there are still several proteins, functions and pathways that are potentially interesting for understanding hypertrophy or have been identified previously. For example, a large number of cytoskeleton proteins were identified in both experiments in more than one repeat. The modulation of a variety of cytoskeletal proteins was also observed in the DIGE experiment, suggesting that salbutamol has an effect within the structure of the cells, including organisation of the cytoskeleton as well as filament forming.

By continuing the analysis into the functional groups and biological processes being affected, several interesting results were revealed. There were a number of proteins indentified at both 24 and 96 hr, which are involved in the acetyl-CoA conversion and fatty acid regulation as well as the regulation of transcription and translation. Energy production is another important process during cell differentiation or growth and was highlighted in both the DIGE study and the transcriptomic study as being modulated. At 24 hr several proteins involved in energy regulation and ATP synthesis were identified as up-regulated, such as ATP synthase, glycine amininotransferase and 5'-AMP-activated protein kinase. However, when ascertaining the protein functions it was observed that many proteins can have a role in a range of processes, making it more difficult to determine which processes and pathways are actually being modulated. For example, Eukaryotic initiation factor 5A isoform was one of the translation and transcriptional proteins identified. This protein has several different functions attributed to it, including a role in cell proliferation (Hanauske-Abel *et al.*, 1994; Parreiras-e-Silva *et al.*, 2010) and

apoptosis (Jin *et al.*, 2003) as well as its interaction with specific sequences of mRNAs, suggesting a possible role in mRNA decay (Zuk & Jacobson, 1998).

Within chapter 4 the possible benefits of pre-fractionation were investigated. As the proteome is large and complex there is a suitable argument for using pre-fractionation before mass spectrometry analysis to simplify the samples allowing more information to be gained per run. However, there is the potential for large sample losses in the preparation and clean up stages. An OFFGEL fractionator was used to separate the whole cell sample lysates by their isoelectric point in solution via both protein and peptide space. Due to the clean up strategy needed when separating proteins, it was determined that the full experiment would be performed in peptide space. The resulting data showed that a large amount of protein identifications could be made after the pre-separation, however, as this may be due to the substantially greater number of total runs on the mass spectrometer when compared to not performing pre-fractionation.

Chapter 5 undertook the task of identifying whether a different proteomic approach could increase the understanding of the β-agonist mode of action as well as helping link the transcriptomic data with proteomic data. Labelling techniques have great advantages in proteomics, however, they still come with their own limitations leading to the development of label free mass spectrometry based techniques. It is this methodology which was applied to the salbutamol treated and control samples collected. By identifying the peak intensities for the peptides, identified proteins from the control and treated samples could be compared to establish which, if any, were changing. Analysis via MaxQuant, followed by a statistical evaluation using

Student's t-test enabled the detection of around 20 proteins which had been modulated. Once again the proteomic data gained was limited compared to that of the transcriptomic study and the proteins identified did not correlate well with those found in the SILAC experiment, highlighting a potential area of concern. It is essential that two methodologies investigating a component with a system, such as proteins, within the same model should be able obtain accurate results which correlate well. Although, there is a lack of correlation with exact proteins, the functions of those identified were related across the two experiments, suggesting that one run per complex sample is not sufficient to obtain all the information.

Despite the small number of proteins observed as being modulated it is probable that these proteins have an important role in the growth and change of the muscle cells after treatment with a β-agonist. For example, inverted formin-2, which was identified as up-regulated at 96 hr, may have a role in the changes occurring in the cytoskeleton of the muscle cells. Formins are actin-binding proteins that have multiple effects on actin dynamics and microtubules (Chhabra & Higgs, 2006; Paul & Pollard, 2009). It is known that actin filaments and microtubules have important roles within a number of different cellular processes including; cell migration, membrane transport and cell division (Rodriguez *et al.*, 2003). From the analysis of the DIGE experiment, a large amount of actin and actin regulatory proteins were identified as being modulated. It is possible that β-agonists cause a response via actin and its associated proteins within in rat muscle cells but the affect is not as great in mouse cells as in rat cells, explaining the lack of actin identified in either the SILAC or label free investigations.

Another interesting protein identified as modulated within this study was histone deacetylase 2, a key protein in gene transcription and protein function through its ability to alter the acetylation status of histones and other proteins. The removal of acetyl groups from histone tails by histone deacetylase 2 promotes transcriptional repression by allowing chromatin compaction (Haberland *et al.*, 2009). As this protein was seen to be down-regulated in this study it suggests that an increase in transcription may be occurring. A number of histones were identified in both the SILAC and DIGE investigations, adding weight to the hypothesis that an effect on transcription occurs when muscle cells are subjected to treatment with salbutamol. Several other transcriptional proteins or proteins involved in the regulation of transcription were identified over all four studies suggesting this process is extremely important in the hypertrophic process.

When a comparison of the data from the two proteomic studies was undertaken only histone H1.2 and C33 antigen were identified as modulated in both experiments. Histone H1.2 was identified as up-regulated at 24 hr in the SILAC experiment but down-regulated at 96 hr in the label free experiment. Whereas, C33 antigen was observed as being down-regulated at 96 hr in the SILAC experiment and up-regulated at 24 hr in the label free experiment. This suggests that, not only are different mechanisms at work at different times, but that the two proteomic techniques were able to identify different proteins with each repeat of the experiment. This can be observed when the number of total identifications made by the SILAC and label free experiments are compared. The label free approach enabled over twice as many identifications of proteins present in the muscle cells than the SILAC method. Although, the 95% of the proteins identified by SILAC were found

in the label free experiment only two of proteins identified as modulated from the SILAC experiement corresponded to the label free data. This suggests that the processes enabling modulation to be determined is the area which is producing the limitation when correlating data. Histones are known to be involved in DNA structure (Littau *et al.*, 1965) and are vital to chromatin structure in mammals (Fan *et al.*, 2005). As previously discussed a number of histones were identified in the DIGE experiment, suggesting that this protein and others with similar functions have an important role in the growth of muscle cells after salbutamol treatment. As histone deacetylase 2 was also identified as modulated in the label free experiment, it could indicate that transcription is also taking place after treatment. It has been observed that Histone deacetylase can be induced by heat shock protein 70 in cardiac hypertrophy in mice (Kee *et al.*, 2008). In the SILAC study, Hsp70/Hsp90-organizing protein was identified as up-regulated at 96 hr. Both Hsp70 and Hsp90 bind the co-chaperone Hsp70/Hsp90 organizing protein, which can help coordinate the heat shock proteins in folding protein substrates (Chen & Smith, 1998). This suggested that although Hsp 70 was not identified as modulated in either experiment, it may have been present but not identified and be involved in the process of hypertrophy. Although Hsp 70 was not identified in the DIGE study Hsp 60 was observed as modulated. These proteins often work together when a cell is under stress and have been observed to increase within muscle undergoing oxidative stress (Khassaf *et al.*, 2001).

Although there was limited proteins which corresponded across both proteomic techniques using the mouse muscle cells, there were several proteins identified in both experiments with similar functions. For example, proteins involved in the

metabolic processes surrounding acetyl-CoA and fatty acid metabolism such as citrate synthase and fatty acid synthase were identified using label free and 3-ketoacyl-CoA thiolase and long chain 3-hydroxyacyl-CoA dehydrogenase were identified using SILAC. Transportation of proteins was observed as a function which had a number of proteins from both experiments fall under, such as vacuolar protein sorting-associated protein 18 homolog which is associated with endocytosis and ADP ribosylation factor which is a regulator of vesicle biogenesis in intracellular traffic. Both proteomic experimental procedures highlighted proteins with a role in the regulation of transcription and translation. For example, heterogeneous nuclear ribonucleoprotein A0 was determined to be down-regulated at 96 hr in the SILAC study and DNA methyltransferase 1-associated protein 1 was identified as down-regulated at 24 hr in the label free analysis.

Overall, a number of interesting processes have been highlighted via both proteomic techniques and the transcriptomic investigation. However, the correlation between exact protein identifications and their respective genes is low. Having confidence in the method and software being applied to any study is exceptionally important. It is essential that the choice of method will not affect the coverage and content of data obtained and it is possible a number of techniques need to be applied to each investigation. The aim of Chapter 6 was to determine the correlation between the different techniques employed within this study, as well as the software packages available for data analysis. Through a range of methods; Pearson correlation, manual comparison and heatmap clustering, a limitation of a system wide study has been highlighted which lies in the ability of the current techniques to understand the mechanisms between gene transcription and protein expression. However, it would

seem that between the two proteomic techniques, SILAC has the edge on label free when comparing to RNA-Seq data. This advantage was again seen when the relative abundance of each protein was compared with their corresponding absolute abundance. The most important conclusion for systems biology comes from comparing the two proteomic techniques, as it becomes apparent that the label free technique can identify more proteins than SILAC and that even though the redundancy between them was high they are unable to identify the same modulations.

Software packages which are available for processing a range of data from proteomic techniques were also compared. MaxQuant has the ability to identify proteins and determine modulation of both labelled and label free data, whereas Progenesis only has the capacity to analyse label free data. However, Dr Andy Jones and his group at the University of Liverpool have recently developed a new tool which allows Progenesis to analyse labelled data. As part of the development process, the SILAC and label free data from this study were used to test the software (Qi *et al.*, 2012). By comparing the results obtained from both MaxQuant and Progenesis when analysing the same label free data this study was able to determine any correlations between the software packages. The correlation, determined via a Pearson correlation, was lower than expected with the software packages identifying a number of different proteins. These differences may arise from the different parameters within each software package and without a comprehensive knowledge of the workings of these packages bias within the results could appear.

A further area under investigation within this chapter was normalisation, which is a very critical step in quantitative experiments and usually the first to be undertaken. This process allows for corrections from technical effects, such as but not exclusive too, sample mixing errors, incomplete isotope incorporation or differences in ionization between independent LC-MS/MS experiments, all of which can influence the results. The basis for normalisation relies on the assumption that majority of proteins will remain unchanged allowing parameters such as total ion counts, total number of spectra or average ratios of the most abundant proteins to be the requirements for normalisation. In stable isotope labelling, including the study in Chapter 4, a control mixture of labelled and unlabeled untreated samples is usually used to determine any biological variation of ratios and mixing errors (Kierszniowska *et al.*, 2009). For gel electrophoresis methods such as DIGE (Chapter 2) a mixture of both control and treated samples are labelled with a third florescent dye and run on the same gel as the labelled control and treated samples. Another method for normalisation examines the log transformed data, which can be taken to reduce the variance between ratios. By plotting the data on a graph and determining its distribution the requirement for normalisation can be assessed. Other potential methods include quantile normalization (Callister *et al.*, 2006), variance stabilization (Kreil *et al.*, 2004) or a "spectral index" combining different features of each data point used for label free proteomics (Griffin *et al.*, 2010). With the range of options for normalisation being extensive, and different processes preferred for each biological technique, it brings the question of bias within results to light. If more than one biological technique is employed within a study to ensure the accuracy of the results obtained as well as increasing the breadth of knowledge, then using different methods to normalise the data could in turn affect the results.

Overall, in this study, a large amount of proteomic and genomic information was gained about the effects of salbutamol on skeletal muscle cells. Although the data between the four experiments do not correlate perfectly the value of this data should not be underestimated. From the genomic data a large response was observed after treatment with salbutamol which increased with time, peaking at 24 hr. Although a significantly smaller response was seen in the proteome, the functions of the proteins identified were related to the genes found to be modulated. As there will be a time lag between gene transcription and protein expression, the resulting small proteomic change may be due to the effect of the large genomic modulation at 24 hr not being processed until a later time point and therefore not identified within this study. The amount of differential expression per gene also needs to be taken into account. Although the fold changes of the genes modulated were significant, if for example, a gene has a fold change of 1.5 it may result in a smaller fold change within the protein expression, meaning this protein would not be identified as modulating in a proteomic investigation. The DIGE data is not directly comparable to the label free and SILAC work due to using a rat primary cell line rather than mouse, but it gives weight to the proteomic data gained with the other proteomic techniques as proteins with similar functions were also identified. This thesis also gives an insight in to the reliability and consistency between different methodologies investigating the same samples. As each technique has its own limitations it could be considered that the only way to fully eradicate these issues is to utilise a wide range of techniques and biological disciplines. However, with each method using slightly different analysis software, with varying criteria and investigating different areas, being able to correlate the data efficiently becomes the next challenge. A consideration into the implementation of each process should also be applied. Although the general

methodologies of these techniques are well used any small variation within the process could affect the results. This becomes important when comparing data with previous studies published. As there are no set guidelines on the experimental information to be included for publication, the depth of detail varies and it is often difficult to ascertain how similar techniques have been performed. Therefore, a Minimum Information About a Proteomic Experiment (MIAPE) for GE (gel electrophoresis) and GI (gel information) was proposed using the DIGE experiment from this study as the basis (Kenyani *et al.*, 2011).

To fully understand the phenomenon of the varying results from the different techniques there are several scenarios that need to be considered. Firstly, determining whether this is a biological or technical phenomenon is essential. Achieving this decision is not possible with this data alone, but the possible scenarios can be explored and with further work would eventually be concluded.

Firstly, the possibility that this phenomenon is due to the methodologies used will be considered. It has been clearly stated that there is a decidedly large gap in the knowledge between genes being transcribed and proteins being expressed. This could suggest that one of these techniques is not performing optimally. However, all of the techniques are regarded highly and used extensively for many areas of research. There are still, however, a number of limitations with all techniques. For proteomics in particular, the inability for identical sample sets to correlate with each other to a high degree when analysed via different methods provides good evidence for the case that results may be bias depending on the method chosen. This issue is not necessarily related to the techniques themselves but the bioinformatic tools

available. Not only are there a range of software products available, they also offer a range of statistical and normalisation options. If the same criteria are not applied to all data sets variation could occur. In some cases this is easily avoided as the criteria are easily changed but in other cases software programmes have their own algorithms for performing such techniques as normalisation which, without the appropriate computational background, can be difficult to ascertain and understand.

To further this area of research, a more extensive investigation in to the efficiency of both Progenesis and MaxQuant needs to be undertaken. These software packages have different criteria for picking which peptides to identify, for example, MaxQuant will use all shared peptides whereas Progenesis only uses unique peptides. Rerunning the data using only unique peptides in both cases may result in a better correlation between the two. A further point to be highlighted is that in MaxQuant only those proteins with peptides seen in all four repeats are processed to determine the fold change. As the samples are extremely complex it would be possible for different peptides to be identified in different runs which means that those that are seen in two or more would be removed but could be actual changes.

Another point that needs to be investigated rests on the choice to exclude those proteins without a reading in one of the two conditions. The assumption that if there is no reading means the peptide has not been detected and not that there is none of the corresponding protein in that sample will again have an effect on the results. This assumption is also applied to the genomic data removing all genes with 0 counts. These proteins or genes could be changing but without the corresponding data they will not be identified. This lends weight to the argument for the need for more mass

spectrometry runs per sample when dealing with complex samples. In theory, pre-fractionation should combat issues such as this but with the huge loss of sample in clean up steps the disadvantages outweigh the benefits.

In the DIGE investigation the inability to reach a firm conclusion was partly due to the range of pathway analysis results listing a variety of different pathways. Although these packages allow helpful visualisation, the results depend on whether the software has been manually curated or not. As many of the available packages are designed for genomic information they become difficult to use efficiently when adapting them to proteomic data. As genes often code for more than one protein a large number of pathways would be identified which may not be the pathways actually modulated.

When considering the biological aspect to the phenomenon there were again several scenarios. The first to consider is that, although it seems that a significant change has occurred at the gene level, these changes are not large enough to produce a great change at the protein level. The second involves analysing the cell line itself. The cell line used was differentiated before treatment to examine a system as close to that of *in vitro* muscle as possible. However, it is unclear whether treatment after differentiation will have resulted in a stunted effect. As these cells are an immortalised cell line they may not have the ability to increase in size once they have reached this point. This would explain the limited proteomic data. Further work using a primary cell line of differentiated muscle would allow evidence to be collected to support one of the potential theories. It would also be interesting to

perform a DIGE experiment on the C2C12 cell line as this investigation was able to determine a larger proteomic result.

Although this thesis was unable to reveal conclusive evidence on the mechanisms of the β-agonist salbutamol many potentially interesting results were obtained, such as the role of the MAPK pathway and the importance of cytoskeletal proteins in the hypertrophic response to salbutamol. The biological mechanisms of β-agonists are extremely interesting and a snap shot has been gained through the extensive research undertaken in this project. However, more importantly the potential of the techniques used are great and with the increasing capacity of sequencing, mass spectrometers and bioinformatical tools the growth of biological knowledge will continue. Along with these biological findings a number of insights in to the importance of experimental procedures and analysis have arisen.

Those in the field of genomics would argue that there are many limitations to the proteomic techniques available, leaving the mechanisms of a cell open for miss interpretation. Conversely, the proteomic researchers would counter that gene expression does not allow for any post transcriptional modifications that may occur as well as the fact that each gene can code for a number of proteins meaning the understanding of biology of the whole system is not complete. However, from this study it could be concluded that both fields need to work together towards fully understanding the process between transcription and protein expression. It is also essential that researchers in these areas work closely with bioinformatians to be able to gain the most from the data sets produced, but to also fully understand the bioinformatic analysis that is occurring, as without this knowledge many mistakes

could potential occur. Applying a system wide approach has highlighted the breadth of data that can be obtained, however, without continuing development of these techniques and their ability to interlink, a true representation of a whole system cannot be achieved. By efficiently combining the research of the all the "omics" areas as well as different techniques within each area our understanding into cellular biology is going to continue to grow to give great insights.

## References


Abe M, Sugiura T, Takahashi M, Ishii K, Shimoda M, Shirasuna K: **A novel function of CD82/KAI-1 on E-cadherin-mediated homophilic cellular adhesion of cancer cells.** *Cancer Lett* 2008, **266**:163-170.

Adams GR, Haddad F: **The relationships among IGF-1, DNA content, and protein accumulation during skeletal muscle hypertrophy.** *J Appl Physiol* 1996, **81**:2509-2516.

Ahlquist RP: **A study of the adrenotropic receptors.** *Am J Physiol* 1948, **153**:586-600.

Alban A, David SO, Bjorkesten L, Andersson C, Sloge E, Lewis S, Currie I: **A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard.** *Proteomics* 2003, **3**:36-44.

Ali M, Umstead TM, Haque R, Mikerov AN, Freeman WM, Floros J, Phelps DS: **Differences in the BAL proteome after Klebsiella pneumoniae infection in wild type and SP-A-/- mice.** *Proteome Sci* 2010, **8**:34.

Amanchy R, Kalume DE, Pandey A: **Stable isotope labeling with amino acids in cell culture (SILAC) for studying dynamics of protein abundance and posttranslational modifications.** *Sci STKE* 2005, **2005**:pl2.

Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11**:R106.

Anderson DB: **Never too late. Resolving the grief of suicide.** *J Psychosoc Nurs Ment Health Serv* 1991, **29**:29-31.

Andrés V, Walsh K: **Myogenin expression, cell cycle withdrawal, and phenotypic differentiation are temporally separable events that precede cell fusion upon myogenesis.** *J Cell Biol* 1996, **132**:657-666.

Aranguiz-Urroz P, Canales J, Copaja M, Troncoso R, Vicencio JM, Carrillo C, Lara H, Lavandero S, Diaz-Araya G: **Beta(2)-adrenergic receptor regulates cardiac fibroblast autophagy and collagen degradation.** *Biochim Biophys Acta* 2011, **1812**:23-31.

Arch JR, Ainsworth AT, Cawthorne MA, Piercy V, Sennitt MV, Thody VE, Wilson C, Wilson S: **Atypical beta-adrenoceptor on brown adipocytes as target for anti-obesity drugs.** *Nature* 1984, **309**:163-165.

Arvas M, Pakula T, Smit B, Rautio J, Koivistoinen H, Jouhten P, Lindfors E, Wiebe M, Penttila M, Saloheimo M: **Correlation of gene expression and protein production rate - a system wide study.** *BMC Genomics* 2011, **12**:616.

Awede BL, Thissen JP, Lebacq J: **Role of IGF-I and IGFBPs in the changes of mass and phenotype induced in rat soleus muscle by clenbuterol.** *Am J Physiol Endocrinol Metab* 2002, **282**:E31-37.

Baluk P, McDonald DM: **The beta 2-adrenergic receptor agonist formoterol reduces microvascular leakage by inhibiting endothelial gap formation.** *Am J Physiol*

Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B: **Quantitative mass spectrometry in proteomics: a critical review.** *Anal Bioanal Chem* 2007, **389**:1017-1031.

Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B: **Quantitative mass spectrometry in proteomics: a critical review.** *Anal Bioanal Chem* 2007, **389**:1017-1031.

Baracos VE: **Management of muscle wasting in cancer-associated cachexia: understanding gained from experimental studies.** *Cancer* 2001, **92**:1669-1677.

Barger G, Dale HH: **Chemical structure and sympathomimetic action of amines.** *J Physiol* 1910, **41**:19-59.

Barker BM, Kroll K, Vödisch M, Mazurie A, Kniemeyer O, Cramer RA: **Transcriptomic and proteomic analyses of the Aspergillus fumigatus hypoxia response using an oxygen-controlled fermenter.** *BMC Genomics* 2012, **13**:62.

Barr FA, Silljé HH, Nigg EA: **Polo-like kinases and the orchestration of cell division.** *Nat Rev Mol Cell Biol* 2004, **5**:429-440.

Bastepe M, Gunes Y, Perez-Villamil B, Hunzelman J, Weinstein LS, Jüppner H: **Receptor-mediated adenylyl cyclase activation through XLalpha(s), the extra-large variant of the stimulatory G protein alpha-subunit.** *Mol Endocrinol* 2002, **16**:1912-1919.

Becker CH, Bern M: **Recent developments in quantitative proteomics.** *Mutat Res* 2011, **722**:171-182.

Beermann DH, Butler WR, Hogue DE, Fishell VK, Dalrymple RH, Ricks CA, Scanes CG: **Cimaterol-induced muscle hypertrophy and altered endocrine status in lambs.** *J Anim Sci* 1987, **65**:1514-1524.

Bendall SC, Hughes C, Stewart MH, Doble B, Bhatia M, Lajoie GA: **Prevention of amino acid conversion in SILAC experiments with embryonic stem cells.** *Mol Cell Proteomics* 2008, **7**:1587-1597.

Bennett ST, Barnes C, Cox A, Davies L, Brown C: **Toward the 1,000 dollars human genome.** *Pharmacogenomics* 2005, **6**:373-382.

Bentley DR: **Whole-genome re-sequencing.** *Curr Opin Genet Dev* 2006, **16**:545-552.

Berry MP, Graham CM, McNab FW, Xu Z, Bloch SA, Oni T, Wilkinson KA, Banchereau R, Skinner J, Wilkinson RJ, et al: **An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis.** *Nature* 2010, **466**:973-977.

Bhavsar PK, Brand NJ, Felkin LE, Luther PK, Cullen ME, Yacoub MH, Barton PJ: **Clenbuterol induces cardiac myocyte hypertrophy via paracrine signalling and fibroblast-derived IGF-1.** *J Cardiovasc Transl Res* 2010, **3**:688-695.

Black DL: **Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology.** *Cell* 2000, **103**:367-370.

Blagoev B, Mann M: **Quantitative proteomics to study mitogen-activated protein kinases.** *Methods* 2006, **40**:243-250.

Blagoev B, Ong SE, Kratchmarova I, Mann M: **Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics.** *Nat Biotechnol* 2004, **22**:1139-1145.

Bloemen PG, van den Tweel MC, Henricks PA, Engels F, Kester MH, van de Loo PG, Blomjous FJ, Nijkamp FP: **Increased cAMP levels in stimulated neutrophils inhibit their adhesion to human bronchial epithelial cells.** *American Journal of Physiology - Lung Cellular and Molecular Physiology* 1997, **272**:L580-L587.

Bogoyevitch MA, Andersson MB, Gillespie-Brown J, Clerk A, Glennon PE, Fuller SJ, Sugden PH: **Adrenergic receptor stimulation of the mitogen-activated protein kinase cascade and cardiac hypertrophy.** *Biochem J* 1996, **314 ( Pt 1)**:115-121.

Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19**:185-193.

Bowden JJ, Sulakvelidze I, McDonald DM: **Inhibition of neutrophil and eosinophil adhesion to venules of rat trachea by beta 2-adrenergic agonist formoterol.** *J Appl Physiol* 1994, **77**:397-405.

Bowden JJ, Sulakvelidze I, McDonald DM: **Inhibition of neutrophil and eosinophil adhesion to venules of rat trachea by beta 2-adrenergic agonist formoterol.** *J Appl Physiol* 1994, **77**:397-405.

Bresnick E, Beakey JF, Levinson L, Segal MS: **Evaluation of Therapeutic Substances Employed for the Relief of Bronchospasm. V. Adrenergic Agents.** *J Clin Invest* 1949, **28**:1182-1189.

Briolant S, Almeras L, Belghazi M, Boucomont-Chapeaublanc E, Wurtz N, Fontaine A, Granjeaud S, Fusai T, Rogier C, Pradines B: **Plasmodium falciparum proteome changes in response to doxycycline treatment.** *Malar J* 2010, 9:141.

Bristow MR, Ginsburg R, Umans V, Fowler M, Minobe W, Rasmussen R, Zera P, Menlove R, Shah P, Jamieson S, et al.: **Beta 1- and beta 2-adrenergic-receptor subpopulations in nonfailing and failing human ventricular myocardium: coupling of both receptor subtypes to muscle contraction and selective beta 1-receptor down-regulation in heart failure.** *Circ Res* 1986, 59:297-309.

Brown DT, Izard T, Misteli T: **Mapping the interaction surface of linker histone H1(0) with the nucleosome of native chromatin in vivo.** *Nat Struct Mol Biol* 2006, 13:250-255.

Brushia RJ, Walsh DA: **Phosphorylase kinase: the complexity of its regulation is reflected in the complexity of its structure.** *Front Biosci* 1999, 4:D618-641.

Buckingham M: **Making muscle in mammals.** *Trends Genet* 1992, 8:144-148.

Buckingham M: **Skeletal muscle formation in vertebrates.** *Curr Opin Genet Dev* 2001, 11:440-448.

Bullard JH, Purdom E, Hansen KD, Dudoit S: **Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments.** *BMC Bioinformatics* 2010, 11:94.

Burniston JG, Kenyani J, Wastling JM, Burant CF, Qi NR, Koch LG, Britton SL: **Proteomic analysis reveals perturbed energy metabolism and elevated oxidative stress in hearts of rats with inborn low aerobic capacity.** *Proteomics* 2011, 11:3369-3379.

Burniston JG, McLean L, Beynon RJ, Goldspink DF: **Anabolic effects of a non-myotoxic dose of the beta2-adrenergic receptor agonist clenbuterol on rat plantaris muscle.** *Muscle Nerve* 2007, 35:217-223.

Burniston JG, Tan LB, Goldspink DF: **beta2-Adrenergic receptor stimulation in vivo induces apoptosis in the rat heart and soleus muscle.** *J Appl Physiol* 2005, 98:1379-1386.

Busquets S, Figueras MT, Fuster G, Almendro V, Moore-Carrasco R, Ametller E, Argiles JM, Lopez-Soriano FJ: **Anticachectic effects of formoterol: a drug for potential treatment of muscle wasting.** *Cancer Res* 2004, 64:6725-6731.

Buur T, Clausen T, Holmberg E, Johansson U, Waldeck B: **Desensitization by terbutaline of beta-adrenoceptors in the guinea-pig soleus muscle: biochemical alterations associated with functional changes.** *Br J Pharmacol* 1982, 76:313-317.

Cabane C, Englaro W, Yeow K, Ragno M, Dérijard B: **Regulation of C2C12 myogenic terminal differentiation by MKK3/p38alpha pathway.** *Am J Physiol Cell Physiol* 2003, **284**:C658-666.

Cairns SP, Dulhunty AF: **The effects of beta-adrenoceptor activation on contraction in isolated fast- and slow-twitch skeletal muscle fibres of the rat.** *Br J Pharmacol* 1993, **110**:1133-1141.

Cairns SP, Westerblad H, Allen DG: **Changes of tension and [Ca2+]i during beta-adrenoceptor activation of single, intact fibres from mouse skeletal muscle.** *Pflugers Arch* 1993, **425**:150-155.

Callister SJ, Barry RC, Adkins JN, Johnson ET, Qian WJ, Webb-Robertson BJ, Smith RD, Lipton MS: **Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics.** *J Proteome Res* 2006, **5**:277-286.

Cannon SC: **Sodium channel gating: no margin for error.** *Neuron* 2002, **34**:853-854.

Carbo N, Lopez-Soriano J, Tarrago T, Gonzalez O, Llovera M, Lopez-Soriano FJ, Argiles JM: **Comparative effects of beta2-adrenergic agonists on muscle waste associated with tumour growth.** *Cancer Lett* 1997, **115**:113-118.

Carbone R, Fré S, Iannolo G, Belleudi F, Mancini P, Pelicci PG, Torrisi MR, Di Fiore PP: **eps15 and eps15R are essential components of the endocytic pathway.** *Cancer Res* 1997, **57**:5498-5504.

Carstairs JR, Nimmo AJ, Barnes PJ: **Autoradiographic visualization of beta-adrenoceptor subtypes in human lung.** *Am Rev Respir Dis* 1985, **132**:541-547.

Carter WJ, Dang AQ, Faas FH, Lynch ME: **Effects of clenbuterol on skeletal muscle mass, body composition, and recovery from surgical stress in senescent rats.** *Metabolism* 1991, **40**:855-860.

Carter WJ, Lynch ME: **Comparison of the effects of salbutamol and clenbuterol on skeletal muscle mass and carcass composition in senescent rats.** *Metabolism* 1994, **43**:1119-1125.

Carvalho PC, Fischer JS, Chen EI, Yates JR, Barbosa VC: **PatternLab for proteomics: a tool for differential shotgun proteomics.** *BMC Bioinformatics* 2008, **9**:316.

Casey A, Constantin-Teodosiu D, Howell S, Hultman E, Greenhaff PL: **Creatine ingestion favorably affects performance and muscle metabolism during maximal exercise in humans.** *Am J Physiol* 1996, **271**:E31-37.

Celis JE, Gromov P: **2D protein electrophoresis: can it be perfected?** *Curr Opin Biotechnol* 1999, **10**:16-21.

Challiss RA, Leighton B, Wilson S, Thurlby PL, Arch JR: **An investigation of the beta-adrenoceptor that mediates metabolic responses to the novel agonist BRL28410 in rat soleus muscle.** *Biochem Pharmacol* 1988, **37**:947-950.

Chelius D, Bondarenko PV: **Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry.** *J Proteome Res* 2002, **1**:317-323.

Chen S, Smith DF: **Hop as an adaptor in the heat shock protein 70 (Hsp70) and hsp90 chaperone machinery.** *J Biol Chem* 1998, **273**:35194-35200.

Chen Y, Bornfeldt KE, Arner A, Jennische E, Malmqvist U, Uvelius B, Arnqvist HJ: **Increase in insulin-like growth factor I in hypertrophying smooth muscle.** *Am J Physiol* 1994, **266**:E224-229.

Chernogubova E, Hutchinson DS, Nedergaard J, Bengtsson T: **Alpha1- and beta1-adrenoceptor signaling fully compensates for beta3-adrenoceptor deficiency in brown adipocyte norepinephrine-stimulated glucose uptake.** *Endocrinology* 2005, **146**:2271-2284.

Chhabra ES, Higgs HN: **INF2 Is a WASP homology 2 motif-containing formin that severs actin filaments and accelerates both polymerization and depolymerization.** *J Biol Chem* 2006, **281**:26754-26767.

Chien KR, Knowlton KU, Zhu H, Chien S: **Regulation of cardiac gene expression during myocardial growth and hypertrophy: molecular studies of an adaptive physiologic response.** *FASEB J* 1991, **5**:3037-3046.

Choi SH, Kim TY, Park SG, Cha GH, Shin DW, Chai JY, Lee YH: **Proteomic analysis of Toxoplasma gondii KI-1 tachyzoites.** *Korean J Parasitol* 2010, **48**:195-201.

Choi S, Liu X, Li P, Akimoto T, Lee SY, Zhang M, Yan Z: **Transcriptional profiling in mouse skeletal muscle following a single bout of voluntary running: evidence of increased cell proliferation.** *J Appl Physiol* 2005, **99**:2406-2415.

Choo JJ, Horan MA, Little RA, Rothwell NJ: **Muscle wasting associated with endotoxemia in the rat: modification by the beta 2-adrenoceptor agonist clenbuterol.** *Biosci Rep* 1989, **9**:615-621.

Choo JJ, Horan MA, Little RA, Rothwell NJ: **Anabolic effects of clenbuterol on skeletal muscle are mediated by beta 2-adrenoceptor activation.** *Am J Physiol* 1992, **263**:E50-56.

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF: **Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1.** *Science* 1997, **277**:1996-2000.

Chung AS, Guan YJ, Yuan ZL, Albina JE, Chin YE: **Ankyrin repeat and SOCS box 3 (ASB3) mediates ubiquitination and degradation of tumor necrosis factor receptor II.** *Mol Cell Biol* 2005, **25:**4716-4726.

Clausen T, Everts ME: **Regulation of the Na,K-pump in skeletal muscle.** *Kidney Int* 1989, **35:**1-13.

Clausen T, Flatman JA: **The effect of catecholamines on Na-K transport and membrane potential in rat soleus muscle.** *J Physiol* 1977, **270:**383-414.

Cockman MD, Jones MB, Prenger MC, Sheldon RJ: **Magnetic resonance imaging of denervation-induced muscle atrophy: effects of clenbuterol in the rat.** *Muscle Nerve* 2001, **24:**1647-1658.

Collier TS, Randall SM, Sarkar P, Rao BM, Dean RA, Muddiman DC: **Comparison of stable-isotope labeling with amino acids in cell culture and spectral counting for relative quantification of protein expression.** *Rapid Commun Mass Spectrom* 2011, **25:**2524-2532.

Collier TS, Sarkar P, Franck WL, Rao BM, Dean RA, Muddiman DC: **Direct comparison of stable isotope labeling by amino acids in cell culture and spectral counting for quantitative proteomics.** *Anal Chem* 2010, **82:**8696-8702.

Collins S, Altschmied J, Herbsman O, Caron MG, Mellon PL, Lefkowitz RJ: **A cAMP response element in the beta 2-adrenergic receptor gene confers transcriptional autoregulation by cAMP.** *J Biol Chem* 1990, **265:**19330-19335.

Collins S, Bouvier M, Bolanowski MA, Caron MG, Lefkowitz RJ: **cAMP stimulates transcription of the beta 2-adrenergic receptor gene in response to short-term agonist exposure.** *Proc Natl Acad Sci U S A* 1989, **86:**4853-4857.

Collomp K, Le Panse B, Portier H, Lecoq AM, Jaffre C, Beaupied H, Richard O, Benhamou L, Courteix D, De Ceaurriz J: **Effects of acute salbutamol intake during a Wingate test.** *Int J Sports Med* 2005, **26:**513-517.

Colombo F, Gosselin H, El-Helou V, Calderone A: **Beta-adrenergic receptor-mediated DNA synthesis in neonatal rat cardiac fibroblasts proceeds via a phosphatidylinositol 3-kinase dependent pathway refractory to the antiproliferative action of cyclic AMP.** *J Cell Physiol* 2003, **195:**322-330.

Communal C, Singh K, Sawyer DB, Colucci WS: **Opposing effects of beta(1)- and beta(2)-adrenergic receptors on cardiac myocyte apoptosis : role of a pertussis toxin-sensitive G protein.** *Circulation* 1999, **100:**2210-2212.

Conolly ME, Greenacre JK: **The lymphocyte beta-adrenoceptor in normal subjects and patients with bronchial asthma: the effect of different forms of treatment on receptor function.** *J Clin Invest* 1976, **58:**1307-1316.

Coombs KM: **Quantitative proteomics of complex mixtures.** *Expert Rev Proteomics* 2011,

Cooper G: **Cytoskeletal networks and the regulation of cardiac contractility: microtubules, hypertrophy, and cardiac dysfunction.** *Am J Physiol Heart Circ Physiol* 2006, **291**:H1003-1014.

Costelli P, Garcia-Martinez C, Llovera M, Carbo N, Lopez-Soriano FJ, Agell N, Tessitore L, Baccino FM, Argiles JM: **Muscle protein waste in tumor-bearing rats is effectively antagonized by a beta 2-adrenergic agonist (clenbuterol). Role of the ATP-ubiquitin-dependent proteolytic pathway.** *J Clin Invest* 1995, **95**:2367-2372.

Cox J, Mann M: **Is proteomics the new genomics?** *Cell* 2007, **130**:395-398.

Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.** *Nat Biotechnol* 2008, **26**:1367-1372.

Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV, Mann M: **A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics.** *Nat Protoc* 2009, **4**:698-705.

Criswell DS, Powers SK, Herb RA: **Clenbuterol-induced fiber type transition in the soleus of adult rats.** *Eur J Appl Physiol Occup Physiol* 1996, **74**:391-396.

Crivelli G, Millet GP, Gremion G, Borrani F: **Effects of salbutamol on the contractile properties of human skeletal muscle before and after fatigue.** *Acta Physiol (Oxf)* 2011.

Cubría JC, Reguera R, Balaña-Fouce R, Ordóñez C, Ordóñez D: **Polyamine-mediated heart hypertrophy induced by clenbuterol in the mouse.** *J Pharm Pharmacol* 1998, **50**:91-96.

Cui Z, Chen X, Lu B, Park SK, Xu T, Xie Z, Xue P, Hou J, Hang H, Yates JR, Yang F: **Preliminary quantitative profile of differential protein expression between rat L6 myoblasts and myotubes by stable isotope labeling with amino acids in cell culture.** *Proteomics* 2009, **9**:1274-1292.

Cullum VA, Farmer JB, Jack D, Levy GP: **Salbutamol: a new, selective beta-adrenoceptive receptor stimulant.** *Br J Pharmacol* 1969, **35**:141-151.

Cutillas PR, Vanhaesebroeck B: **Quantitative profile of five murine core proteomes using label-free functional proteomics.** *Mol Cell Proteomics* 2007, **6**:1560-1573.

Dale HH: **On some physiological actions of ergot.** *J Physiol* 1906, **34**:163-206.

Darby AC, Hall N: **Fast forward genetics.** *Nat Biotechnol* 2008, **26**:1248-1249.

Datta PK, Chytil A, Gorska AE, Moses HL: **Identification of STRAP, a novel WD domain protein in transforming growth factor-beta signaling.** *J Biol Chem* 1998, **273**:34671-34674.

Dautel F, Kalkhof S, Trump S, Michaelson J, Beyer A, Lehmann I, von Bergen M: **DIGE-based protein expression analysis of B[a]P-exposed hepatoma cells reveals a complex stress response including alterations in oxidative stress, cell cycle control, and cytoskeleton motility at toxic and subacute concentrations.** *J Proteome Res* 2011, **10**:379-393.

Davis MA, Ireton RC, Reynolds AB: **A core function for p120-catenin in cadherin turnover.** *J Cell Biol* 2003, **163**:525-534.

Davis PB, Silski CL, Kercsmar CM, Infeld M: **Beta-adrenergic receptors on human tracheal epithelial cells in primary culture.** *Am J Physiol* 1990, **258**:C71-76.

de Angelis L, Zhao J, Andreucci JJ, Olson EN, Cossu G, McDermott JC: **Regulation of vertebrate myotome development by the p38 MAP kinase-MEF2 signaling pathway.** *Dev Biol* 2005, **283**:171-179.

de Hoog CL, Mann M: **Proteomics.** *Annu Rev Genomics Hum Genet* 2004, **5**:267-293.

Deeb SJ, D'Souza R, Cox J, Schmidt-Supprian M, Mann M: **Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles.** *Mol Cell Proteomics* 2012.

Delahunty C, Yates JR: **Protein identification using 2D-LC-MS/MS.** *Methods* 2005, **35**:248-255.

Delday MI, Maltin CA: **Clenbuterol increases the expression of myogenin but not myoD in immobilized rat muscles.** *Am J Physiol* 1997, **272**:E941-944.

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4**:P3.

Desaphy JF, Pierno S, De Luca A, Didonna P, Camerino DC: **Different ability of clenbuterol and salbutamol to block sodium channels predicts their therapeutic use in muscle excitability disorders.** *Mol Pharmacol* 2003, **63**:659-670.

DeSouza LV, Taylor AM, Li W, Minkoff MS, Romaschin AD, Colgan TJ, Siu KW: **Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues.** *J Proteome Res* 2008, **7**:3525-3534.

DeSouza L, Diehl G, Rodrigues MJ, Guo J, Romaschin AD, Colgan TJ, Siu KW: **Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry.** *J Proteome Res* 2005, **4**:377-386.

Devalia JL, Sapsford RJ, Rusznak C, Toumbis MJ, Davies RJ: **The effects of salmeterol and salbutamol on ciliary beat frequency of cultured human bronchial epithelial cells, in vitro.** *Pulm Pharmacol* 1992, **5**:257-263.

Dhingra V, Gupta M, Andacht T, Fu ZF: **New frontiers in proteomics research: a perspective.** *Int J Pharm* 2005, **299**:1-18.

Doherty KR, Demonbreun AR, Wallace GQ, Cave A, Posey AD, Heretis K, Pytel P, McNally EM: **The endocytic recycling protein EHD2 interacts with myoferlin to regulate myoblast fusion.** *J Biol Chem* 2008, **283**:20252-20260.

Dong MQ, Venable JD, Au N, Xu T, Park SK, Cociorva D, Johnson JR, Dillin A, Yates JR: **Quantitative mass spectrometry identifies insulin signaling targets in C. elegans.** *Science* 2007, **317**:660-663.

Donoghue P, Staunton L, Mullen E, Manning G, Ohlendieck K: **DIGE analysis of rat skeletal muscle proteins using nonionic detergent phase extraction of young adult versus aged gastrocnemius tissue.** *J Proteomics* 2010, **73**:1441-1453.

Downie D, Delday MI, Maltin CA, Sneddon AA: **Clenbuterol increases muscle fiber size and GATA-2 protein in rat skeletal muscle in utero.** *Mol Reprod Dev* 2008, **75**:785-794.

Drees F, Pokutta S, Yamada S, Nelson WJ, Weis WI: **Alpha-catenin is a molecular switch that binds E-cadherin-beta-catenin and regulates actin-filament assembly.** *Cell* 2005, **123**:903-915.

Drexler HC, Ruhs A, Konzer A, Mendler L, Bruckskotten M, Looso M, Günther S, Boettger T, Krüger M, Braun T: **On marathons and sprints: an integrated quantitative proteomics and transcriptomics analysis of differences between slow and fast muscle fibers.** *Mol Cell Proteomics* 2011.

Dumont EA, Hofstra L, van Heerde WL, van den Eijnde S, Doevendans PA, DeMuinck E, Daemen MA, Smits JF, Frederik P, Wellens HJ, et al: **Cardiomyocyte death induced by myocardial ischemia and reperfusion: measurement with recombinant human annexin-V in a mouse model.** *Circulation* 2000, **102**:1564-1568.

Duncan ND, Williams DA, Lynch GS: **Deleterious effects of chronic clenbuterol treatment on endurance and sprint exercise performance in rats.** *Clin Sci (Lond)* 2000, **98**:339-347.

Emerson CP: **Myogenesis and developmental control genes.** *Curr Opin Cell Biol* 1990, **2**:1065-1075.

Emery PW, Rothwell NJ, Stock MJ, Winter PD: **Chronic effects of beta 2-adrenergic agonists on body composition and protein synthesis in the rat.** *Biosci Rep* 1984, **4**:83-91.

Emorine LJ, Marullo S, Briend-Sutren MM, Patey G, Tate K, Delavier-Klutchko C, Strosberg AD: **Molecular characterization of the human beta 3-adrenergic receptor.** *Science* 1989, **245**:1118-1121.

Everts ME, Retterstol K, Clausen T: **Effects of adrenaline on excitation-induced stimulation of the sodium-potassium pump in rat skeletal muscle.** *Acta Physiol Scand* 1988, **134**:189-198.

Fan Y, Nikitina T, Zhao J, Fleury TJ, Bhattacharyya R, Bouhassira EE, Stein A, Woodcock CL, Skoultchi AI: **Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation.** *Cell* 2005, **123**:1199-1212.

Fitton AR, Berry MS, McGregor AD: **Preservation of denervated muscle form and function by clenbuterol in a rat model of peripheral nerve injury.** *J Hand Surg Br* 2001, **26**:335-346.

Florens L, Carozza MJ, Swanson SK, Fournier M, Coleman MK, Workman JL, Washburn MP: **Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors.** *Methods* 2006, **40**:303-311.

Foth BJ, Zhang N, Chaal BK, Sze SK, Preiser PR, Bozdech Z: **Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum.** *Mol Cell Proteomics* 2011.

Frankfurt OS, Krishan A: **Identification of apoptotic cells by formamide-induced dna denaturation in condensed chromatin.** *J Histochem Cytochem* 2001, **49**:369-378.

Fraterman S, Zeiger U, Khurana TS, Rubinstein NA, Wilm M: **Combination of peptide OFFGEL fractionation and label-free quantitation facilitated proteomics profiling of extraocular muscle.** *Proteomics* 2007, **7**:3404-3416.

Fraterman S, Zeiger U, Khurana TS, Wilm M, Rubinstein NA: **Quantitative proteomics profiling of sarcomere associated proteins in limb and extraocular muscle allotypes.** *Mol Cell Proteomics* 2007, **6**:728-737.

Frerichs O, Fansa H, Ziems P, Schneider W, Keilhoff G: **Regeneration of peripheral nerves after clenbuterol treatment in a rat model.** *Muscle Nerve* 2001, **24**:1687-1691.

Friso G, Olinares PD, van Wijk KJ: **The workflow for quantitative proteome analysis of chloroplast development and differentiation, chloroplast mutants, and protein interactions by spectral counting.** *Methods Mol Biol* 2011, **775**:265-282.

Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, Menzel C, Chen W, Li Y, Zeng R, Khaitovich P: **Estimating accuracy of RNA-Seq and microarrays with proteomics.** *BMC Genomics* 2009, **10**:161.

Fulton AB, Isaacs WB: **Titin, a huge, elastic sarcomeric protein with a probable role in morphogenesis.** *Bioessays* 1991, **13**:157-161.

Galant SP, Duriseti L, Underwood S, Allred S, Insel PA: **Beta adrenergic receptors of polymorphonuclear particulates in bronchial asthma.** *J Clin Invest* 1980, **65**:577-585.

Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP: **Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS.** *Proc Natl Acad Sci U S A* 2003, **100**:6940-6945.

Gokce E, Shuford CM, Franck WL, Dean RA, Muddiman DC: **Evaluation of normalization methods on GeLC-MS/MS label-free spectral counting data to correct for variation during proteomic workflows.** *J Am Soc Mass Spectrom* 2011, **22**:2199-2208.

Goldberg AL: **Functions of the proteasome: from protein degradation and immune surveillance to cancer therapy.** *Biochem Soc Trans* 2007, **35**:12-17.

Gorelick-Feldman J, Cohick W, Raskin I: **Ecdysteroids elicit a rapid Ca2+ flux leading to Akt activation and increased protein synthesis in skeletal muscle cells.** *Steroids* 2010, **75**:632-637.

Gotshalk LA, Kraemer WJ, Mendonca MA, Vingren JL, Kenny AM, Spiering BA, Hatfield DL, Fragala MS, Volek JS: **Creatine supplementation improves muscular performance in older women.** *Eur J Appl Physiol* 2008, **102**:223-231.

Graumann J, Hubner NC, Kim JB, Ko K, Moser M, Kumar C, Cox J, Schöler H, Mann M: **Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins.** *Mol Cell Proteomics* 2008, **7**:672-683.

Gréen A, Lönn A, Peterson KH, Ollinger K, Rundquist I: **Translocation of histone H1 subtypes between chromatin and cytoplasm during mitosis in normal human fibroblasts.** *Cytometry A* 2010, **77**:478-484.

Griffin NM, Yu J, Long F, Oh P, Shore S, Li Y, Koziol JA, Schnitzer JE: **Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis.** *Nat Biotechnol* 2010, **28**:83-89.

Griffin TJ, Gygi SP, Ideker T, Rist B, Eng J, Hood L, Aebersold R: **Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae.** *Mol Cell Proteomics* 2002, **1**:323-333.

Griffith M, Griffith OL, Mwenifumbo J, Goya R, Morrissy AS, Morin RD, Corbett R, Tang MJ, Hou YC, Pugh TJ, et al: **Alternative expression analysis by RNA sequencing.** *Nat Methods* 2010, **7**:843-847.

Gruhler A, Olsen JV, Mohammed S, Mortensen P, Faergeman NJ, Mann M, Jensen ON: **Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway.** *Mol Cell Proteomics* 2005, **4**:310-327.

Gu Y, Seidel A: **Influence of salbutamol and isoproterenol on the production of TNF and reactive oxygen species by bovine alveolar macrophages and calcitriol differentiated HL-60 cells.** *Immunopharmacol Immunotoxicol* 1996, **18**:115-128.

Guasconi V, Puri PL: **Chromatin: the interface between extrinsic cues and the epigenetic regulation of muscle regeneration.** *Trends Cell Biol* 2009, **19**:286-294.

Guo K, Wang J, Andrés V, Smith RC, Walsh K: **MyoD-induced expression of p21 inhibits cyclin-dependent kinase activity upon myocyte terminal differentiation.** *Mol Cell Biol* 1995, **15**:3823-3829.

Gutteridge A, Pir P, Castrillo JI, Charles PD, Lilley KS, Oliver SG: **Nutrient control of eukaryote cell growth: a systems biology study in yeast.** *BMC Biol* 2010, **8**:68.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R: **Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.** *Nat Biotechnol* 1999, **17**:994-999.

Gygi SP, Rochon Y, Franza BR, Aebersold R: **Correlation between protein and mRNA abundance in yeast.** *Mol Cell Biol* 1999, **19**:1720-1730.

Haberland M, Montgomery RL, Olson EN: **The many roles of histone deacetylases in development and physiology: implications for disease and therapy.** *Nat Rev Genet* 2009, **10**:32-42.

Hakimov HA, Walters S, Wright TC, Meidinger RG, Verschoor CP, Gadish M, Chiu DK, Strömvik MV, Forsberg CW, Golovan SP: **Application of iTRAQ to catalogue the skeletal muscle proteome in pigs and assessment of effects of gender and diet dephytinization.** *Proteomics* 2009, **9**:4000-4016.

Hamza N, Silvestre F, Mhetli M, Khemis IB, Dieu M, Raes M, Cahu C, Kestemont P: **Differential protein expression profile in the liver of pikeperch (Sander lucioperca) larvae fed with increasing levels of phospholipids.** *Comp Biochem Physiol Part D Genomics Proteomics* 2010, **5**:130-137.

Hanauske-Abel HM, Park MH, Hanauske AR, Popowicz AM, Lalande M, Folk JE: **Inhibition of the G1-S transition of the cell cycle by inhibitors of deoxyhypusine hydroxylation.** *Biochim Biophys Acta* 1994, **1221**:115-124.

Hanna MG, Stewart J, Schapira AH, Wood NW, Morgan-Hughes JA, Murray NM: **Salbutamol treatment in a patient with hyperkalaemic periodic paralysis due to a mutation in the skeletal muscle sodium channel gene (SCN4A).** *J Neurol Neurosurg Psychiatry* 1998, **65**:248-250.

Hannan NJ, Stephens AN, Rainczuk A, Hincks C, Rombauts LJ, Salamonsen LA: **2D-DiGE analysis of the human endometrial secretome reveals differences between receptive and nonreceptive states in fertile and infertile women.** *J Proteome Res* 2010, **9**:6256-6264.

Hegde PS, White IR, Debouck C: **Interplay of transcriptomics and proteomics.** *Curr Opin Biotechnol* 2003, **14**:647-651.

Hendrickson EL, Xia Q, Wang T, Leigh JA, Hackett M: **Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics.** *Analyst* 2006, **131**:1335-1341.

Herrera NM, Jr., Zimmerman AN, Dykstra DD, Thompson LV: **Clenbuterol in the prevention of muscle atrophy: a study of hindlimb-unweighted rats.** *Arch Phys Med Rehabil* 2001, **82**:930-934.

Hershko A, Ciechanover A: **The ubiquitin system.** *Annu Rev Biochem* 1998, **67**:425-479.

Hertel C, Stachelin M, Perkins JP: **Evidence for intravesicular beta-adrenergic receptors in membrane fractions from desensitized cells: binding of the hydrophilic ligand CGP-12177 only in the presence of alamethicin.** *J Cyclic Nucleotide Protein Phosphor Res* 1983, **9**:119-128.

183. Hesketh JE, Campbell GP, Lobley GE, Maltin CA, Acamovic F, Palmer RM: **Stimulation of actin and myosin synthesis in rat gastrocnemius muscle by clenbuterol; evidence for translational control.** *Comp Biochem Physiol C* 1992, **102**:23-27.

Higgs RE, Knierman MD, Gelfanova V, Butler JP, Hale JE: **Comprehensive label-free method for the relative quantification of proteins from biological samples.** *J Proteome Res* 2005, **4**:1442-1450.

Hinkle RT, Hodge KM, Cody DB, Sheldon RJ, Kobilka BK, Isfort RJ: **Skeletal muscle hypertrophy and anti-atrophy effects of clenbuterol are mediated by the beta2-adrenergic receptor.** *Muscle Nerve* 2002, **25**:729-734.

Hirst SJ, Martin JG, Bonacci JV, Chan V, Fixman ED, Hamid QA, Herszberg B, Lavoie JP, McVicker CG, Moir LM, et al: **Proliferative aspects of airway smooth muscle.** *J Allergy Clin Immunol* 2004, **114**:S2-17.

Hon JK, Yacoub MH: **Bridge to recovery with the use of left ventricular assist device and clenbuterol.** *Ann Thorac Surg* 2003, **75**:S36-41.

Hong HM, Song EJ, Oh E, Kabir MH, Lee C, Yoo YS: **Endothelin-1- and isoproterenol-induced differential protein expression and signaling pathway in HL-1 cardiomyocytes.** *Proteomics* 2011, **11**:283-297.

Horth P, Miller CA, Preckel T, Wenz C: **Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis.** *Mol Cell Proteomics* 2006, **5**:1968-1974.

Hoshino D, Yoshida Y, Holloway GP, Lally J, Hatta H, Bonen A: **Clenbuterol, a β2-adrenergic agonist, reciprocally alters PGC-1 alpha and RIP140 and reduces fatty acid and pyruvate oxidation in rat skeletal muscle.** *Am J Physiol Regul Integr Comp Physiol* 2012, **302**:R373-384.

Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R: **The Orbitrap: a new mass spectrometer.** *J Mass Spectrom* 2005, **40**:430-443.

Huey KA, Burdette S, Zhong H, Roy RR: **Early response of heat shock proteins to functional overload of the soleus and plantaris in rats and mice.** *Exp Physiol* 2010, **95**:1145-1155.

Huntgeburth M, Tiemann K, Shahverdyan R, Schlüter KD, Schreckenberg R, Gross ML, Mödersheim S, Caglayan E, Müller-Ehmsen J, Ghanem A, et al: **Transforming growth factor $\beta_1$ oppositely regulates the hypertrophic and contractile response to β-adrenergic stimulation in the heart.** *PLoS One* 2011, **6**:e26628.

Hutchinson DS, Bengtsson T: **alpha1A-adrenoceptors activate glucose uptake in L6 muscle cells through a phospholipase C-, phosphatidylinositol-3 kinase-, and atypical protein kinase C-dependent pathway.** *Endocrinology* 2005, **146**:901-912.

Hwang SI, Lundgren DH, Mayya V, Rezaul K, Cowan AE, Eng JK, Han DK: **Systematic characterization of nuclear proteome during apoptosis: a quantitative proteomic study by differential extraction and stable isotope labeling.** *Mol Cell Proteomics* 2006, **5**:1131-1145.

Ikeda F, Dikic I: **Atypical ubiquitin chains: new molecular signals. 'Protein Modifications: Beyond the Usual Suspects' review series.** *EMBO Rep* 2008, **9**:536-542.

Imai Y, Nakada A, Hashida R, Sugita Y, Tanaka T, Tsujimoto G, Matsumoto K, Akasawa A, Saito H, Oshida T: **Cloning and characterization of the highly expressed ETEA gene from blood cells of atopic dermatitis patients.** *Biochem Biophys Res Commun* 2002, **297**:1282-1290.

Ingalls CP, Barnes WS, Smith SB: **Interaction between clenbuterol and run training: effects on exercise performance and MLC isoform content.** *J Appl Physiol* 1996, **80**:795-801.

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M: **Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.** *Mol Cell Proteomics* 2005, **4**:1265-1272.

Jiang H, English AM: **Quantitative analysis of the yeast proteome by incorporation of isotopically labeled leucine.** *J Proteome Res* 2002, **1**:345-350.

Jin BF, He K, Wang HX, Wang J, Zhou T, Lan Y, Hu MR, Wei KH, Yang SC, Shen BF, Zhang XM: **Proteomic analysis of ubiquitin-proteasome effects: insight into the function of eukaryotic initiation factor 5A.** *Oncogene* 2003, **22**:4819-4830.

Johnson M: **Beta2-adrenoceptors: mechanisms of action of beta2-agonists.** *Paediatr Respir Rev* 2001, **2**:57-62.

Johnson M, Butchers PR, Coleman RA, Nials AT, Strong P, Sumner MJ, Vardey CJ, Whelan CJ: **The pharmacology of salmeterol.** *Life Sci* 1993, **52**:2131-2143.

Johnson M: **Effects of [beta]2-agonists on resident and infiltrating inflammatory cells.** *Journal of Allergy and Clinical Immunology* 2002, **110**:S282-S290.

Kariman K: **Beta-adrenergic receptor binding in lymphocytes from patients with asthma.** *Lung* 1980, **158**:41-51.

Karp NA, Kreil DP, Lilley KS: **Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis.** *Proteomics* 2004, **4**:1421-1432.

Kash JC, Tumpey TM, Proll SC, Carter V, Perwitasari O, Thomas MJ, Basler CF, Palese P, Taubenberger JK, García-Sastre A, et al: **Genomic analysis of increased host immune and cell death responses induced by 1918 influenza virus.** *Nature* 2006, **443**:578-581.

Kassel KM, Wyatt TA, Panettieri RA, Jr., Toews ML: **Inhibition of human airway smooth muscle cell proliferation by beta 2-adrenergic receptors and cAMP is PKA independent: evidence for EPAC involvement.** *Am J Physiol Lung Cell Mol Physiol* 2008, **294**:L131-138.

Kee HJ, Eom GH, Joung H, Shin S, Kim JR, Cho YK, Choe N, Sim BW, Jo D, Jeong MH, et al: **Activation of histone deacetylase 2 by inducible heat shock protein 70 in cardiac hypertrophy.** *Circ Res* 2008, **103**:1259-1269.

Kenrick KG, Margolis J: **Isoelectric focusing and gradient gel electrophoresis: a two-dimensional technique.** *Anal Biochem* 1970, **33**:204-207.

Kenyani J, Medina-Aunon JA, Martinez-Bartolomé S, Albar JP, Wastling JM, Jones AR: **A DIGE study on the effects of salbutamol on the rat muscle proteome - an exemplar of best practice for data sharing in proteomics.** *BMC Res Notes* 2011, **4**:86.

Kepler TB, Crosby L, Morgan KT: **Normalization and analysis of DNA microarray data by self-consistency and local regression.** *Genome Biol* 2002, **3**:RESEARCH0037.

Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU: **Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli.** *Cell* 2005, **122**:209-220.

Khassaf M, Child RB, McArdle A, Brodie DA, Esanu C, Jackson MJ: **Time course of responses of human skeletal muscle to oxidative stress induced by nondamaging exercise.** *J Appl Physiol* 2001, **90**:1031-1035.

Kierszniowska S, Walther D, Schulze WX: **Ratio-dependent significance thresholds in reciprocal 15N-labeling experiments as a robust tool in detection of candidate proteins responding to biological treatment.** *Proteomics* 2009, **9**:1916-1924.

Kim J, Eckhart AD, Eguchi S, Koch WJ: **Beta-adrenergic receptor-mediated DNA synthesis in cardiac fibroblasts is dependent on transactivation of the epidermal growth factor receptor and subsequent activation of extracellular signal-regulated kinases.** *J Biol Chem* 2002, **277**:32116-32123.

.Kirkpatrick DS, Gerber SA, Gygi SP: **The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications.** *Methods* 2005, **35**:265-273.

Kislinger T, Cox B, Kannan A, Chung C, Hu P, Ignatchenko A, Scott MS, Gramolini AO, Morris Q, Hallett MT, et al: **Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling.** *Cell* 2006, **125**:173-186.

Kitano H: **Computational systems biology.** *Nature* 2002, **420**:206-210.

Klopfleisch R, Gruber AD: **Transcriptome and proteome research in veterinary science: what is possible and what questions can be asked?** *ScientificWorldJournal* 2012, **2012**:254962.

Kobilka BK, Dixon RA, Frielle T, Dohlman HG, Bolanowski MA, Sigal IS, Yang-Feng TL, Francke U, Caron MG, Lefkowitz RJ: **cDNA for the human beta 2-adrenergic receptor: a protein with multiple membrane-spanning domains and encoded by a gene whose chromosomal location is shared with that of the receptor for platelet-derived growth factor.** *Proc Natl Acad Sci U S A* 1987, **84**:46-50.

Kosako H, Nagano K: **Quantitative phosphoproteomics strategies for understanding protein kinase-mediated signal transduction pathways.** *Expert Rev Proteomics* 2011, **8**:81-94.

Kreil DP, Karp NA, Lilley KS: **DNA microarray normalization methods can remove bias from differential protein expression analysis of 2D difference gel electrophoresis results.** *Bioinformatics* 2004, **20**:2026-2034.

Kubota T, Koga K, Araki H, Odajima H, Nishima S, Miyamoto H, Tanaka H, Sindou M: **[The relationships of mononuclear leukocyte beta-adrenergic receptors to aerobic capacity and exercise-induced asthma in asthmatic children].** *Arerugi* 2000, **49**:40-51.

Kugelberg E: **Adaptive transformation of rat soleus motor units during growth.** *J Neurol Sci* 1976, **27**:269-289.

Kunisada K, Hirota H, Fujio Y, Matsui H, Tani Y, Yamauchi-Takihara K, Kishimoto T: **Activation of JAK-STAT and MAP kinases by leukemia inhibitory factor through gp130 in cardiac myocytes.** *Circulation* 1996, **94**:2626-2632.

Lackner DH, Schmidt MW, Wu S, Wolf DA, Bahler J: **Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast.** *Genome Biol* 2012, **13**:R25.

Lam HT, Josserand J, Lion N, Girault HH: **Modeling the isoelectric focusing of peptides in an OFFGEL multicompartment cell.** *J Proteome Res* 2007, **6**:1666-1676.

Lands AM, Arnold A, McAuliff JP, Luduena FP, Brown TG, Jr.: **Differentiation of receptor systems activated by sympathomimetic amines.** *Nature* 1967, **214**:597-598.

Lee GE, Kim JH, Taylor M, Muller MT: **DNA methyltransferase 1-associated protein (DMAP1) is a co-repressor that stimulates DNA methylation globally and locally at sites of double strand break repair.** *J Biol Chem* 2010, **285**:37630-37640.

Leonhardt H, Page AW, Weier HU, Bestor TH: **A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei.** *Cell* 1992, **71**:865-873.

Li CJ, Li RW, Wang YH, Elsasser TH: **Pathway analysis identifies perturbation of genetic networks induced by butyrate in a bovine kidney epithelial cell line.** *Funct Integr Genomics* 2007, **7**:193-205.

Li E, Bestor TH, Jaenisch R: **Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.** *Cell* 1992, **69**:915-926.

Liao X, Che X, Zhao W, Zhang D, Bi T, Wang G: **The β-adrenoceptor antagonist, propranolol, induces human gastric cancer cell apoptosis and cell cycle arrest via inhibiting nuclear factor κB signaling.** *Oncol Rep* 2010, **24**:1669-1676.

Liaubet L, Lobjois V, Faraut T, Tircazes A, Benne F, Iannuccelli N, Pires J, Glénisson J, Robic A, Le Roy P, et al: **Genetic variability of transcript abundance in pig peri-mortem skeletal muscle: eQTL localized genes involved in stress response, cell death, muscle disorders and metabolism.** *BMC Genomics* 2011, **12**:548.

Littau VC, Burdick CJ, Allfrey VG, Mirsky SA: **The role of histones in the maintenance of chromatin structure.** *Proc Natl Acad Sci U S A* 1965, **54**:1204-1212.

Lluís F, Perdiguero E, Nebreda AR, Muñoz-Cánoves P: **Regulation of skeletal muscle gene expression by p38 MAP kinases.** *Trends Cell Biol* 2006, **16**:36-44.

Lohse MJ: **Molecular mechanisms of membrane receptor desensitization.** *Biochim Biophys Acta* 1993, **1179**:171-188.

Lulich KM, Goldie RG, Paterson JW: **Beta-adrenoceptor function in asthmatic bronchial smooth muscle.** *Gen Pharmacol* 1988, **19**:307-311.

Lynch GS, Hayes A, Campbell SP, Williams DA: **Effects of beta 2-agonist administration and exercise on contractile activation of skeletal muscle fibers.** *J Appl Physiol* 1996, **81**:1610-1618.

MacLennan PA, Edwards RH: **Effects of clenbuterol and propranolol on muscle mass. Evidence that clenbuterol stimulates muscle beta-adrenoceptors to induce hypertrophy.** *Biochem J* 1989, **264**:573-579.

MacKay, K, Qi, D & Jones, AR: **A methodolodyfor controlling false discovery rates in label-free quantitative proteomics through the use of multiple software packages.** *Submitted to Mollecular and Cellular Proteomics*

Mak JC, Nishikawa M, Barnes PJ: **Glucocorticosteroids increase beta 2-adrenergic receptor transcription in human lung.** *Am J Physiol* 1995, **268**:L41-46.

Mak JC, Nishikawa M, Shirasaki H, Miyayasu K, Barnes PJ: **Protective effects of a glucocorticoid on downregulation of pulmonary beta 2-adrenergic receptors in vivo.** *J Clin Invest* 1995, **96**:99-106.

Makarov A: **Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.** *Anal Chem* 2000, **72**:1156-1162.

Makarov A, Denisov E, Kholomeev A, Balschun W, Lange O, Strupat K, Horning S: **Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer.** *Anal Chem* 2006, **78**:2113-2120.

Malarkey K, Belham CM, Paul A, Graham A, McLees A, Scott PH, Plevin R: **The regulation of tyrosine kinase signalling pathways by growth factor and G-protein-coupled receptors.** *Biochem J* 1995, **309 ( Pt 2)**:361-375.

Maltin CA, Delday MI, Hay SM: **The effect of clenbuterol administration in utero and throughout lactation on pre- and post-natal muscle development in the rat.** *Growth Dev Aging* 1990, **54**:143-150.

Maltin CA, Delday MI, Watson JS, Heys SD, Nevison IM, Ritchie IK, Gibson PH: **Clenbuterol, a beta-adrenoceptor agonist, increases relative muscle strength in orthopaedic patients.** *Clin Sci (Lond)* 1993, **84**:651-654.

Maltin CA, Hay SM, Delday MI, Reeds PJ, Palmer RM: **Evidence that the hypertrophic action of clenbuterol on denervated rat muscle is not propranolol-sensitive.** *Br J Pharmacol* 1989, **96**:817-822.

Mancini A, Sirabella D, Zhang W, Yamazaki H, Shirao T, Krauss RS: **Regulation of myotube formation by the actin-binding factor drebrin.** *Skelet Muscle* 2011, **1**:36.

Mann K, Mann M: **In-depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos.** *Proteome Sci* 2011, **9**:7.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**:376-380.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays.** *Genome Res* 2008, **18**:1509-1517.

Markou T, Cieslak D, Gaitanaki C, Lazou A: **Differential roles of MAPKs and MSK1 signalling pathways in the regulation of c-Jun during phenylephrine-induced cardiac myocyte hypertrophy.** *Mol Cell Biochem* 2009, **322**:103-112.

Martineau L, Little RA, Rothwell NJ, Fisher MI: **Clenbuterol, a beta 2-adrenergic agonist, reverses muscle wasting due to scald injury in the rat.** *Burns* 1993, **19**:26-34.

Martinez-Esteso MJ, Selles-Marchart S, Vera-Urbina JC, Pedreno MA, Bru-Martinez R: **DIGE analysis of proteome changes accompanying large resveratrol production by grapevine (Vitis vinifera cv. Gamay) cell cultures in response to methyl-beta-cyclodextrin and methyl jasmonate elicitors.** *J Proteomics* 2011.

McDaneld TG, Hannon K, Moody DE: **Ankyrin repeat and SOCS box protein 15 regulates protein synthesis in skeletal muscle.** *Am J Physiol Regul Integr Comp Physiol* 2006, **290**:R1672-1682.

McElligott MA, Mulder JE, Chaung LY, Barreto A, Jr.: **Clenbuterol-induced muscle growth: investigation of possible mediation by insulin.** *Am J Physiol* 1987, **253**:E370-375.

McMillan DN, Noble BS, Maltin CA: **The effect of the beta-adrenergic agonist clenbuterol on growth and protein metabolism in rat muscle cell cultures.** *J Anim Sci* 1992, **70**:3014-3023.

Menon B, Johnson JN, Ross RS, Singh M, Singh K: **Glycogen synthase kinase-3beta plays a pro-apoptotic role in beta-adrenergic receptor-stimulated apoptosis in adult rat ventricular myocytes: Role of beta1 integrins.** *J Mol Cell Cardiol* 2007, **42**:653-661.

Mersmann HJ: **Overview of the effects of beta-adrenergic receptor agonists on animal growth including mechanisms of action.** *J Anim Sci* 1998, **76**:160-172.

Meszaros JG, Gonzalez AM, Endo-Mochizuki Y, Villegas S, Villarreal F, Brunton LL: **Identification of G protein-coupled signaling pathways in cardiac fibroblasts: cross talk between G(q) and G(s).** *Am J Physiol Cell Physiol* 2000, **278**:C154-162.

Meyerson M, Gabriel S, Getz G: **Advances in understanding cancer genomes through second-generation sequencing.** *Nat Rev Genet* 2010, **11**:685-696.

Moresi V, Carrer M, Grueter CE, Rifki OF, Shelton JM, Richardson JA, Bassel-Duby R, Olson EN: **Histone deacetylases 1 and 2 regulate autophagy flux and skeletal muscle homeostasis in mice.** *Proc Natl Acad Sci U S A* 2012, **109**:1649-1654.

Morozova O, Marra MA: **Applications of next-generation sequencing technologies in functional genomics.** *Genomics* 2008, **92**:255-264.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**:621-628.

Mortusewicz O, Schermelleh L, Walter J, Cardoso MC, Leonhardt H: **Recruitment of DNA methyltransferase I to DNA repair sites.** *Proc Natl Acad Sci U S A* 2005, **102**:8905-8909.

Mujika I, Padilla S: **Creatine supplementation as an ergogenic aid for sports performance in highly trained athletes: a critical review.** *Int J Sports Med* 1997, **18**:491-496.

Mukherjee A, Murray RD, Shalet SM: **Impact of growth hormone status on body composition and the skeleton.** *Horm Res* 2004, **62 Suppl 3**:35-41.

Munday DC, Surtees R, Emmott E, Dove BK, Digard P, Barr JN, Whitehouse A, Matthews D, Hiscox JA: **Using SILAC and quantitative proteomics to investigate the interactions between viral and host proteomes.** *Proteomics* 2012, **12**:666-672.

Munoz NM, Rabe KF, Vita AJ, McAllister K, Mayer D, Weiss M, Leff AR: **Paradoxical blockade of beta adrenergically mediated inhibition of stimulated eosinophil secretion by salmeterol.** *J Pharmacol Exp Ther* 1995, **273**:850-854.

Murgia M, Serrano AL, Calabria E, Pallafacchina G, Lomo T, Schiaffino S: **Ras is involved in nerve-activity-dependent regulation of muscle genes.** *Nat Cell Biol* 2000, **2**:142-147.

Nagira M, Imai T, Ishikawa I, Uwabe KI, Yoshie O: **Mouse homologue of C33 antigen (CD82), a member of the transmembrane 4 superfamily: complementary DNA, genomic structure, and expression.** *Cell Immunol* 1994, **157**:144-157.

Navegantes LC, Resano NM, Migliorini RH, Kettelhut IC: **Catecholamines inhibit Ca(2+)-dependent proteolysis in rat skeletal muscle through beta(2)-adrenoceptors and cAMP.** *Am J Physiol Endocrinol Metab* 2001, **281**:E449-454.

Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, Lee A, van Sluyter SC, Haynes PA: **Less label, more free: approaches in label-free quantitative mass spectrometry.** *Proteomics* 2011, **11**:535-553.

Nelson MM, Jones AR, Carmen JC, Sinai AP, Burchmore R, Wastling JM: **Modulation of the host cell proteome by the intracellular apicomplexan parasite Toxoplasma gondii.** *Infect Immun* 2008, **76**:828-844.

Nesvizhskii AI, Keller A, Kolker E, Aebersold R: **A statistical model for identifying proteins by tandem mass spectrometry.** *Anal Chem* 2003, **75**:4646-4658.

Nevzorova J, Bengtsson T, Evans BA, Summers RJ: **Characterization of the beta-adrenoceptor subtype involved in mediation of glucose transport in L6 cells.** *Br J Pharmacol* 2002, **137**:9-18.

Nevzorova J, Evans BA, Bengtsson T, Summers RJ: **Multiple signalling pathways involved in beta2-adrenoceptor-mediated glucose uptake in rat skeletal muscle cells.** *Br J Pharmacol* 2006, **147**:446-454.

Ngala RA, O'Dowd J, Wang SJ, Agarwal A, Stocker C, Cawthorne MA, Arch JR: **Metabolic responses to BRL37344 and clenbuterol in soleus muscle and**

**C2C12 cells via different atypical pharmacologies and beta2-adrenoceptor mechanisms.** *Br J Pharmacol* 2008, **155**:395-406.

Nirmalan N, Sims PF, Hyde JE: **Quantitative proteomics of the human malaria parasite Plasmodium falciparum and its application to studies of development and inhibition.** *Mol Microbiol* 2004, **52**:1187-1199.

Nishimura K, Murozumi K, Shirahata A, Park MH, Kashiwagi K, Igarashi K: **Independent roles of eIF5A and polyamines in cell proliferation.** *Biochem J* 2005, **385**:779-785.

O'Connor BJ, Aikman SL, Barnes PJ: **Tolerance to the nonbronchodilator effects of inhaled beta 2-agonists in asthma.** *N Engl J Med* 1992, **327**:1204-1208.

O'Connor BJ, Fuller RW, Barnes PJ: **Nonbronchodilator effects of inhaled beta 2 agonists. Greater protection against adenosine monophosphate- than methacholine-induced bronchoconstriction in asthma.** *Am J Respir Crit Care Med* 1994, **150**:381-387.

O'Farrell PH: **High resolution two-dimensional electrophoresis of proteins.** *Journal of Biological Chemistry* 1975, **250**:4007-4021.

Oe M, Ohnishi-Kameyama M, Nakajima I, Muroya S, Shibata M, Ojima K, Kushibiki S, Chikuni K: **Proteome analysis of whole and water-soluble proteins in masseter and semitendinosus muscles of Holstein cows.** *Anim Sci J* 2011, **82**:181-186.

Oishi Y, Imoto K, Ogata T, Taniguchi K, Matsumoto H, Roy RR: **Clenbuterol induces expression of multiple myosin heavy chain isoforms in rat soleus fibres.** *Acta Physiol Scand* 2002, **176**:311-318.

Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG: **Comparison of label-free methods for quantifying human proteins by shotgun proteomics.** *Mol Cell Proteomics* 2005, **4**:1487-1502.

Olsen JV, de Godoy LM, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M: **Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap.** *Mol Cell Proteomics* 2005, **4**:2010-2021.

Olson EN, Williams RS: **Calcineurin signaling and muscle remodeling.** *Cell* 2000, **101**:689-692.

Onaran HO, Costa T, Rodbard D: **Beta gamma subunits of guanine nucleotide-binding proteins and regulation of spontaneous receptor activity: thermodynamic model for the interaction between receptors and guanine nucleotide-binding protein subunits.** *Mol Pharmacol* 1993, **43**:245-256.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: **Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.** *Mol Cell Proteomics* 2002, **1**:376-386.

Ong SE, Mann M: **A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC).** *Nat Protoc* 2006, **1**:2650-2660.

Ono M, Handa K, Withers DA, Hakomori S: **Motility inhibition and apoptosis are induced by metastasis-suppressing gene product CD82 and its analogue CD9, with concurrent glycosylation.** *Cancer Res* 1999, **59**:2335-2339.

Oshlack A, Robinson MD, Young MD: **From RNA-seq reads to differential expression results.** *Genome Biol* 2010, **11**:220.

Ostrom RS, Naugle JE, Hase M, Gregorian C, Swaney JS, Insel PA, Brunton LL, Meszaros JG: **Angiotensin II enhances adenylyl cyclase signaling via Ca2+/calmodulin. Gq-Gs cross-talk regulates collagen production in cardiac fibroblasts.** *J Biol Chem* 2003, **278**:24461-24468.

Paasch U, Heidenreich F, Pursche T, Kuhlisch E, Kettner K, Grunewald S, Kratzsch J, Dittmar G, Glander HJ, Hoflack B, Kriegel TM: **Identification of increased amounts of eppin protein complex components in sperm cells of diabetic and obese individuals by difference gel electrophoresis.** *Mol Cell Proteomics* 2011.

Pan C, Gnad F, Olsen JV, Mann M: **Quantitative phosphoproteome analysis of a mouse liver cell line reveals specificity of phosphatase inhibitors.** *Proteomics* 2008, **8**:4534-4546.

Panchaud A, Affolter M, Moreillon P, Kussmann M: **Experimental and computational approaches to quantitative proteomics: status quo and outlook.** *J Proteomics* 2008, **71**:19-33.

Pancholi V: **Multifunctional alpha-enolase: its role in diseases.** *Cell Mol Life Sci* 2001, **58**:902-920.

Park KS, Mohapatra DP, Misonou H, Trimmer JS: **Graded regulation of the Kv2.1 potassium channel by variable phosphorylation.** *Science* 2006, **313**:976-979.

Park T, Yi SG, Kang SH, Lee S, Lee YS, Simon R: **Evaluation of normalization methods for microarray data.** *BMC Bioinformatics* 2003, **4**:33.

Parreiras-e-Silva LT, Luchessi AD, Reis RI, Oliver C, Jamur MC, Ramos RG, Oliveira EB, Curi R, Costa-Neto CM: **Evidences of a role for eukaryotic translation initiation factor 5A (eIF5A) in mouse embryogenesis and cell differentiation.** *J Cell Physiol* 2010, **225**:500-505.

Patel VJ, Thalassinos K, Slade SE, Connolly JB, Crombie A, Murrell JC, Scrivens JH: **A comparison of labeling and label-free mass spectrometry-based proteomics approaches.** *J Proteome Res* 2009, **8**:3752-3759.

Paul AS, Pollard TD: **Review of the mechanism of processive actin filament elongation by formins.** *Cell Motil Cytoskeleton* 2009, **66**:606-617.

Pearen MA, Ryall JG, Lynch GS, Muscat GE: **Expression profiling of skeletal muscle following acute and chronic beta2-adrenergic stimulation: implications for hypertrophy, metabolism and circadian rhythm.** *BMC Genomics* 2009, **10**:448.

Pearson K: **DETERMINATION OF THE COEFFICIENT OF CORRELATION.** *Science* 1909, **30**:23-25.

Pelkonen O, Tuimala R, Kauppila A: **Placental transfer of clenbuterol early in human pregnancy.** *Eur J Clin Pharmacol* 1982, **22**:403-406.

Pellegrino MA, D'Antona G, Bortolotto S, Boschi F, Pastoris O, Bottinelli R, Polla B, Reggiani C: **Clenbuterol antagonizes glucocorticoid-induced atrophy and fibre type transformation in mice.** *Exp Physiol* 2004, **89**:89-100.

Penn RB: **Embracing emerging paradigms of G protein-coupled receptor agonism and signaling to address airway smooth muscle pathobiology in asthma.** *Naunyn Schmiedebergs Arch Pharmacol* 2008, **378**:149-169.

Pennica D, Wood WI, Chien KR: **Cardiotrophin-1: a multifunctional cytokine that signals via LIF receptor-gp 130 dependent pathways.** *Cytokine Growth Factor Rev* 1996, **7**:81-91.

Perdiguero E, Ruiz-Bonilla V, Gresh L, Hui L, Ballestar E, Sousa-Victor P, Baeza-Raja B, Jardí M, Bosch-Comas A, Esteller M, et al: **Genetic analysis of p38 MAP kinases in myogenesis: fundamental role of p38alpha in abrogating myoblast proliferation.** *EMBO J* 2007, **26**:1245-1256.

Petrak J, Ivanek R, Toman O, Cmejla R, Cmejlova J, Vyoral D, Zivny J, Vulpe CD: **Deja vu in proteomics. A hit parade of repeatedly identified differentially expressed proteins.** *Proteomics* 2008, **8**:1744-1749.

Pieper FR, Slobbe RL, Ramaekers FC, Cuypers HT, Bloemendal H: **Upstream regions of the hamster desmin and vimentin genes regulate expression during in vitro myogenesis.** *EMBO J* 1987, **6**:3611-3618.

Podwojski K, Eisenacher M, Kohl M, Turewicz M, Meyer HE, Rahnenführer J, Stephan C: **Peek a peak: a glance at statistics for quantitative label-free proteomics.** *Expert Rev Proteomics* 2010, **7**:249-261.

Polla B, Bottinelli R, Sandoli D, Sardi C, Reggiani C: **Cortisone-induced changes in myosin heavy chain distribution in respiratory and hindlimb muscles.** *Acta Physiol Scand* 1994, **151**:353-361.

Polla B, Cappelli V, Morello F, Pellegrino MA, Boschi F, Pastoris O, Reggiani C: **Effects of the beta(2)-agonist clenbuterol on respiratory and limb muscles of weaning rats.** *Am J Physiol Regul Integr Comp Physiol* 2001, **280**:R862-869.

Porteus B, Kocharunchitt C, Nilsson RE, Ross T, Bowman JP: **Utility of gel-free, label-free shotgun proteomics approaches to investigate microorganisms.** *Appl Microbiol Biotechnol* 2011, **90**:407-416.

Post SR, Hammond HK, Insel PA: **Beta-adrenergic receptors and receptor signaling in heart failure.** *Annu Rev Pharmacol Toxicol* 1999, **39**:343-360.

Powell DW, Weaver CM, Jennings JL, McAfee KJ, He Y, Weil PA, Link AJ: **Cluster analysis of mass spectrometry data reveals a novel component of SAGA.** *Mol Cell Biol* 2004, **24**:7249-7259.

Price AH, Clissold SP: **Salbutamol in the 1980s. A reappraisal of its clinical efficacy.** *Drugs* 1989, **38**:77-122.

Puri PL, Wu Z, Zhang P, Wood LD, Bhakta KS, Han J, Feramisco JR, Karin M, Wang JY: **Induction of terminal differentiation by constitutive activation of p38 MAP kinase in human rhabdomyosarcoma cells.** *Genes Dev* 2000, **14**:574-584.

Qi D, Brownridge P, Xia D, Mackay K, Gonzalez-Galarza FF, Kenyani J, Harman V, Beynon RJ, Jones AR: **A Software Toolkit and Interface for Performing Stable Isotope Labeling and Top3 Quantification Using Progenesis LC-MS.** *OMICS* 2012, **16**:489-495.

Raddatz K, Albrecht D, Hochgrafe F, Hecker M, Gotthardt M: **A proteome map of murine heart and skeletal muscle.** *Proteomics* 2008, **8**:1885-1897.

Reeds PJ, Hay SM, Dorward PM, Palmer RM: **The effect of beta-agonists and antagonists on muscle growth and body composition of young rats (Rattus sp.).** *Comp Biochem Physiol C* 1988, **89**:337-341.

Reeds PJ, Hay SM, Dorwood PM, Palmer RM: **Stimulation of muscle growth by clenbuterol: lack of effect on muscle protein biosynthesis.** *Br J Nutr* 1986, **56**:249-258.

Ricart-Firinga C, Stevens L, Canu MH, Nemirovskaya TL, Mounier Y: **Effects of beta(2)-agonist clenbuterol on biochemical and contractile properties of unloaded soleus fibers of rat.** *Am J Physiol Cell Physiol* 2000, **278**:C582-588.

Roatta S, Arendt-Nielsen L, Farina D: **Sympathetic-induced changes in discharge rate and spike-triggered average twitch torque of low-threshold motor units in humans.** *J Physiol* 2008, **586**:5561-5574.

Roatta S, Farina D: **Sympathetic actions on the skeletal muscle.** *Exerc Sport Sci Rev* 2010, **38**:31-35.

Robinson JS, Graham TR, Emr SD: **A putative zinc finger protein, Saccharomyces cerevisiae Vps18p, affects late Golgi functions required for vacuolar protein sorting and efficient alpha-factor prohormone maturation.** *Mol Cell Biol* 1991, **11**:5813-5824.

Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26**:139-140.

Robinson MD, Oshlack A: **A scaling normalization method for differential expression analysis of RNA-seq data.** *Genome Biol* 2010, **11**:R25.

Rodriguez OC, Schaefer AW, Mandato CA, Forscher P, Bement WM, Waterman-Storer CM: **Conserved microtubule-actin interactions in cell movement and morphogenesis.** *Nat Cell Biol* 2003, **5**:599-609.

Ros A, Faupel M, Mees H, Oostrum J, Ferrigno R, Reymond F, Michel P, Rossier JS, Girault HH: **Protein purification by Off-Gel electrophoresis.** *Proteomics* 2002, **2**:151-156.

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, et al: **Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.** *Mol Cell Proteomics* 2004, **3**:1154-1169.

ROSS SB: **IN VIVO INACTIVATION OF CATECHOLAMINES IN MICE.** *Acta Pharmacol Toxicol (Copenh)* 1963, **20**:267-273.

Rothwell NJ, Stock MJ: **Modification of body composition by clenbuterol in normal and dystrophic (mdx) mice.** *Biosci Rep* 1985, **5**:755-760.

Rothwell NJ, Stock MJ, Sudera DK: **Changes in tissue blood flow and beta-receptor density of skeletal muscle in rats treated with the beta2-adrenoceptor agonist clenbuterol.** *Br J Pharmacol* 1987, **90**:601-607.

Ryall JG, Schertzer JD, Alabakis TM, Gehrig SM, Plant DR, Lynch GS: **Intramuscular beta2-agonist administration enhances early regeneration and functional repair in rat skeletal muscle after myotoxic injury.** *J Appl Physiol* 2008, **105**:165-172.

Ryu S, Gallis B, Goo YA, Shaffer SA, Radulovic D, Goodlett DR: **Comparison of a label-free quantitative proteomic method based on peptide ion current area to the isotope coded affinity tag method.** *Cancer Inform* 2008, **6**:243-255.

Salem M, Kenney PB, Rexroad CE, Yao J: **Proteomic signature of muscle atrophy in rainbow trout.** *J Proteomics* 2010, **73**:778-789.

Sanders VM, Baker RA, Ramer-Quinn DS, Kasprowicz DJ, Fuchs BA, Street NE: **Differential expression of the beta2-adrenergic receptor by Th1 and Th2 clones: implications for cytokine production and B cell help.** *J Immunol* 1997, **158**:4200-4210.

Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proc Natl Acad Sci U S A* 1977, **74**:5463-5467.

Sato S, Nomura S, Kawano F, Tanihata J, Tachiyashiki K, Imaizumi K: **Effects of the beta2-agonist clenbuterol on beta1- and beta2-adrenoceptor mRNA expressions of rat skeletal and left ventricle muscles.** *J Pharmacol Sci* 2008, **107**:393-400.

Saunders PA, Cooper JA, Roodell MM, Schroeder DA, Borchert CJ, Isaacson AL, Schendel MJ, Godfrey KG, Cahill DR, Walz AM, et al: **Quantification of active caspase 3 in apoptotic cells.** *Anal Biochem* 2000, **284**:114-124.

Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**:467-470.

Schmid A, Renaud JF, Lazdunski M: **Short term and long term effects of beta-adrenergic effectors and cyclic AMP on nitrendipine-sensitive voltage-dependent Ca2+ channels of skeletal muscle.** *J Biol Chem* 1985, **260**:13041-13046.

Schmidt F, Strozynski M, Salus SS, Nilsen H, Thiede B: **Rapid determination of amino acid incorporation by stable isotope labeling with amino acids in cell culture (SILAC).** *Rapid Commun Mass Spectrom* 2007, **21**:3919-3926.

Schmidt MW, Houseman A, Ivanov AR, Wolf DA: **Comparative proteomic and transcriptomic profiling of the fission yeast Schizosaccharomyces pombe.** *Mol Syst Biol* 2007, **3**:79.

Schoenfeld N, Bauer MK, Grimm S: **The metastasis suppressor gene C33/CD82/KAI1 induces apoptosis through reactive oxygen intermediates.** *FASEB J* 2004, **18**:158-160.

Schulze WX, Usadel B: **Quantitation in mass-spectrometry-based proteomics.** *Annu Rev Plant Biol* 2010, **61**:491-516.

Schulze WX, Usadel B: **Quantitation in mass-spectrometry-based proteomics.** *Annu Rev Plant Biol* 2010, **61**:491-516.

Schuster SC: **Next-generation sequencing transforms today's biology.** *Nat Methods* 2008, **5**:16-18.

Segal MS, Beakey JF: **The use of isuprel for the management of bronchial asthma.** *Bull New Engl Med Cent* 1947, **9**:62-67.

Seldon PM, Barnes PJ, Meja K, Giembycz MA: **Suppression of lipopolysaccharide-induced tumor necrosis factor-alpha generation from human peripheral blood monocytes by inhibitors of phosphodiesterase 4: interaction with stimulants of adenylyl cyclase.** *Mol Pharmacol* 1995, 48:747-757.

Seldon PM, Stevens DA, Adcock IM, O'Connor BJ, Barnes PJ, Giembycz MA: **Albuterol does not antagonize the inhibitory effect of dexamethasone on monocyte cytokine release.** *Am J Respir Crit Care Med* 1998, **157**:803-809.

Seth A, Gonzalez FA, Gupta S, Raden DL, Davis RJ: **Signal transduction within the nucleus by mitogen-activated protein kinase.** *J Biol Chem* 1992, **267**:24796-24804.

Shadforth IP, Dunkley TP, Lilley KS, Bessant C: **i-Tracker: for quantitative proteomics using iTRAQ.** *BMC Genomics* 2005, **6**:145.

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM: **Accurate multiplex polony sequencing of an evolved bacterial genome.** *Science* 2005, **309**:1728-1732.

Shi H, Zeng C, Ricome A, Hannon KM, Grant AL, Gerrard DE: **Extracellular signal-regulated kinase pathway is differentially involved in beta-agonist-induced hypertrophy in slow and fast muscles.** *Am J Physiol Cell Physiol* 2007, **292**:C1681-1689.

Shiio Y, Aebersold R: **Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry.** *Nat Protoc* 2006, **1**:139-145.

Sichletidis L, Daskalopoulou E, Kyriazis G, Kosmidou I, Koupidou S, Pechlivanidis T, Chloros D: **Comparative efficacy of salbutamol and salmeterol in exercise-induced asthma.** *J Int Med Res* 1993, **21**:81-88.

Sichletidis L, Daskalopoulou E, Kyriazis G, Kosmidou I, Koupidou S, Pechlivanidis T, Chloros D: **Comparative efficacy of salbutamol and salmeterol in exercise-induced asthma.** *J Int Med Res* 1993, **21**:81-88.

Sillence MN, Matthews ML: **Classical and atypical binding sites for beta-adrenoceptor ligands and activation of adenylyl cyclase in bovine skeletal muscle and adipose tissue membranes.** *Br J Pharmacol* 1994, **111**:866-872.

Silva JC, Denny R, Dorschel C, Gorenstein MV, Li GZ, Richardson K, Wall D, Geromanos SJ: **Simultaneous qualitative and quantitative analysis of the Escherichia coli proteome: a sweet tale.** *Mol Cell Proteomics* 2006, **5**:589-607.

Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ: **Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition.** *Mol Cell Proteomics* 2006, **5**:144-156.

Singh K, Xiao L, Remondino A, Sawyer DB, Colucci WS: **Adrenergic regulation of cardiac myocyte apoptosis.** *J Cell Physiol* 2001, **189**:257-265.

Slack JP, Grupp IL, Luo W, Kranias EG: **Phospholamban ablation enhances relaxation in the murine soleus.** *Am J Physiol* 1997, **273**:C1-6.

Sneddon AA, Delday MI, Maltin CA: **Amelioration of denervation-induced atrophy by clenbuterol is associated with increased PKC-alpha activity.** *Am J Physiol Endocrinol Metab* 2000, **279**:E188-195.

Sneddon AA, Delday MI, Steven J, Maltin CA: **Elevated IGF-II mRNA and phosphorylation of 4E-BP1 and p70(S6k) in muscle showing clenbuterol-induced anabolism.** *Am J Physiol Endocrinol Metab* 2001, **281**:E676-682.

Soic-Vranic T, Bobinac D, Bajek S, Jerkovic R, Malnar-Dragojevic D, Nikolic M: **Effect of salbutamol on innervated and denervated rat soleus muscle.** *Braz J Med Biol Res* 2005, **38**:1799-1805.

Soma LR, Uboh CE, Guan F, Luo Y, Teleis D, Runbo L, Birks EK, Tsang DS, Habecker P: **Tissue distribution of clenbuterol in the horse.** *J Vet Pharmacol Ther* 2004, **27**:91-98.

Soppa GK, Smolenski RT, Latif N, Yuen AH, Malik A, Karbowska J, Kochan Z, Terracciano CM, Yacoub MH: **Effects of chronic administration of clenbuterol on function and metabolism of adult rat cardiac muscle.** *Am J Physiol Heart Circ Physiol* 2005, **288**:H1468-1476.

Spurlock DM, McDaneld TG, McIntyre LM: **Changes in skeletal muscle gene expression following clenbuterol administration.** *BMC Genomics* 2006, **7**:320.

Stella R, Biancotto G, Krogh M, Angeletti R, Pozza G, Sorgato MC, James P, Andrighetto I: **Protein expression changes in skeletal muscle in response to growth promoter abuse in beef cattle.** *J Proteome Res* 2011, **10**:2744-2757.

Stewart AG, Tomlinson PR, Wilson JW: **Beta 2-adrenoceptor agonist-mediated inhibition of human airway smooth muscle cell proliferation: importance of the duration of beta 2-adrenoceptor stimulation.** *Br J Pharmacol* 1997, **121**:361-368.

Strader CD, Candelore MR, Hill WS, Sigal IS, Dixon RA: **Identification of two serine residues involved in agonist activation of the beta-adrenergic receptor.** *J Biol Chem* 1989, **264**:13572-13578.

Stuck BJ, Lenski M, Böhm M, Laufs U: **Metabolic switch and hypertrophy of cardiomyocytes following treatment with angiotensin II are prevented by AMP-activated protein kinase.** *J Biol Chem* 2008, **283**:32562-32569.

Sui J, Zhang J, Ching CB, Chen WN: **Comparative proteomic analysis of extracellular proteins reveals secretion of T-kininogen from vascular smooth muscle cells in response to incubation with s-enantiomer of propranolol.** *Mol Pharm* 2008, **5**:885-890.

Suzuki J, Ohno I, Nawata J, Miura S, Ikeda J, Shirato K: **Overexpression of insulin-like growth factor-I in hearts of rats with isoproterenol-induced cardiac hypertrophy.** *J Cardiovasc Pharmacol* 1999, **34**:635-644.

Swa HL, Blackstock WP, Lim LH, Gunaratne J: **Quantitative proteomics profiling of murine mammary gland cells unravels impact of annexin-1 on DNA-damage response, cell adhesion and migration.** *Mol Cell Proteomics* 2012.

Tafelmeyer P, Laurent C, Lenormand P, Rousselle JC, Marsollier L, Reysset G, Zhang R, Sickmann A, Stinear TP, Namane A, Cole ST: **Comprehensive proteome analysis of Mycobacterium ulcerans and quantitative comparison of mycolactone biosynthesis.** *Proteomics* 2008, **8**:3124-3138.

Takahashi Y, Karbowski M, Yamaguchi H, Kazi A, Wu J, Sebti SM, Youle RJ, Wang HG: **Loss of Bif-1 suppresses Bax/Bak conformational change and mitochondrial apoptosis.** *Mol Cell Biol* 2005, **25**:9369-9382.

Tannu NS, Rao VK, Chaudhary RM, Giorgianni F, Saeed AE, Gao Y, Raghow R: **Comparative proteomes of the proliferating C(2)C(12) myoblasts and fully differentiated myotubes reveal the complexity of the skeletal muscle differentiation program.** *Mol Cell Proteomics* 2004, **3**:1065-1082.

Tao WA, Aebersold R: **Advances in quantitative proteomics via stable isotope tagging and mass spectrometry.** *Curr Opin Biotechnol* 2003, **14**:110-118.

Tarazona S, García-Alcalde F, Dopazo J, Ferrer A, Conesa A: **Differential expression in RNA-seq: A matter of depth.** *Genome Res* 2011, **21**:2213-2223.

Tashiro N: **Effects of isoprenaline on contractions of directly stimulated fast and slow skeletal muscles of the guinea-pig.** *Br J Pharmacol* 1973, **48**:121-131.

Taulet N, Comunale F, Favard C, Charrasse S, Bodin S, Gauthier-Rouvière C: **N-cadherin/p120 catenin association at cell-cell contacts occurs in cholesterol-rich membrane domains and is required for RhoA activation and myogenesis.** *J Biol Chem* 2009, **284**:23137-23145.

Teng X, Xiao H: **Perspectives of DNA microarray and next-generation DNA sequencing technologies.** *Sci China C Life Sci* 2009, **52**:7-16.

Tewari AK, Popova-Butler A, El-Mahdy MA, Zweier JL: **Identification of differentially expressed proteins in blood plasma of control and cigarette smoke-exposed mice by 2-D DIGE/MS.** *Proteomics* 2011, **11**:2051-2062.

Thaler R, Rumpler M, Spitzer S, Klaushofer K, Varga F: **Mospd1, a new player in mesenchymal versus epidermal cell differentiation.** *J Cell Physiol* 2011, **226**:2505-2515.

Thomas L, Leung S, Cumming M, Shaw M, Albert N, McCallum J, McManus MT: **Genotypic variation in sulphur assimilation and metabolism of onion (Allium cepa L.). II: Characterisation of ATP sulphurylase activity.** *Phytochemistry* 2011.

Toigo M, Donohoe S, Sperrazzo G, Jarrold B, Wang F, Hinkle R, Dolan E, Isfort RJ, Aebersold R: **ICAT-MS-MS time course analysis of atrophying mouse skeletal muscle cytosolic subproteome.** *Mol Biosyst* 2005, **1**:229-241.

Turner JD, Rotwein P, Novakofski J, Bechtel PJ: **Induction of mRNA for IGF-I and -II during growth hormone-stimulated muscle hypertrophy.** *Am J Physiol* 1988, **255**:E513-517.

Unlu M, Morgan ME, Minden JS: **Difference gel electrophoresis: a single gel method for detecting changes in protein extracts.** *Electrophoresis* 1997, **18**:2071-2077.

van Baak MA, Mayer LH, Kempinski RE, Hartgens F: **Effect of salbutamol on muscle strength and endurance performance in nonasthmatic men.** *Med Sci Sports Exerc* 2000, **32**:1300-1306.

Van Hoof D, Pinkse MW, Oostwaard DW, Mummery CL, Heck AJ, Krijgsveld J: **An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics.** *Nat Methods* 2007, **4**:677-678.

van Koppen CJ, Jakobs KH: **Arrestin-independent internalization of G protein-coupled receptors.** *Mol Pharmacol* 2004, **66**:365-367.

Vaniotis G, Del Duca D, Trieu P, Rohlicek CV, Hébert TE, Allen BG: **Nuclear β-adrenergic receptors modulate gene expression in adult rat heart.** *Cell Signal* 2011, **23**:89-98.

Varshavsky A: **The N-end rule: functions, mysteries, uses.** *Proc Natl Acad Sci U S A* 1996, **93**:12142-12149.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270**:484-487.

Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE, Hieter P, Vogelstein B, Kinzler KW: **Characterization of the yeast transcriptome.** *Cell* 1997, **88**:243-251.

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH: **Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing.** *Mol Ecol* 2008, **17**:1636-1647.

von Maltzahn J, Bentzinger CF, Rudnicki MA: **Wnt7a-Fzd7 signalling directly activates the Akt/mTOR anabolic growth pathway in skeletal muscle.** *Nat Cell Biol* 2011.

von Zastrow M, Kobilka BK: **Ligand-regulated internalization and recycling of human beta 2-adrenergic receptors between the plasma membrane and endosomes containing transferrin receptors.** *J Biol Chem* 1992, **267**:3530-3538.

Wang M, You J, Bemis KG, Tegeler TJ, Brown DP: **Label-free mass spectrometry-based protein quantification technologies in proteomic analysis.** *Brief Funct Genomic Proteomic* 2008, **7**:329-339.

Wang P, Clausen T: **Treatment of attacks in hyperkalaemic familial periodic paralysis by inhalation of salbutamol.** *Lancet* 1976, **1**:221-223.

Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10**:57-63.

Warren CM, Geenen DL, Helseth DL, Xu H, Solaro RJ: **Sub-proteomic fractionation, iTRAQ, and OFFGEL-LC-MS/MS approaches to cardiac proteomics.** *J Proteomics* 2010, **73**:1551-1561.

Warriss PD, Brown SN, Rolph TP, Kestin SC: **Interactions between the beta-adrenergic agonist salbutamol and genotype on meat quality in pigs.** *J Anim Sci* 1990, **68**:3669-3676.

Washburn MP, Ulaszek RR, Yates JR, 3rd: **Reproducibility of quantitative proteomic analyses of complex biological mixtures by multidimensional protein identification technology.** *Anal Chem* 2003, **75**:5054-5061.

Weintraub H, Davis R, Tapscott S, Thayer M, Krause M, Benezra R, Blackwell TK, Turner D, Rupp R, Hollenberg S: **The myoD gene family: nodal point during specification of the muscle cell lineage.** *Science* 1991, **251**:761-766.

Wiener MC, Sachs JR, Deyanova EG, Yates NA: **Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures.** *Anal Chem* 2004, **76**:6085-6096.

Wilcox A, Katsanakis KD, Bheda F, Pillay TS: **Asb6, an adipocyte-specific ankyrin and SOCS box protein, interacts with APS to enable recruitment of elongins B and C to the insulin receptor signaling complex.** *J Biol Chem* 2004, **279**:38881-38888.

Wilhelm BT, Landry JR: **RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing.** *Methods* 2009, **48**:249-257.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J: **Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution.** *Nature* 2008, **453**:1239-1243.

Wilkins MR, Williams KL: **The extracellular matrix of the Dictyostelium discoideum slug.** *Experientia* 1995, **51**:1189-1196.

Wilson C, Wilson S, Piercy V, Sennitt MV, Arch JR: **The rat lipolytic beta-adrenoceptor: studies using novel beta-adrenoceptor agonists.** *Eur J Pharmacol* 1984, **100**:309-319.

Wu Y, Bauman WA, Blitzer RD, Cardozo C: **Testosterone-induced hypertrophy of L6 myoblasts is dependent upon Erk and mTOR.** *Biochem Biophys Res Commun* 2010, **400**:679-683.

Wu Z, Woodring PJ, Bhakta KS, Tamura K, Wen F, Feramisco JR, Karin M, Wang JY, Puri PL: **p38 and extracellular signal-regulated kinases regulate the myogenic program at multiple steps.** *Mol Cell Biol* 2000, **20**:3951-3964.

Xiao K, Allison DF, Buckley KM, Kottke MD, Vincent PA, Faundez V, Kowalczyk AP: **Cellular levels of p120 catenin function as a set point for cadherin expression levels in microvascular endothelial cells.** *J Cell Biol* 2003, **163**:535-545.

Xiao RP, Cheng H, Zhou YY, Kuschel M, Lakatta EG: **Recent advances in cardiac beta(2)-adrenergic signal transduction.** *Circ Res* 1999, **85**:1092-1100.

Yaffe D, Saxel O: **Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle.** *Nature* 1977, **270**:725-727.

Yamada S, Pokutta S, Drees F, Weis WI, Nelson WJ: **Deconstructing the cadherin-catenin-actin complex.** *Cell* 2005, **123**:889-901.

Yamaguchi W, Fujimoto E, Higuchi M, Tabata I: **A DIGE proteomic analysis for high-intensity exercise-trained rat skeletal muscle.** *J Biochem* 2010, **148**:327-333.

Yamamoto DL, Hutchinson DS, Bengtsson T: **Beta(2)-Adrenergic activation increases glycogen synthesis in L6 skeletal muscle cells through a signalling pathway independent of cyclic AMP.** *Diabetologia* 2007, **50**:158-167.

Yan JX, Devenish AT, Wait R, Stone T, Lewis S, Fowler S: **Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli.** *Proteomics* 2002, **2**:1682-1698.

Yang J, Takahashi Y, Cheng E, Liu J, Terranova PF, Zhao B, Thrasher JB, Wang HG, Li B: **GSK-3beta promotes cell survival by modulating Bif-1-dependent autophagy and cell death.** *J Cell Sci* 2010, **123**:861-870.

Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucleic Acids Res* 2002, **30**:e15.

Yang YT, McElligott MA: **Multiple actions of beta-adrenergic agonists on skeletal muscle and adipose tissue.** *Biochem J* 1989, **261**:1-10.

Yazawa H, Murakami T, Li HM, Back T, Kurosaka K, Suzuki Y, Shorts L, Akiyama Y, Maruyama K, Parsoneault E, et al: **Hydrodynamics-based gene delivery of naked DNA encoding fetal liver kinase-1 gene effectively suppresses the growth of pre-existing tumors.** *Cancer Gene Ther* 2006, **13**:993-1001.

Yimlamai T, Dodd SL, Borst SE, Park S: **Clenbuterol induces muscle-specific attenuation of atrophy through effects on the ubiquitin-proteasome pathway.** *J Appl Physiol* 2005, **99**:71-80.

Yogosawa S, Hatakeyama S, Nakayama KI, Miyoshi H, Kohsaka S, Akazawa C: **Ubiquitylation and degradation of serum-inducible kinase by hVPS18, a RING-H2 type ubiquitin ligase.** *J Biol Chem* 2005, **280**:41619-41627.

Young JF, Larsen LB, Malmendal A, Nielsen NC, Straadt IK, Oksbjerg N, Bertram HC: **Creatine-induced activation of antioxidative defence in myotube cultures revealed by explorative NMR-based metabonomics and proteomics.** *J Int Soc Sports Nutr* 2010, **7**:9.

Young KG, Thurston SF, Copeland S, Smallwood C, Copeland JW: **INF1 is a novel microtubule-associated formin.** *Mol Biol Cell* 2008, **19**:5168-5180.

Zaugg M, Xu W, Lucchinetti E, Shafiq SA, Jamali NZ, Siddiqui MA: **Beta-adrenergic receptor subtypes differentially affect apoptosis in adult rat ventricular myocytes.** *Circulation* 2000, **102**:344-350.

Zeman RJ, Ludemann R, Easton TG, Etlinger JD: **Slow to fast alterations in skeletal muscle fibers caused by clenbuterol, a beta 2-receptor agonist.** *Am J Physiol* 1988, **254**:E726-732.

Zeman RJ, Peng H, Etlinger JD: **Clenbuterol retards loss of motor function in motor neuron degeneration mice.** *Exp Neurol* 2004, **187**:460-467.

Zetterlund A, Linden M, Larsson K: **Effects of beta2-agonists and budesonide on interleukin-1beta and leukotriene B4 secretion: studies of human monocytes and alveolar macrophages.** *J Asthma* 1998, **35**:565-573.

Zhang KM, Hu P, Wang SW, Feher JJ, Wright LD, Wechsler AS, Spratt JA, Briggs FN: **Salbutamol changes the molecular and mechanical properties of canine skeletal muscle.** *J Physiol* 1996, **496 ( Pt 1)**:211-220.

Zhao J, Grant SF: **Advances in whole genome sequencing technology.** *Curr Pharm Biotechnol* 2011, **12**:293-305.

Zhu W, Smith JW, Huang CM: **Mass spectrometry-based label-free quantitative proteomics.** *J Biomed Biotechnol* 2010, **2010**:840518.

Zuk D, Jacobson A: **A single amino acid substitution in yeast eIF-5A results in mRNA stabilization.** *EMBO J* 1998, **17**:2914-2925.

Zybailov B, Coleman MK, Florens L, Washburn MP: **Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling.** *Anal Chem* 2005, **77**:6218-6224.

## Appendices

**Appendix I** Table 1. List of MASCOT identifications from the DIGE analysis of mouse muscle cells after 96 hr of salbutamol treatment.

Table 2. List of MASCOT identifications from the DIGE analysis of mouse muscle cells after 24 hr of salbutamol treatment.

Table 3. List of MASCOT identifications from the statistical analysis of DIGE.

Table 4. List of DAVID clusters and processes for the proteins identified as modulated at 24 hr.

Table 5. List of DAVID clusters and processes for the proteins identified as modulated at 96 hr.

Table 6. List of DAVID clusters and processes for the proteins identified as modulated run against the proteins identified in the statistical test.

**Appendix II** Table 1. List of DAVID process identified for the genes modulated at 2 hr.

Table 2. List of DAVID process identified for the genes modulated at 6 hr.

Table 3. List of DAVID process identified for the genes modulated at 24 hr.

Table 4. List of DAVID process identified for the genes modulated at 96hr.

Table 5. Visualisation of how the 785 genes are differently

**Appendix III** Table 1. Log ratios of all genes and proteins identified from RNA-Seq, Label free and SILAC experiments.

Table 2. Log rations of only genes present in both the RNASeq data and SILAC data at 24 hr.

Table 3. Log rations of only genes present in both the RNASeq data and SILAC data at 96 hr.

Table 4. Log rations of only genes present in both the RNASeq data and label free data at 24 hr.

Table 5. Log rations of only genes present in both the RNASeq data and label free data at 96 hr.

Figure 1. Pearson correlation of the data from the label free experiment analysed through either MaxQuant or Progenesis at 24 hr.

Figure 2. Pearson correlation of the data from the label free experiment analysed through either MaxQuant or Progenesis at 96 hr.

Appendices I-III can be found on the enclosed CD

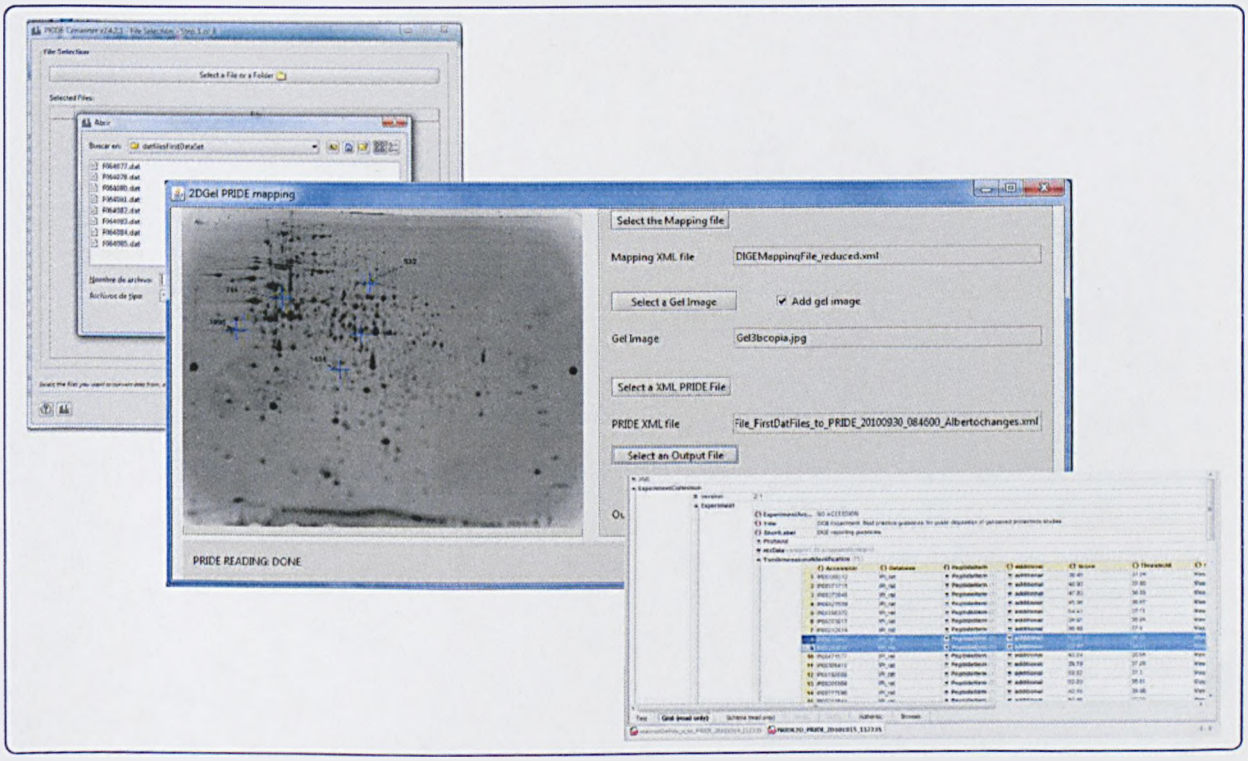
**Papers in support of this thesis**

Kenyani J, Medina-Aunon JA, Martinez-Bartolomé S, Albar JP, Wastling JM, Jones AR. **A DIGE study on the effects of salbutamol on the rat muscle proteome - an exemplar of best practice for data sharing in proteomics.** *BMC Res Notes.* 2011 Mar 28(4):86.

Burniston JG, Kenyani J, Wastling JM, Burant CF, Qi NR, Koch LG, Britton SL. **Proteomic analysis reveals perturbed energy metabolism and elevated oxidative stress in hearts of rats with inborn low aerobic capacity.** *Proteomics.* 2011 Aug;11(16):3369-79.

Qi D, Brownridge P, Xia D, Mackay K, Gonzalez-Galarza FF, Kenyani J, Harman V, Beynon RJ and Jones AR. **A software toolkit and interface for performing stable isotope labelling and top3 quantification using Progenesis LC-MS.** *Proteomics* 2011, 11(33):3369-3379.

BMC
Research Notes




# A DIGE study on the effects of salbutamol on the rat muscle proteome - an exemplar of best practice for data sharing in proteomics

Kenyani *et al.*

BioMed Central

Kenyani *et al. BMC Research Notes* 2011, 4:86
http://www.biomedcentral.com/1756-0500/4/86 (28 March 2011)

Kenyani et al. BMC Research Notes 2011, 4:86
http://www.biomedcentral.com/1756-0500/4/86

BMC Research Notes

DATA NOTE Open Access

# A DIGE study on the effects of salbutamol on the rat muscle proteome - an exemplar of best practice for data sharing in proteomics

Jenna Kenyani[1†], J Alberto Medina-Aunon[2†], Salvador Martinez-Bartolomé[2], Juan-Pablo Albar[2], Jonathan M Wastling[1] and Andrew R Jones[3*]

**Abstract**

**Background:** Proteomic techniques allow researchers to perform detailed analyses of cellular states and many studies are published each year, which highlight large numbers of proteins quantified in different samples. However, currently few data sets make it into public databases with sufficient metadata to allow other groups to verify findings, perform data mining or integrate different data sets. The Proteomics Standards Initiative has released a series of "Minimum Information About a Proteomics Experiment" guideline documents (MIAPE modules) and accompanying data exchange formats. This article focuses on proteomic studies based on gel electrophoresis and demonstrates how the corresponding MIAPE modules can be fulfilled and data deposited in public databases, using a new experimental data set as an example.

**Findings:** We have performed a study of the effects of an anabolic agent (salbutamol) at two different time points on the protein complement of rat skeletal muscle cells, quantified by difference gel electrophoresis. In the DIGE study, a total of 31 non-redundant proteins were identified as being potentially modulated at 24 h post treatment and 110 non redundant proteins at 96 h post-treatment. Several categories of function have been highlighted as strongly enriched, providing candidate proteins for further study. We also use the study as an example of best practice for data deposition.

**Conclusions:** We have deposited all data sets from this study in public databases for further analysis by the community. We also describe more generally how gel-based protein identification data sets can now be deposited in the PRoteomics IDEntifications database (PRIDE), using a new software tool, the PRIDESpotMapper, which we developed to work in conjunction with the PRIDE Converter application. We also demonstrate how the ProteoRed MIAPE generator tool can be used to create and share a complete and compliant set of MIAPE reports for this experiment and others.

## Introduction

A variety of high-throughput experimental techniques are available for studying how the protein complement of a sample (the proteome) changes under different cellular conditions, such as during disease processes. The changes observed in individual proteins, or groups of proteins, as experimental conditions vary allow researchers to begin understanding the underlying molecular mechanisms in the cell. Gel electrophoresis (GE) has been employed to study proteins for over four decades [1]. GE is frequently applied in two dimensions, whereby proteins are separated by charge followed by molecular weight [2]. More recently, the difference in-gel electrophoresis technique (DIGE) [3] has improved the relative quantification of proteins on 2-D gels. In DIGE, the whole proteomes of different samples are labelled with different fluorescent dyes, mixed and applied to a single gel, thus reducing gel to gel variability in protein migration. Despite the relative age of gel-based proteomic techniques, and recent advances in liquid chromatography-mass spectrometry (LC-MS) for protein quantification, gel-based techniques are still commonly

* Correspondence: andrew.jones@liv.ac.uk
† Contributed equally
[3]Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool, UK
Full list of author information is available at the end of the article

used. For all proteomic techniques, it has been widely documented that the protocols employed can influence the results, for example introducing variability in the set of proteins detected or the estimation of their individual abundances. It is thus important to capture and report a detailed set of information (termed metadata) about how experiments were performed and analysed to allow groups to verify findings, employ similar protocols in their own labs or compare data sets generated in different experiments.

The Human Proteome Organisation - Proteomics Standards Initiative (HUPO-PSI, [4]) was created to help scientists share their data, deposit data sets in public databases and provide tools to assist other groups in performing large scale analysis of public proteomic data sets. In 2007, the PSI published the Minimum Information About a Proteomics Experiment (MIAPE) specification [5]. From this root document, a set of MIAPE modules for proteomics techniques were delivered: gel electrophoresis [6], gel image informatics [7], mass spectrometry [8], mass spectrometry informatics [9], column chromatography [10], capillary electrophoresis [11] and protein-protein or molecular interactions [12]. Each MIAPE module contains a minimal checklist of items that should be reported for the given technique. The items can be reported using plain language, for example describing specific points within the experimental protocols or the data analysis that has been performed, to allow other groups to interpret the published results without ambiguity as to how they were generated. The PSI has also developed data exchange formats, typically represented in Extensible Markup Language (XML). One of these, GelML [13], captures the data related to gel electrophoresis experiments. There are a number of public databases storing protein identification data from proteomics, including PRIDE [14], PeptideAtlas [15], Peptidome [16], the GPMDB [17] and the Swiss2D-PAGE database storing GE experiments [18]. However the widely used protein identification repositories (PRIDE, PeptideAtlas etc) are primarily focussed on LC-MS studies and historically have either no GE data sets or no simple mechanism for deposition of data derived from gel-based experiments.

In this article we demonstrate how MIAPE GE (gel electrophoresis) and GI (gel informatics) compliant reports can be created easily in practice, through the MIAPE Generator tool [19], developed by ProteoRed - the Spanish network for proteomics. We have also developed a new tool, the PRIDESpotMapper, to work alongside the PRIDE Converter software [20] to enable GE studies to be captured in the PRIDE XML format and be submitted to the public PRIDE repository. The provision of both the MIAPE report and the public PRIDE record, enables other groups to download the complete data sets, including raw gel images, mass spectra and protein identifications, along with complete descriptions of the experimental protocols.

We have performed a study on the effects of salbutamol (an anabolic agent) on the proteome of rat muscle cells. Salbutamol is a type of $beta_2$ adrenergic agonist, which is known to cause hypertrophy in muscle but the underlying molecular mechanisms are not well understood. The aims of the study are to use proteomic technologies to model changes in the development of skeletal muscle cells *in vitro* in the presence of salbutamol and to identify novel proteins and pathways within these cells that interact with these agents, and therefore could be potential targets for their action. DIGE was used to compare control and treated samples at 24 h and 96 h after addition of salbutamol. Gel spots with changed abundance were subjected to tandem mass spectrometry for protein identification. Bioinformatics analysis was performed using the Gene Ontology (GO) [21] and the DAVID tool [22] for determining categories of functions that appear to be enriched at the different time points.

In the supplementary material [Additional file 1], we include the protocols employed in the DIGE study, as they would be reported in a standard journal article. We have also used the ProteoRed MIAPE Generator to create MIAPE GE and GI compliant reports (described in [19]) and we use these examples to demonstrate how a standard set of materials and methods map into the MIAPE reports generated, to act as a practical guide to MIAPE for proteome scientists. We have also deposited the MS data sets and identifications in PRIDE, using the PRIDESpotMapper and PRIDE Converter, for public access and review.

### Software development

The PRIDE Converter software [20] enables conversion from a variety of mass spectra and search engine file formats into the PRIDE XML format that can subsequently be used for uploading spectra and peptide/protein identifications into the PRIDE database. However, the PRIDE Converter has been designed primarily for "shotgun proteomics" experimental designs, where peptide to protein inference is performed across all input spectra, which is not well suited to gel-based studies. The software is capable of loading multiple identification files (e.g. Mascot dat files or Sequest .out files), but in its internal processing, the resulting proteins are inferred from a combined list containing all the identified peptides. For gel-based studies, typically each identification file (say one Mascot dat file) comes from a single gel spot and its identified peptides should not be combined with those from other spots. The PRIDE Converter also has no mechanism for uploading gel image

coordinates, or additional information regarding protein quantification. To overcome these limitations, a custom version of the PRIDE Converter was developed by the PRIDE team, where every identified peptide was annotated with the name of the source gel spot. Simultaneously, we developed a new application called "the PRIDESpotMapper" as a complement to the PRIDE Converter for gel-based experiments. This was implemented in Java and modifies the PRIDE XML file generated using the custom PRIDE Converter, dividing the identified proteins according to the source identification file for each gel spot. Starting from a PRIDE XML file and either an XML or Excel spot map (see [Additional File 2] for the format specifications) the application ensures that records are created for each identified protein, derived from peptide identifications from each input file independently.

Once all the resulting files coming from the search engine (Mascot for this version) are joined in a single PRIDE XML using the PRIDE Converter, the execution of the PRIDESpotMapper is straightforward (Figure 1). First, either the XML or Excel spot map file should be entered. Second, the gel image can be loaded from a local file or from a URI, for example if gel images have been loaded into the ProteoRed MIAPE Generator database [19]. Third, the previously created PRIDE XML file is required. The application merges the two data files (Spot map file and PRIDE XML file) to create a new PRIDE XML file (internally called 2D PRIDE XML file),

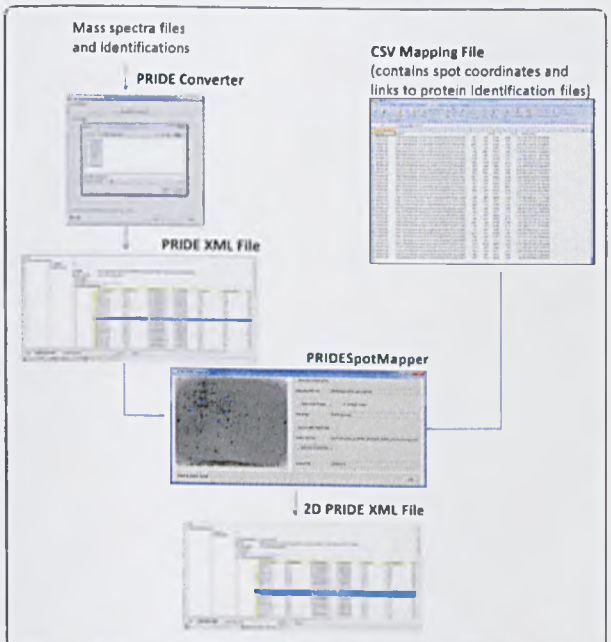



**Figure 1 The workflow used to construct a PRIDE XML file containing gel spot data, via the modified PRIDE Converter and the PRIDESpotMapper.**

in which each spot is linked to one protein only with its corresponding peptides, alongside gel spot coordinates and relative quantification data. The file is then saved on the local drive, ready for upload to the PRIDE database.

## Results

The DIGE gels were analyzed as described in the Supplementary Methods [Additional file 1] and sets of spots were identified as being differentially expressed at the 24 h time point (versus the untreated control) and 96 h time point (versus control). The protein(s) contained within those spots were then identified by tandem mass spectrometry (MS/MS). In most spots, more than one protein was identified, indicating that some co-migration of proteins occurred (and the high sensitivity of MS/MS). As such it is not always possible to link exact quantitative differences between conditions to specific proteins, although general conclusions can be made about the groups of proteins that have changed between conditions. The proteins that were identified with high confidence were further analyzed using the DAVID tool [22], which highlighted several functions that were strongly enriched (discussed below).

At the 24 h time point 17 spots of interest were identified: 4 spots were down-regulated, 13 spots up-regulated from which 31 non-redundant proteins were identified [Additional file 3]. 23% of the proteins identified are cytoskeletal and also are mapped in pathways involved with skeletal development (based on Gene Ontology terms). At the 96 h time point 35 spots of interest were identified - 11 spots were down-regulated, 24 spots up-regulated [Additional file 4]. From these spots 110 non-redundant proteins were identified. 25% of these proteins are cytoskeletal proteins. Several of these proteins, e.g. vimentin and desmin, are known to be involved in the skeletal development pathway. There are several proteins modulated in both sample sets, such as beta-enolase that is involved in glycolysis. Some proteins were found in more than one spot, which could suggest the presence of post-translational modifications, such as phosphorylation.

We have performed gene ontology enrichment analysis on the two data sets using DAVID (24 h [Additional file 5] and 96 h [Additional file 6]). At 24 h the main functional clusters enriched were: "contractile fiber", "cytoskeleton", "calcium ion binding" and "collagen biosynthetic process". At 96 h post treatment, the main enriched functional categories were "cytoskeleton", "tubulin", "microtubule-based movement", "GTPase activity" "cellular protein complex assembly", "contractile fiber" and "regulation of ATPase activity" amongst several others.

In summary, proteins involved in ion binding and transport, nucleosome assembly, cell interactions,

protein binding and structural proteins seem to be modulated at 96 h, whereas only structural and energy production proteins are affected after 24 h. An immediate effect of the anabolic agent is to produce a structural effect that needs a great deal of energy. It appears a more complex later effect is observed, involving a number of cellular pathways.

Salbutamol is shown to cause muscle hypertrophy, which suggests it may have a similar mode of action as other $beta_2$-adrenergic agonists. This effect on the muscle cell is rapid, and is clearly seen using high magnification microscopy. As anticipated, many of the differentially expressed proteins identified are cytoskeletal. A significant number are also involved in transcription or translation. Skeletal development pathways are activated both at the early and later time points. Up-regulation of ATP synthesis, glycolysis and phosphorylation also seems to be occurring.

### Data deposition and MIAPE report generation

MIAPE guideline documents describe the metadata that should be captured about a given proteomic technique, for instance detailing the minimum information that should be reported about the experimental protocols. The MIAPE Generator tool has been developed to assist in the generation of MIAPE-compliant reports, and is freely accessible from http://www.proteored.org/. The tool guides users through each stage of the report creation process, capturing all details required by the underlying MIAPE module. The tool's user interface is based on a series of web forms for data entry, built on top of a relational database. These forms follow a hierarchical structure according to the original sections of each MIAPE module. Each document is always linked to a project, which can be viewed and accessed only by the project owner, until the document is ready for public access. The tool has a template system such that protocols can be re-used in different reports to avoid repetition in data entry, and drop-down boxes are provided as applicable, containing controlled vocabulary or ontology terms to capture standard terminology for techniques, units and so on, to allow reports to be compared automatically. The generated reports are stored in the database and can be exported in a variety of formats. In this instance, the MIAPE Generator tool was used to create two reports for each time point, capturing the methods detailed above concerning gel electrophoresis sections (in the MIAPE GE report) and concerning the gel image informatics sections (in the MIAPE GI report) - see "Availability and requirements".

The benefits of producing the MIAPE reports, in addition or instead of a traditional Materials and Methods, are as follows. The report has a standard structure, requesting key details for each stage of the process, ensuring that the experimenter does not fail to report any information that may be important for reproducing the protocols in another lab. As one example, the MIAPE GE document requests that the gel recipe are provided (section 3.2.2) if the gel was not purchased pre-cast. Similarly, the MIAPE GI document requests details should be provided on software parameters and algorithms used with different software packages, which can affect the results obtained. The MIAPE GE/GI specifications also request that raw data should be provided and linked to the report, in this case, the original gel images. This could potentially be hugely valuable if researchers are interested in performing more detailed examination of results, for example to test if a specific protein is differentially regulated, using different statistical assumptions than the researchers who generated the data.

### Deposition of protein identification data in PRIDE

The PRIDE database has become one of the leading public repositories for proteomics results. However, to date, few gel studies have been deposited in PRIDE due to the lack of appropriate tools. We have deposited two files - one for the 24 h and one for 96 h time point, each containing the protein identifications for each gel spot. Each protein identification has a link to the gel image in the MIAPE database, along with X/Y coordinates and quantification information, in terms of the ratio detected by DIGE across treated versus control samples. As far as we are aware, this is the first deposition of a complete quantitative DIGE data set in PRIDE. The PRIDE records can be accessed at http://www.ebi.ac.uk/pride/ under accessions 16472 and 16473. Data files download from PRIDE can be visualised using the PRIDEViewer software [23].

## Discussion and Conclusions

Each year there are many hundreds of proteomics studies published in the literature, in which gel electrophoresis is used to separate, identify and perform relative quantification of the proteins present in complex samples. However, few of these data sets have ever made it into the public domain, beyond lists of protein spots provided in tables within articles or as supplementary material in spreadsheets. The Proteomics Standards Initiative has released several tools and guideline documents designed to improve the public accessibility of proteomics data, including minimum reporting guidelines (MIAPE documents) and XML formats. The EBI has also developed the PRIDE database to allow proteomics scientists to publish protein identification data sets to the wider community.

While it is possible to include protein identification data in the database behind the MIAPE generator tool,

this is not the standard public repository for this kind of data. Instead, proteomics scientists tend to search the PRIDE database (or GPM, PeptideAtlas, Peptidome, Tranche) for identification data. As such, it is important for gel-based proteomics studies to be deposited in one of these primary data repositories. To date, almost no data sets derived from gel-based experiments have been deposited in any of these databases. We have created the PRIDESpotMapper to work alongside the PRIDE Converter and thus, for the first time, provide a simple route for uploading valid PRIDE XML, containing gel spot information and quantitative values. The PRIDE developers may include gel support in future versions of the PRIDE Converter directly. We will work with the PRIDE team to incorporate the same mechanism presented here for representing gel spot data, to ensure that researchers wishing to share gel-based proteomic data can use the PRIDESpotMapper now, and migrate to a new version of the PRIDE Converter, as and when appropriate.

This article should serve as an exemplar for how researchers can upload gel-based data to PRIDE and use the MIAPE Generator tool for creating MIAPE compliant reports. There is on-going discussion with journal editors as to the requirement for proteomics articles to be MIAPE compliant - and these reports may in due course supplement, or in some cases replace, traditional materials and methods sections of proteomics articles. We encourage further discussion of these issues, for example through the PSI's open mailing lists or attendance at the annual PSI meeting.

For the study described, we have created MIAPE reports describing the gel electrophoresis and the gel image informatics performed, and these have been deposited in the associated database. The database also contains the source gel images, allowing other groups to re-analyse these data, using the same or different software pipelines.

The initial results of the study show that several key pathways are modulated by treatment with salbutamol, with significantly more changes occurring at 96 h post-treatment. This indicates that there is a lag between the treatment and the downstream activation of cellular pathways. There are some limitations of the DIGE results, not least that true quantitative ratios cannot be linked to individual protein identities, since the sensitivity of tandem MS revealed that many spots on the gels contained more than one protein. However, the ontology enrichment analysis shows that many of the protein groups highlighted are likely to be direct or indirect targets for salbutamol, since the enriched functional categories fit our expected hypotheses of the effects of an adrenergic agonist. We are making these data sets freely available as we anticipate that they will be useful to other researchers working in this area for building hypotheses about the mechanism of action of salbutamol on the proteome of muscle tissue.

## Availability and requirements

The software described in this article is accessible from http://proteo.cnb.csic.es/pridespotmapper.

The adapted PRIDE Converter described here is released as a snapshot build (2.4.2), which will be updated periodically when there are major new releases of the main PRIDE Converter software but not for minor updates. The PRIDE Converter is freely available and open-source, released under Apache License 2.0. PRIDE Converter requires Java 1.5 (or above). The current version has been tested on Windows XP, Windows Vista, Linux and Mac OS X.

The PRIDESpotMapper is freely available as a Java jar file for local install or can be run using Java web start. The application has tested using Java Runtime Environment (JRE) 1.6 with the following operating systems: Windows 7, Windows XP, Windows Vista, Linux Red Hat, Linux Ubuntu.

URLs to link direct to these records in the MIAPE Generator database.

24 h timepoint:

http://estrellapolar.cnb.csic.es/proteored/MIAPE/MIAPE_GE.asp?pmCodigoAcceso=415db6c1&pmIDUsuario=2378&pmId=1082

http://estrellapolar.cnb.csic.es/proteored/MIAPE/MIAPE_GI.asp?pmCodigoAcceso=1d9f04d3&pmIDUsuario=2378&pmId=768

96 h timepoint:

http://estrellapolar.cnb.csic.es/proteored/MIAPE/MIAPE_GE.asp?pmCodigoAcceso=ae31268d&pmIDUsuario=2378&pmId=663

http://estrellapolar.cnb.csic.es/proteored/MIAPE/MIAPE_GI.asp?pmCodigoAcceso=a43637ec&pmIDUsuario=2378&pmId=397

**PRIDE access for data sets**

Accessions 16472 and 16473

**Mascot dat files located on Tranche:**

24 h time point:

https://proteomecommons.org/dataset.jsp?id=ziJZS3iGNcd5eMDW3vPpCb5VXJ4oorFWe1xwdIaE97hUxDNcXwtXaf6twotWtsTds4RVu84Obfgw2oLp3k7tRjWXWx8AAAAAAAAAChw%3D%3D

96 h time point:

https://proteomecommons.org/dataset.jsp?id=bZv347BIF4uVOWlwKok4ASHz2OCgPSXwqxfNP4LB2Qqna6tEnYVQNilPsrlQMgIAZBUAxyJCBpCK2kRqq%2BPCoQIv6oAAAAAAAAACjg%3D%3D

Passphrase: ratproteome

## Additional material

**Additional file 1: Supplementary Methods.** The materials and methods from this study

**Additional file 2: Example mapping file for the PRIDESpotMapper.** The csv file can be opened, for example, in spreadsheet software. The file contains information for mapping Mascot dat files to spot coordinates on gel images, and is used to create the final PRIDE XML file.

**Additional file 3: Proteins with changed abundance at 24 h post-treatment.** A summary file containing the set of proteins that had modulated abundance, as detected by DIGE, at 24 h post-treatment with salbutamol.

**Additional file 4: Proteins with changed abundance at 96 h post-treatment.** A summary file containing the set of proteins that had modulated abundance, as detected by DIGE, at 96 h post-treatment with salbutamol.

**Additional file 5: Clusters of functional annotations for proteins at 24 h post-treatment.** The output of clustering functional annotations produced by DAVID, on the protein set at 24 h post-treatment

**Additional file 6: Clusters of functional annotations for proteins at 96 h post-treatment.** The output of clustering functional annotations produced by DAVID, on the protein set at 96 h post-treatment

### List of abbreviations

DIGE: difference in-gel electrophoresis; GE: Gel electrophoresis; GI: gel informatics; GO: Gene ontology; HUPO: Human Proteome Organisation; LC-MS: Liquid Chromatography-Mass Spectrometry; MIAPE: Minimum Information About a Proteomics Experiment; PRIDE: PRoteomics IDEntifications database; PSI: Proteomics Standards Initiative; XML: Extensible Markup Language

### Acknowledgements and Funding

We would like to thank Adrian Thompson at Pfizer for providing rat cells. JK is funded on a BBSRC CASE studentship, in collaboration with Pfizer, whose support we gratefully acknowledge. Work on data standards in ARJ's group is funded by the BBSRC [BB/G010781/1].

### Author details

[1]Institute of Infection and Global Health, University of Liverpool, Crown Street, Liverpool, UK. [2]Spanish Institute for Proteomics (ProteoRed), Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Madrid, Spain. [3]Institute of Integrative Biology, University of Liverpool, Biosciences Building, Crown Street, Liverpool, UK.

### Authors' contributions

JK performed the DIGE study, supervised by JMW. JAMA led the software development work described, assisted by SMB and supervised by JPA. ARJ led the writing of the manuscript and contributed to software development and to analysis of the DIGE results. All authors have contributed text to the manuscript.

### Competing interests

The authors declare that they have no competing interests.

Received: 17 December 2010 Accepted: 28 March 2011
Published: 28 March 2011

### References

1. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 1970, 227:680-685.
2. O'Farrell PH: **High resolution two-dimensional electrophoresis of proteins.** *J Biol Chem* 1975, 250:4007-4021.
3. Unlu M, Morgan ME, Minden JS: **Difference gel electrophoresis: a single gel method for detecting changes in protein extracts.** *Electrophoresis* 1997, 18:2071-2077.
4. **The HUPO Proteomics Standards Initiative.** [http://www.psidev.info/].
5. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK, Jones AR, Zhu W, Apweiler R, Aebersold R, Deutsch EW, *et al*: **The minimum information about a proteomics experiment (MIAPE).** *Nat Biotech* 2007, 25:887-893.
6. Gibson F, Anderson L, Babnigg G, Baker M, Berth M, Binz PA, Borthwick A, Cash P, Day BW, Friedman DB, *et al*: **Guidelines for reporting the use of gel electrophoresis in proteomics.** *Nat Biotech* 2008, 26:863-864.
7. Hoogland C, O'Gorman M, Bogard P, Gibson F, Berth M, Cockell SJ, Ekefjärd A, Forsstrom-Olsson O, Kapferer A, Nilsson M, *et al*: **Guidelines for reporting the use of gel image informatics in proteomics.** *Nat Biotech* 2010.
8. Taylor CF, Binz PA, Aebersold R, Affolter M, Barkovich R, Deutsch EW, Horn DM, Huhmer A, Kussmann M, Lilley K, *et al*: **Guidelines for reporting the use of mass spectrometry in proteomics.** *Nat Biotech* 2008, 26:860-861.
9. Binz PA, Barkovich R, Beavis RC, Creasy D, Horn DM, Julian RK, Seymour SL, Taylor CF, Vandenbrouck Y: **Guidelines for reporting the use of mass spectrometry informatics in proteomics.** *Nat Biotech* 2008, 26:862-862.
10. Jones AR, Carroll K, Knight D, MacLellan K, Domann PJ, Legido-Quigley C, Huang L, Smallshaw L, Mirzaei H, Paton NW: **Guidelines for reporting the use of column chromatography in proteomics.** *Nat Biotech* 2010.
11. Domann PJ, Akashi S, Barbas C, Huang L, Lau W, Legido-Quigley C, McClean S, Neusüss C, Perrett D, Quaglia M, *et al*: **Guidelines for reporting the use of capillary electrophoresis in proteomics.** *Nature Biotechnology* 2010.
12. Orchard S, Salwinski L, Kerrien S, Montecchi-Palazzi L, Oesterheld M, Stumpflen V, Ceol A, Chatr-aryamontri A, Armstrong J, Woollard P, *et al*: **The minimum information required for reporting a molecular interaction experiment (MIMIx).** *Nat Biotech* 2007, 25:894-898.
13. Gibson F, Hoogland C, Martinez-Bartolomé S, Medina-Aunon JA, Albar JP, Babnigg G, Wipat A, Hermjakob H, Almeida JS, Stanislaus R, *et al*: **The Gel Electrophoresis Markup Language (GelML) from the Proteomics Standards Initiative.** *PROTEOMICS* 2010, 10:3073-3081.
14. Jones P, Cote RG, Martens L, Quinn AF, Taylor CF, Derache W, Hermjakob H, Apweiler R: **PRIDE: a public repository of protein and peptide identifications for the proteomics community.** *Nucl Acids Res* 2006, 34:D659-663.
15. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, Chen S, Eddes J, Loevenich SN, Aebersold R: **The PeptideAtlas project.** *Nucl Acids Res* 2006, 34:D655-658.
16. Slotta DJ, Barrett T, Edgar R: **NCBI Peptidome: a new public repository for mass spectrometry peptide identifications.** *Nat Biotech* 2009, 27:600-601.
17. Craig R, Cortens JP, Beavis RC: **Open Source System for Analyzing, Validating, and Storing Protein Identification Data.** *Journal of Proteome Research* 2004, 3:1234-1242.
18. Hoogland C, Mostaguir K, Sanchez JC, Hochstrasser DF, Appel RD: **SWISS-2DPAGE, ten years later.** *PROTEOMICS* 2004, 4:2352-2356.
19. Martínez-Bartolomé S, Medina-Aunon JA, Jones AR, Albar JP: **Semi-automatic tool to describe, store and compare proteomics experiments based on MIAPE compliant reports.** *PROTEOMICS* 2010, 10:1256-1260.
20. Barsnes H, Vizcaino JA, Eidhammer I, Martens L: **PRIDE Converter: making proteomics data-sharing easy.** *Nat Biotech* 2009, 27:598-599.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al*: **Gene Ontology: tool for the unification of biology.** *Nat Genet* 2000, 25:25-29.
22. Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protocols* 2008, 4:44-57.
23. Medina-Aunon JA, Carazo JM, Albar JP: **PRIDEViewer: A novel user-friendly interface to visualize PRIDE XML files.** *PROTEOMICS* 2010, 11:334-337.

doi:10.1186/1756-0500-4-86
**Cite this article as:** Kenyani *et al.*: **A DIGE study on the effects of salbutamol on the rat muscle proteome - an exemplar of best practice for data sharing in proteomics.** *BMC Research Notes* 2011 **4**:86.

Published in final edited form as:
*Proteomics*. 2011 August ; 11(16): 3369–3379. doi:10.1002/pmic.201000593.

# Proteomic analysis reveals perturbed energy metabolism and elevated oxidative stress in hearts of rats with inborn low aerobic capacity

**Jatin G. Burniston**[1,2,*], **Jenna Kenyani**[3], **Jonathan M. Wastling**[3], **Charles F. Burant**[4], **Nathan R. Qi**[4], **Lauren G. Koch**[4,5], and **Steven L. Britton**[4,5]

[1]Research Institute for Sport and Exercise Sciences, Liverpool John Moores University, Liverpool, L3 3AF, UK

[2]Institute for Health Research, Liverpool John Moores University, Liverpool, L3 3AF, UK

[3]Comparative Molecular Medicine, School of Veterinary Science, University of Liverpool, Liverpool, L69 7ZJ, UK

[4]Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109-2200, USA.

[5]Department of Anesthesiology, University of Michigan, Ann Arbor, MI 48109-2200, USA.

## Abstract

Selection on running capacity has created rat phenotypes of high capacity runners (HCR) that have enhanced cardiac function and low capacity runners (LCR) that exhibit risk factors of metabolic syndrome. We analysed hearts of HCR and LCR from generation 22 of selection using DIGE and identified proteins from MS database searches. The running capacity of HCR was 6-fold greater than LCR. DIGE resolved 957 spots and proteins were unambiguously identified in 369 spots. Protein expression profiling detected 67 statistically significant ($P<0.05$; false discovery rate <10 %, calculated using q-values) differences between HCR and LCR. Hearts of HCR rats exhibited robust increases in the abundance of each enzyme of the beta-oxidation pathway. In contrast, LCR hearts were characterised by the modulation of enzymes associated with ketone body or amino acid metabolism. LCR also exhibited enhanced expression of antioxidant enzymes such as catalase and greater phosphorylation of alpha B-crystallin at serine 59, which is a common point of convergence in cardiac stress signalling. Thus proteomic analysis revealed selection on low running capacity is associated with perturbations in cardiac energy metabolism and provided the first evidence that the LCR cardiac proteome is exposed to greater oxidative stress.

### Keywords

2D Gel Electrophoresis; Mass Spectrometry; Animal Selection Model

*Corresponding author: Dr Jatin G. Burniston. Muscle Physiology and Proteomics Laboratory, Research Institute for Sport and Exercise Sciences, Tom Reilly Building, Liverpool John Moores University, Byrom Street, LIVERPOOL, L3 3AF United Kingdom. Tel: 0151 904 6265. Fax: 0151 904 6284 j.burniston@ljmu.ac.uk.

Supporting information.

2D gel image and mass spectrometry protein identification table including x,y spot coordinates submitted to The World 2D-PAGE repository (http://world-2dpage.expasy.org/repository/), accession number 0022.

## 1. Introduction

Divergent selection on running capacity has produced two populations of rats, high capacity runners (HCR) and low capacity runners (LCR), which have the potential to illuminate mechanisms underlying diseases associated with low cardiorespiratory fitness. Genetic heterogeneity of HCR-LCR lines is maintained by a rotational mating paradigm that minimizes inbreeding [1], and concurrent breeding of LCR and HCR enables the lines to serve as reciprocal controls for unknown environmental changes. Compared to inbred strains, in which essentially all loci are fixed, these outbred selected lines maintain genetic complexity, allowing allelic variants to be enriched by selection pressure [2]. As a result, the HCR-LCR model is ideally suited to discovering the resultant effects of epistatic interactions, modifier genes or synergistic actions associated with different levels of aerobic capacity.

Sedentary HCR and LCR exhibit substantial differences in aerobic capacity, disease susceptibility and mortality rate. LCR are relatively hypertensive, have poorer vascular function and higher fasting blood-levels of glucose, insulin, triacylglycerides and free fatty acids [3]. LCR are also more susceptible to environmental challenges such as ischaemia/reperfusion [4], high-fat diet [5] or hypoxia [6]. Exposure to hypoxia (7 % oxygen) induces cardiac pump failure more rapidly in LCR than HCR [6]. That is, LCR experience fatal irregular ventricular contractions within 8 min of hypoxia, in contrast, pump failure in HCR occurs after 20 min of hypoxia and is characterised by a gradual diminution of cardiac systolic performance. At the cellular level, cardiomyocytes from LCR are significantly shorter, contract less quickly and relax more slowly than those isolated from HCR [6]. These differences are accompanied by greater expression of β-myosin heavy chain (β-MyHC) and differences in $Ca^{++}$ transients, but the expression of proteins associated with $Ca^{++}$ handling; sarco-endoplasmic reticulum ATPase, phospholamban, $Na^{+}$-$Ca^{++}$ exchanger or calsequestrin is similar in LCR and HCR hearts [6]. Therefore the mechanisms responsible for contractile dysfunction associated with selection on low running capacity are not yet fully understood.

Molecular differences that regress with low running capacity may provide further mechanistic insight to the elevated disease risk and cardiac dysfunction observed in LCR rats and individuals with low aerobic capacity. Microarray analysis [7] of HCR-LCR hearts detected 1540 differences in gene expression and highlighted enrichment of genes involved in lipid metabolism in HCR, whereas, glucose metabolism and microtubule-based processes were enriched in LCR. However, alterations in gene expression do not necessarily result in concomitant changes in protein abundance and post-translational processes can give rise to protein species that are undetectable in microarray analyses. The protein complement of a cell (i.e. its proteome) defines that cell and dictates its functional characteristics. Therefore it is of mechanistic value to verify if the differences in gene expression are also evident at the protein level, and determine protein modifications that regress with differences in running capacity.

## 2. Materials and Methods

### 2.1 Animal model

The inception of HCR-LCR strains from a founder population of genetically heterogeneous N:NIH rats has been described in detail [1]. In the current work, hearts from adult male HCR and LCR rats (n = 6 in each group) from generation 22 were used. Animals were housed (2 per cage) in accordance with the University of Michigan Committee guidelines on the use and care of animals. Environmental conditions were 20 ± 2 °C, 40-50 % relative humidity with a 12 h light (0600-1800) and dark cycle. Food and water were available ad

libitum. Home cage activity was measured gravimetrically during a 72 h period, as described previously [8]. Non-fasting blood samples were collected prior to the animals being killed. Serum glucose levels were measured using an Accu-Chek glucose meter (Roche Diagnostics, Basel, Switzerland), serum insulin concentrations were determined by ELISA using a commercially available rodent assay (Linco Research Inc. St Charles, MO).

## 2.2 Muscle processing

Animals were killed by cervical dislocation and their hearts isolated and washed with saline before being blotted dry and weighed. Cardiac septa were pulverised in liquid nitrogen then homogenised on ice in Lysis Buffer: 7 M urea, 2 M thiourea, 4 % (w/v) CHAPS, 30 mM Tris, pH 8.5 at 4° C, containing Complete™ protease inhibitor (Roche Diagnostics). After centrifugation at 12,000 g, 4 °C for 45 min supernatants were decanted, precipitated in acetone and resuspended in DIGE Lysis Buffer. Insoluble proteins were incubated in 1 % (v/v) SDS and at 90 °C for 30 min then washed twice in 0.1 % (w/v) Rapigest SF (Waters Corporation; Milford, MA) in 50 mM ammonium bicarbonate using 5 kDa $M_r$ cut-off spin columns. Protein suspensions were concentrated to ~50 µl then incubated at 80 °C for 15 min. DTT was added (final concentration 1 mM) and incubated at 60 °C for 15 min followed by incubation in the presence of 5 mM iodoacetamide at 4 °C, protected from light. Sequencing grade trypsin (Promega; Madison, WI) was added at a protein ratio of ~1:50 and digestion allowed to proceed at 37 °C overnight. Digestion was terminated by the addition of 2 µl concentrated HCl and peptide solutions centrifuged at 13,000 g for 5 min to clear supernatants in preparation for iTRAQ labelling and LC-MALDI MS.

The protein/ peptide concentrations were measured using the Bradford assay (Sigma, Poole, Dorset, UK). Samples prepared for DIGE were adjusted to 5 µg $\mu l^{-1}$ in lysis buffer.

## 2.3 DIGE of soluble cardiac proteins

Fifty microgram aliquots of each sample and the pooled internal standard were labelled with 400 pmol CyDye DIGE Fluor minimal dyes (GE Healthcare, Chalfont St Giles, UK), consistent with previous work [9]. Cy3 and Cy5 labelling was alternated between LCR and HCR samples in a 'balanced' design. Labelled LCR and HCR aliquots and pooled Cy2-labelled standard were combined with Rehydration Buffer: 7 M urea, 2 M thiourea, 2 % (w/v) CHAPS, 20 mM DTT and 0.5% (v/v) ampholytes. IPG strips (24 cm pH 3-11 non-linear; Immoboline Drystrip, GE Healthcare) were rehydrated overnight in 450 µl Rehydration Buffer that contained the combined Cyanine-labelled samples. Isoelectric focusing (maximum 50 µA per strip) was performed on an IPGPhor II (GE Healthcare) at 20 °C using the protocol (total 44000 Vh): 3 h at 300 V, 3 h at 600 V, 3 h at 1000 V, gradient to 8000 V in 3 h then 4 h at 8000 V. IPG strips were equilibrated in 50 mM Tris-HCl pH 8.8, containing 6 M urea, 30 % v/v glycerol, 2 % w/v SDS and bromophenol blue. DTT (10 mg $ml^{-1}$) was present in the first equilibration and iodoacetamide (25 mg $ml^{-1}$) in the second. Proteins were electrophoresed (Ettan Dalt six; GE healthcare) through denaturing 12.5 % poly-acrylamide gels at 20 °C; at 5 W per gel for 30 min, then 17 W per gel until the tracking dye reached the bottom edge.

## 2.4 SameSpots protein expression profiling

Gels were digitised (16-bit greyscale, 100 µm pixel size) immediately after electrophoresis using a fluorescence scanner (Typhoon 9400, GE Healthcare) at wavelengths appropriate for Cy2, Cy3 and Cy5 dyes. Gel images (Cy2, Cy3 and Cy5, total = 18) were aligned using SameSpots (Nonlinear Dynamics, Newcastle, UK). Prominent spots (mean ± SD per gel image: 548 ± 36) were used as vectors to warp each image to a common reference gel. A mask was used to remove gel artefacts (109 spots deleted) and features with an average normalised volume <10,000 (147 spots) or spot area < 150 (33 spots) were disregarded. In

total, 957 spots were included in the subsequent statistical analysis. As a measure of technical variation, when ranked by coefficient of variation, 70 % of spots in Cy2 images (pooled standard) had a coefficient of variation less than 14 %.

Log transformed spot volumes, expressed relative to the pooled standard, were used to investigate differences in expression between LCR and HCR groups by one-way analysis of variance. To control false discovery rate (FDR), the p-value distribution was used to calculate q-values and a criterion FDR of 10 % was set. This statistical approach, which considers the biological variation across each spot, is more sophisticated than arbitrary implementation of a threshold based on fold-change. Associations between serum insulin levels and differentially expressed spots (i.e. $p<0.05$, FDR <10 %) were investigated by Pearson correlation of HCR and LCR data using Prism 4.0 (GraphPad Software Inc. La Jolla, CA).

### 2.5 Identification of gel spots using MALDI-MS/MS and ESI-MS/MS

Spots were robotically excised from colloidal Coomassie-stained (Bio-Safe; Bio-Rad, Hercules, CA, USA) preparative gels loaded with 1.5 mg protein (pooled standard) and proteins identified by MADLI-MS/MS analysis of in-gel tryptic digests, as described previously [10]. Peak lists were searched against the Swiss-Prot database (57.1) restricted to 'Rattus' (7347 sequences) using a locally implemented Mascot (www.matrixscience.com) server (version 2.2.03). The enzyme specificity was trypsin allowing 1 missed cleavage, carbamidomethyl modification of cysteine (fixed), oxidation of methionine (variable) and an m/z error of ± 0.5 Da. Identification was accepted based on a significant Mowse score; the threshold ($P<0.05$) was 51 using the described database constraints. Up to 3 confirmatory MS/MS spectra were automatically acquired from digests with Mowse scores <120. Selection of MS/MS fragment ion spectra (average m/z) was restricted to 42 ions over 6 segments encompassing 5 – 95 % of the precursor ion m/z. MS/MS ion lists were searched against the Swiss-Prot database (tolerance ± 0.5 Da parent, ± 0.8 Da fragments) using the constraints described for peptide mass fingerprinting.

Selected spots were analysed using nano LC-ESI-MS/MS. Tandem electrospray mass spectra were recorded using a quadrupole-high capacity ion-trap (HCT Ultra ETD II; Bruker Daltonics, Bremen, Germany) coupled to a nano-flow HPLC system (Ultimate 3000; Dionex, Sunnyvale, CA). Five microlitres of sample (in-gel digest) was injected and peptides eluted using an ACN/ 0.1 % FA gradient of 2-40 % B in 30 min at a flow rate of 300 nl $min^{-1}$ (column Acclaim PepMap 100, C18, 3μm, 100 Å, 75μm * 15 cm, Dionex). An online nanospray source (Bruker Daltonics) was used equipped with a fused silica emitter (PicoTip FS360-20-10-D, New Objective). The capillary voltage was –1300 V and a survey scan from 300 m/z to 1500 m/z was used to select peptides with charge states of +2 or +3 using Enhanced Scan mode (8100 (m/z)/s). Data dependent MS/MS analysis in alternating CID/ETD mode was performed selecting the two most abundant precursor ions with active exclusion enabled. Data were searched against the Swiss-Prot database (tolerance ± 1.3 Da parent, ± 0.5 Da fragments) using the constraints described for peptide mass fingerprinting with phosphorylation of serine, threonine or tyrosine residues as possible modifications.

### 2.6 Analysis of insoluble cardiac proteins using iTRAQ and LC-MALDI MS/MS

Ten microgram aliquots of LCR and HCR samples and pooled internal standard were labelled with iTRAQ reagents according to the manufacturer's instructions (Applied Biosystems; Foster City, CA). HCR and LCR samples were labelled with iTRAQ 114, 115 or 116 in a 'balanced' design and combined with the pooled standard labelled with iTRAQ 117 to produce 4-plex experiments. Samples, labelled with iTRAQ reagents, were analysed in triplicate, as described previously [11]. LC-MALDI automation software selected 500

precursors in order of descending intensity from chromatographic peaks of at least 20 seconds width. MS/MS spectra were analysed for protein identification and parsed with spectra of iTRAQ reporter ions (m/z 114, 115,116 and 117) collected using the instrument's low-mass zoom feature. Data were searched against the Swiss-Prot database using the constraints described for peptide mass fingerprinting with the addition of quantitation (iTRAQ 4-plex) as a possible modification. Proteins with 2 or more iTRAQ peptides with ions scores greater than the identity threshold were selected for quantitation. Data expressed relative to the pooled internal standard were averaged from triplicate analyses to create protein lists for each animal. Statistically significant differences were determined using Log transformed expression data.

### 2.7 Muscle enzyme activity

β-(3)-Hydroxyacyl-CoA dehydrogenase activity was measured in muscle homogenates prepared in 100 mM $K_2HPO_4$ and 2 mM EDTA, pH 7.3 [12]. Homogenates were diluted in 100 mM $K_2HPO_4$ pH 7.3 containing 6.4 mM β-NADH and mixed with 5.4 mM S-acetoacetyl-CoA to a final concentration of 97 mM $K_2HPO_4$, 0.1 mM β-NADH, 0.09 mM S-acetoacetyl-CoA. The rate of disappearance of NADH was determined by measuring the change in absorbance at 340 nm for 5 min at 37 °C.

### 2.8 Western blot analysis

Muscle powders were homogenised on ice in 1 % (w/v) SDS containing Complete™ protease inhibitor (Roche Diagnostics) and centrifuged at 12,000 g, 4 °C for 20 min. Supernatants were adjusted to 5 μg $\mu l^{-1}$ protein in 2x Laemmli buffer and incubated at 90 °C for 5 min. Seventy-five microgram aliquots of HCR and LCR proteins were separated by SDS-PAGE and electro-transferred to PVDF membranes. For immunoblotting of 2-DE gels, 500 μg protein prepared for preparative gels was loaded on 13 cm, pH 3-10 IPG strips (GE Healthcare) and transferred to PVDF membranes as described previously [13]. Membranes were blocked with 5 % non-fat dried milk (NFDM) in 0.05 % (v/v) Tween-20 in PBS (PBS-Tw) then incubated overnight at 4 °C with primary antibodies specific for either 4-hydroxynonenal (4-HNE; 1:1000 ab46544; Abcam, UK), alpha B-crystallin (1:10000 ab13497; ABcam) or alpha B-crystallin phosphorylated at serine 59 (1:2000 ab5577; Abcam) in 1 % NFDM PBS-Tw. Secondary antibodies were from Abcam (ab6741 [1:5000] and ab6721 [1:3000]) and immunoreactive proteins were detected by enhanced chemiluminescence (ECL Prime, GE Healthcare) and quantified by densitometry. Equivalent transfer of proteins was assessed by post-staining PVDF membranes with Ponceau S or Direct Blue 71.

### 2.9 On-membrane MALDI-MS/MS

Protein spots visualised by anti-4-HNE immunoblotting after 2-DE were identified by on-membrane MALDI-MS/MS. Membranes were striped by incubation (45 min at 50 °C) in 100 mM DTT, 2 % (w/v) SDS 52.5 mM Tris pH 6.7 and affixed to MALDI target plates using conductive tape. Sequencing grade modified trypsin (0.2 μg/μl in 25 mM Ammonium bicarbonate; Promega) was dispensed at the positions (x,y coordinates) of immunoreactive proteins and prominent 'landmark' spots using a piezoelectric printing device (ChIP, Chemical Inkjet printer, Shimadzu Biotech). One nano-litre spots were printed (10 intervals of 100 pl) at a spacing of 500 μm within the border of each protein position and on-membrane digestion was allowed to proceed overnight at 37 °C in a humidified chamber. Spot positions were over-printed (2 nl spots dispensed in intervals of 100 pl) with matrix solution (5 μg/μl α-cyano-4-hydroxcinnamic acid in 50:50 ACN and 0. 1 % trifluoroacetic acid) and a calibration mix (Proteomass; Sigma) consisting of bradykinin (m/z 756.39), angiotensin II (m/z 1046.54), P14R (m/z 1533.85) and ACTH fragment (m/z 2465.19) was pipetted (0.5 μl) in the target-plate calibration wells and overlaid with 0.5 μl matrix solution.

Membranes were dried at 37 °C for 1 h prior to MALDI-MS/MS analysis directed at each protein coordinate. Data were searched against the Swiss-Prot database using the constraints described for peptide mass fingerprinting.

## 3. Results

### 3.1 Exercise capacity of HCR and LCR

Rats in the founder population ran 355 ± 11 m during a standardised run to exhaustion on a motorised treadmill [1]. Using an identical procedure, HCR rats from generation 22 ran 6.2-fold further than LCR when they were 3 months old (1864 ± 40 vs 299 ± 6 m) and 4-fold further (754 ± 24 vs 185 ± 10 m) prior to isolation of the hearts at 12 months of age. At each assessment HCR performed significantly more running work compared to LCR. Habitual cage activity during either light or dark periods was not significantly different in HCR and LCR (Supplementary Table 1). Serum glucose (mg/dl) was not different (196.3 ± 23 in HCR versus 195.3 ± 22 in LCR) whereas serum insulin (ng/ml) was significantly ($P<0.0001$) greater in LCR (4.84 ± 0.5) compared to HCR (1.32 ± 0.2). Gross and lean body weights of HCR were less than LCR and the absolute wet weight of the hearts was significantly greater in HCR, equating to a 1.5-fold greater heart: lean body mass ratio (Supplementary Table 1).

### 3.2 DIGE analysis of soluble cardiac proteins

Differential analysis was performed on 957 spots and the expression of 67 gel spots was significantly (FDR <10 %) different between HCR and LCR. Mass spectrometry unambiguously identified proteins in 369 gel spots (Supplementary Table 2), including 56 of the differentially expressed spots. A 2D gel map of the proteins identified by mass spectrometry is available in the World 2D-PAGE repository (http://world-2dpage.expasy.org/repository/), accession number 0022 and a representative pseudo-colour image of the DIGE separation is available online (Supplementary Figure 1). This proteome mining work revealed majority of proteins were resolved as multi-spot series. Thus the 369 identified gel-spots comprise 170 gene products and, in total, DIGE and MS analysis identified 39 gene products that were differentially expressed between HCR and LCR (Table 1). Further to differences between HCR and LCR group means, significant correlation was found between individual serum insulin concentrations in LCR and the abundance of coronin-6 (spot 1; $R^2=0.7410$, $P=0.0277$) and propionyl-CoA carboxylase α-chain (spot 23; $R^2$ 0.662, $P=0.0489$) whereas low serum insulin levels in HCR correlated with the abundance of hydroxyacyl-CoA dehydrogenase (spot 80; $R^2=0.907$, $P=0.0034$), succinyl-CoA: 3-ketoacid coenzyme A ($R^2=0.8587$, $P=0.0079$), ventricular myosin regulatory light chain (spot 83; $R^2=0.8295$, $P=0.0116$) and protein disulfide-isomerase A3 (Spot 45; $R^2=0.6869$, $P=0.0415$).

### 3.3 iTRAQ analysis of insoluble cardiac proteins

Twenty-eight proteins met the inclusion criteria for iTRAQ analysis (Supplementary Table 3) and the expression of 6 proteins differed significantly ($P<0.05$) between HCR and LCR (Table 3). iTRAQ detected 11 gene products (3-keto acyl-CoA thiolase, ADP/ATP translocase 1, citrate synthase, cytochrome c oxidase subunit 4 isoform 1, cytochrome c oxidase subunit 5A, cytoplasmic dynein, fatty acid-binding protein, haemoglobin subunit alpha-1/2, haemoglobin subunit beta-1, myosin-6 and myosin-7) not observed using DIGE. Four of these proteins (3-keto acyl-CoA thiolase, ADP/ATP translocase 1, cytochrome c oxidase subunit 4 isoform 1 and cytochrome c oxidase subunit 5A) were differentially expressed between HCR and LCR. Three proteins (ATP synthase subunit alpha, isocitrate dehydrogenase [NADP] and myosin regulatory light chain 2) identified as differentially expressed using DIGE were also detected in the iTRAQ analysis. In each case no significant

difference was detected using iTRAQ, which is consistent with our previous experience [11].

### 3.4 Biochemical and immunoblot analyses

The proteomic analyses point to differences in β-oxidation of fatty acids and exposure to oxidative stress. β-(3)-Hydroxyacyl-CoA dehydrogenase activity is commonly used as a measure of β-oxidation capacity and was 1.3-fold greater in HCR (75.4 ± 3.3 versus 56.7 ± 3.3 μmol/g/min, P=0.002), whereas, proteins associated with exposure to $H_2O_2$ and lipid peroxidation, such as 4-HNE, were detected in LCR hearts. Western blotting detected a prominent 4-HNE reactive band at ~45 kDa that was significantly (P=0.015) more abundant in LCR (Figure 1). Immunoblot analysis of 2-DE separated proteins revealed two 4-HNE immunoreactive spots in LCR hearts (Figure 1D), which were each identified as long-chain specific acyl-CoA dehydrogenase (ACADL) and correspond to spots 651 and 627 of the DIGE map. α B-crystallin is an abundant small heat shock protein in cardiac muscle and was resolved as 3 discrete isoelectric species (spots 20, 640 and 717) using DIGE. Spot 20 was 1.67-fold greater in LCR and nLC-ESI-MS/MS revealed this spot contained peptides phosphorylated at $S^{59}$ that were not present in spot 640 or 717 (Figure 2). Western blotting further confirmed the ratio of phosphorylated ($S^{59}$) to total α B-crystallin was significantly greater in LCR (Figure 2F).

## 4. Discussion

Proteomic analysis revealed selection on running capacity is associated with changes in cardiac energy metabolism and discovered the first evidence that the cardiac proteome of LCR is exposed to greater oxidative stress. These findings closely support the conclusions of our [7] microarray investigation but, in the majority, proteomic analysis discovered new information rather than simply providing protein-level confirmation of earlier reported [7] differences in gene expression. The limited direct correspondence between proteome and transcriptome data is, in part, due to the larger number of genes assayed by microarray compared to DIGE. In addition, it is important to recognise the majority of proteins were resolved as multiple spots using 2-DE, therefore, differences in spot expression may indicate differences in post-translational modification and provide information not evident at the transcriptome level.

The current work revealed conspicuous differences in β-oxidation enzymes suggesting an enhanced capacity for fatty acid oxidation in HCR. Specifically, DIGE detected pronounced modulation of each enzyme of the matrix component of the β-oxidation pathway (Table 1) responsible for oxidation of short-chain fatty acyl-CoA [14]. Accordingly, β-Hydroxyacyl-CoA dehydrogenase activity was significantly greater in HCR and the abundance of this key enzyme closely correlated with low serum insulin concentrations in these animals. Modulation of short-chain specific acyl-CoA dehydrogenase (ACADS) was a prominent feature in the current analysis. ACADS was observed as multiple spots (4, 10, 11 and 14) and the abundance of the most basic species (spot 11) was 2-fold greater in LCR, whereas, spots 4, 10 and 14 were 2-fold greater in HCR. These isoelectric species may represent different post-translational states, for example phosphorylated species of ACADS have been reported [15]. In the current work nLC-ESI-MS/MS analysis identified 84 % of the ACADS sequence (57 unique peptides, MOWSE score = 2806), but did not include phosphorylated residues. Nonetheless, the ACADS spot pattern in HCR closely resembles that of endurance-trained rats [10], and putative post-translational modulation of ACADS is a consistent feature associated with changes in cardiac performance. For example, genetic deletion of phospholamban, which produces a hyper-dynamic cardiac phenotype, is associated with a similar shift in ACADS spot pattern [16], and elevated expression of low-pI ACADS spots in hearts of swim-trained rats correlates with enhanced ACADS activity [17].

Differential expression of β-oxidation enzymes coincided with differences in electron transfer proteins and components of the respiratory chain. This may indicate tighter coupling specifically between β-oxidation of short-chain fatty acyl-CoA and oxidative phosphorylation that is important for enhanced aerobic capacity. Acyl-CoA dehydrogenases reduce FAD (FADH2) and transfer electrons to electron transfer flavoprotein-ubiquinone oxidoreductase via electron transfer flavoprotein-α, driving the reduction of ubiquinone. Both electron transfer flavoprotein-α and electron transfer flavoprotein-ubiquinone oxidoreductase were identified as multi-spot series and the expression of individual spots within these series was significantly different in HCR and LCR hearts. In addition, subunit 2 of cytochrome b-c1, was 1.22-fold less in LCR and iTRAQ detected 1.9-fold lower abundance of subunit 1 of cytochrome b-c1 and ~1.67-fold lesser abundance of cytochrome c oxidase subunits 4 and 5A in LCR. These findings are consistent with enrichment of fatty acid oxidation genes reported in the HCR transcriptome [7]. Furthermore, DIGE provided protein-level confirmation of the elevated mRNA expression of acyl-CoA thioesterase 2 (mitochondrial thioesterase 1: MTE-1) in HCR. MTE-1 hydrolyses fatty acyl-CoA to its free fatty acid anion, which may be exported from the mitochondria [18]. MTE-1 is also modulated in endurance-trained hearts [10] and this adaptation may serve to protect the inner mitochondrial membrane by preventing accumulation of acyl-carnitine, which is a strong surfactant and can damage lipid membranes [19].

The observed differences in metabolic enzymes may be related to impaired peroxisome proliferator activated receptor-α (PPARα) signalling in LCR. The PPAR gene ontology pathway was overrepresented in microarray analysis [7] of HCR-LCR hearts, and some genes of the modulated proteins involved in β-oxidation (Table 1) have PPAR responsive elements [20]. Interestingly, metabolomic profiling of PPARα-/- mice [21] reveals a compensatory increase in leucine metabolism. Degradation of branched-chain amino acids (BCAA), in particular leucine, is ketogenic and DIGE detected modulation of methylcortonyl-CoA carboxylase subunit-α and propionyl-CoA carboxylase α-chain, which contribute to BCAA degradation or β-oxidation of odd chain-length and methyl-branched fatty acyl-CoA. Moreover, succinyl-CoA: 3-ketoacid-coenzyme A transferase 1 (spot 17), which catalyses the rate-limiting step in ketolysis, was more abundant in LCR hearts and low serum insulin concentrations in HCR correlated with low abundance of this enzyme. The second reaction in ketolysis is catalysed by acetyl-CoA acetyltransferase and this protein was detected in 2 spots, the abundance of spot 41 was 1.5-fold less, whereas, spot 152 was 1.2-fold greater in LCR suggesting a shift in post-translational state of this enzyme. As yet, it is not possible to determine whether these findings depict altered amino acid degradation only, or concurrent differences in ketone body and amino acid degradation in LCR.

The lesser abundance of β-oxidation and electron transport chain enzymes in LCR indicates diminished mitochondrial capacity, which may be associated with greater oxidative stress and $H_2O_2$ production that can cause irreversible damage to macromolecules. Indeed, catalase, which scavenges $H_2O_2$, was significantly more abundant in LCR and 2 proteins, aldehyde dehydrogenase-2 and glutathione S-transferase, that defend against lipid peroxidation products such as 4-HNE were also differentially expressed. Phosphorylation of aldehyde dehydrogenase-2 protects against ischaemic damage to the heart [22] and this protein was modulated in a manner suggesting post-translational modification in LCR, whereas, glutathione S-transferase was less abundant in LCR, which is consistent with findings in hearts of spontaneously hypertensive rats [23]. Thyfault et al. [24] reports elevated catalase activity and higher levels of 4-HNE modified proteins associated with hepatic steatosis in LCR. Using a similar approach, we detected a greater abundance of 4-HNE-modified ACADL in LCR heart (Figure 1), which may be linked to the observed modulation of enzymes specifically responsible for β-oxidation of short-chain fatty acids.

Recently, Tweedie et al. [25] reported greater activity of catalase and superoxide dismutase in skeletal muscle of HCR, which is in contrast to the findings in heart and liver. In the current work, superoxide dismutase and other key enzymes of the antioxidant defense system, including glutathione peroxidase, were detected by DIGE but were not different in HCR and LCR.

Selection on running capacity resulted in differential expression of chaperone proteins, including protein disulfide isomerase A3, heat shock protein beta-7 (cardiovascular heat shock protein: cvHSP) and α B-crystallin. The expression of cvHSP was 2.33-fold less in LCR and protein disulfide isomerase A3, which was resolved as 3 spots, was on average 1.6-fold less in LCR. α B-crystallin was also resolved as 3 spots (20, 640 and 717) and the abundance of phosphorylated ($S^{59}$) α B-crystallin (spot 20) was 1.67-fold greater in LCR (Figure 2). Phosphorylation of α B-crystallin at $S^{59}$ is associated with protection against a variety of different environmental stresses. For example, exposure of cardiomyoblasts to $H_2O_2$ activates the p38 MAPK/MSK1 pathway resulting in phosphorylation of α B-crystallin $S^{59}$ [26], similarly ischaemia/ reperfusion injury activates p38 MAPK and is associated with $S^{59}$ phosphorylation [27]. Therefore, the greater phosphorylation of α B-crystallin $S^{59}$ observed here (Figure 2) is consistent with enrichment of the stress-activated p38 MAPK signalling in the LCR cardiac transcriptome [7].

Phosphorylated α B-crystallin is also associated with cytoskeletal stress where upon this chaperone translocates and interacts with desmin [28]. Elevated expression of desmin is a feature of human cardiomyopathy [29] and was also observed in LCR hearts (Table 1). Moreover, cytoskeletal stress induced by pressure overload increases microtubule density [30], while overexpression of α B-crystallin helps to maintain cardiomyocyte microtubule integrity [31]. Coronin-6 exhibited the greatest difference (4.4-fold greater in LCR) in abundance between HCR and LCR and correlated significantly with high serum insulin levels in LCR. Coronin proteins cross-link actin filaments [32], and may also interact with microtubules linking this novel discovery with over-representation of microtubule-based processes in LCR [7]. The role of coronin-6 in cardiac muscle is yet to be described but changes in the expression of coronin 1 are associated with the transition to maladaptive cardiac hypertrophy induced by transgenic overexpression of myotrophin[33]. Changes in microtubule proteins are consistent with elevated oxidative stress [34] and proteasomal dysfunction [35], which may contribute to the development of cardiomyopathy [36]. Accordingly, LCR exhibited greater expression of proteasome activator complex subunit 1 (PA28α). Cardiac expression of PA28α is elevated in diabetic cardiomyopathy induced by streptozotocin (STZ), and Powell et al [37] reports the 11S-activated proteasome, rather than the 26S-proteasome, is associated with the degradation oxidatively modified proteins.

Post-genomic investigations afford comprehensive unbiased analysis that can advance knowledge and generate novel hypotheses. However, despite using sophisticated statistical analyses, the large number of dependent variables often raises concerns regarding false discoveries. Therefore, validation is important. Duplication of proteomic experiments or the inclusion of technical replicates does not provide validation of the biological significance of the original discoveries. Instead, comparison of data across associated biological models is a more relevant approach. Features of the HCR cardiac proteome, including ACADS and MTE-1, were identical to endurance-trained rat hearts [38] and, therefore, may represent key molecular events associated with enhanced aerobic capacity. Similarly, LCR exhibited molecular features consistent with animal models of cardiomyopathy. For example, Turko et al.[39] reports modulation of β-oxidation enzymes and differential expression of voltage-dependent anion channel 1 and catalase in hearts of STZ-induced diabetic rats. Faber et al. [40] found differences in desmin and α B-crystallin expression after pressure-induced cardiomyopathy. Bugger et al.[41] reports robust down regulation of enzymes involved in

fatty acid oxidation after 20 wk of aortic constriction, and Meng et al.[42] reveals changes in fatty acid metabolism contribute to cardiac hypertrophy and precede the development of hypertension in spontaneously hypertensive rats.

The correspondence between LCR and models of cardiomyopathy is testament to the importance of low aerobic capacity in the development of disease. Our discoveries support the thesis that, in normal populations, a decline in physical activity resulting in low aerobic capacity is the antecedent of hypertension and overt cardiac dysfunction. In humans, heart failure is a progressive and complex condition involving both inherited and environmental factors that co-occurs with other complicated risks including hypertension, obesity, and insulin resistance. Thus, the use of acute models or inbred strains such as SHR is limited. The LCR-HCR model retains genetic complexity. LCR manifest many of the features associated with metabolic syndrome, and here we show the cardiac proteome of LCR is exposed to greater oxidative stress and metabolic perturbations that may underlie their diminished cardiac performance. In contrast, despite low levels of habitual physical activity, HCR appear to be uniquely protected from cardiometabolic disease and demonstrate an enhanced capacity to utilise fatty acids for cardiac work similar to endurance-trained animals. As such, further interrogation of this model may provide additional clues or even reveal the molecular mechanisms that underlie the complexity of low heart function as a precursor to heart failure.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by The Institute for Health Research, Liverpool John Moores University (JGB, LGK and SLB) and grants R17718 (LGK and SLB) and 5R01DK7720 (CFB, SLB and NQ) from the National Center for Research Resources of the National Institutes of Health and the Animal Phenotyping Core of the Michigan Metabolomics and Obesity Center.

## Abbreviations

| | |
|---|---|
| **4-HNE** | 4-hydroxynonenal |
| **ACADS** | short-chain specific acyl-CoA dehydrogenase |
| **ACADL** | long-chain specific acyl-CoA dehydrogenase |
| **Acyl-CoA** | acyl-co enzyme A |
| **BCAA** | branched-chain amino acid |
| **cvHSP** | cardiovascular heat shock protein |
| **ETD** | electron transfer dissociation |
| **FAD** | flavin adenine dinucleotide |
| **FDR** | false discovery rate |
| **HCR** | high-capacity runner |
| **LCR** | low-capacity runner |
| **MTE-1** | mitochondrial thioesterase |
| **MAPK** | mitogen activated protein kinase |
| **MSK1** | mitogen and stress activated protein kinase |

| | |
|---|---|
| **MyHC** | myosin heavy chain |
| **NFDM** | non-fat dried milk |
| **PA28α** | Proteasome activator subunit 1 |
| **SHR** | spontaneously hypertensive rat |
| **STZ** | streptozotocin |

## References


1. Koch LG, Britton SL. Artificial selection for intrinsic aerobic endurance running capacity in rats. Physiol Genomics. 2001; 5:45–52. [PubMed: 11161005]
2. Carlborg O, Jacobsson L, Ahgren P, Siegel P, Andersson L. Epistasis and the release of genetic variation during long-term selection. Nat Genet. 2006; 38:418–420. [PubMed: 16532011]
3. Wisloff U, Najjar SM, Ellingsen O, Haram PM, et al. Cardiovascular risk factors emerge after artificial selection for low aerobic capacity. Science. 2005; 307:418–420. [PubMed: 15662013]
4. Lujan HL, Britton SL, Koch LG, DiCarlo SE. Reduced susceptibility to ventricular tachyarrhythmias in rats selectively bred for high aerobic capacity. Am J Physiol Heart Circ Physiol. 2006; 291:H2933–H2941. [PubMed: 16891405]
5. Noland RC, Thyfault JP, Henes ST, Whitfield BR, et al. Artificial selection for high-capacity endurance running is protective against high-fat diet-induced insulin resistance. Am J Physiol Endocrinol Metab. 2007; 293:E31–E41. [PubMed: 17341547]
6. Palpant NJ, Szatkowski ML, Wang W, Townsend D, et al. Artificial selection for whole animal low intrinsic aerobic capacity co-segregates with hypoxia-induced cardiac pump failure. PLoS One. 2009; 4:e6117. [PubMed: 19568432]
7. Bye A, Langaas M, Hoydal MA, Kemi OJ, et al. Aerobic capacity-dependent differences in cardiac gene expression. Physiol Genomics. 2008; 33:100–109. [PubMed: 18171719]
8. Biesiadecki BJ, Brand PH, Koch LG, Britton SL. A gravimetric method for the measurement of total spontaneous activity in rats. Proc Soc Exp Biol Med. 1999; 222:65–69. [PubMed: 10510247]
9. Nelson MM, Jones AR, Carmen JC, Sinai AP, et al. Modulation of the host cell proteome by the intracellular apicomplexan parasite Toxoplasma gondii. Infect Immun. 2008; 76:828–844. [PubMed: 17967855]
10. Burniston JG. Adaptation of the rat cardiac proteome in response to intensity-controlled endurance exercise. Proteomics. 2009; 9:106–115. [PubMed: 19053138]
11. Holloway KV, O'Gorman M, Woods P, Morton JP, et al. Proteomic investigation of changes in human vastus lateralis muscle in response to interval-exercise training. Proteomics. 2009; 9:5155–5174. [PubMed: 19834892]
12. Lynen, F.; Wieland, O. Methods Enzymol. Academic Press; 1955. [beta]-Ketoreductase.; p. 566-573.
13. Burniston JG. Changes in the rat skeletal muscle proteome induced by moderate-intensity endurance exercise. Biochimica et Biophysica Acta Proteins and Proteomics. 2008; 1784:1077–1086.
14. Liang X, Le W, Zhang D, Schulz H. Impact of the intramitochondrial enzyme organization on fatty acid oxidation. Biochem Soc Trans. 2001; 29:279–282. [PubMed: 11356167]
15. Mayr M, Chung YL, Mayr U, McGregor E, et al. Loss of PKC-delta alters cardiac metabolism. Am J Physiol Heart Circ Physiol. 2004; 287:H937–H945. [PubMed: 15277208]
16. Chu G, Kerr JP, Mitton B, Egnaczyk GF, et al. Proteomic analysis of hyperdynamic mouse hearts with enhanced sarcoplasmic reticulum calcium cycling. Faseb J. 2004; 18:1725–1727. [PubMed: 15358683]
17. Sun B, Wang JH, Lv YY, Zhu SS, et al. Proteomic adaptation to chronic high intensity swimming training in the rat heart. Comparative Biochemistry and Physiology-Part D. 2008; 3:108–117.

18. Gerber LK, Aronow BJ, Matlib MA. Activation of a novel long-chain free fatty acid generation and export system in mitochondria of diabetic rat hearts. Am J Physiol Cell Physiol. 2006; 291:C1198–C1207. [PubMed: 16855217]

19. Goni FM, Requero MA, Alonso A. Palmitoylcarnitine, a surface-active metabolite. FEBS Lett. 1996; 390:1–5. [PubMed: 8706815]

20. Finck BN. The PPAR regulatory system in cardiac physiology and disease. Cardiovasc Res. 2007; 73:269–277. [PubMed: 17010956]

21. Makowski L, Noland RC, Koves TR, Xing W, et al. Metabolic profiling of PPARalpha-/- mice reveals defects in carnitine and amino acid homeostasis that are partially reversed by oral carnitine supplementation. FASEB J. 2009; 23:586–604. [PubMed: 18945875]

22. Chen CH, Budas GR, Churchill EN, Disatnik MH, et al. Activation of aldehyde dehydrogenase-2 reduces ischemic damage to the heart. Science. 2008; 321:1493–1495. [PubMed: 18787169]

23. Zhou SG, Wang P, Pi RB, Gao J, et al. Reduced expression of GSTM2 and increased oxidative stress in spontaneously hypertensive rat. Mol Cell Biochem. 2008; 309:99–107. [PubMed: 18008142]

24. Thyfault JP, Rector RS, Uptergrove GM, Borengasser SJ, et al. Rats selectively bred for low aerobic capacity have reduced hepatic mitochondrial oxidative capacity and susceptibility to hepatic steatosis and injury. J Physiol. 2009; 587:1805–1816. [PubMed: 19237421]

25. Tweedie C, Romestaing C, Burelle Y, Safdar A, et al. Lower Oxidative DNA Damage Despite Greater ROS Production in Muscles from Rats Selectively Bred for High Running Capacity. Am J Physiol Regul Integr Comp Physiol. 2010

26. Aggeli IK, Beis I, Gaitanaki C. Oxidative stress and calpain inhibition induce alpha B-crystallin phosphorylation via p38-MAPK and calcium signalling pathways in H9c2 cells. Cell Signal. 2008; 20:1292–1302. [PubMed: 18420382]

27. Whittaker R, Glassy MS, Gude N, Sussman MA, et al. Kinetics of the translocation and phosphorylation of alphaB-crystallin in mouse heart mitochondria during ex vivo ischemia. Am J Physiol Heart Circ Physiol. 2009; 296:H1633–H1642. [PubMed: 19252088]

28. Launay N, Goudeau B, Kato K, Vicart P, Lilienbaum A. Cell signaling pathways to alphaB-crystallin following stresses of the cytoskeleton. Exp Cell Res. 2006; 312:3570–3584. [PubMed: 16979163]

29. Pawlak A, Gil RJ, Walczak E, Seweryniak P. Desmin expression in human cardiomyocytes and selected clinical and echocardiographic parameters in patients with chronic heart failure. Kardiol Pol. 2009; 67:955–961. [PubMed: 19838951]

30. Cooper G. Cytoskeletal networks and the regulation of cardiac contractility: microtubules, hypertrophy, and cardiac dysfunction. Am J Physiol Heart Circ Physiol. 2006; 291:H1003–H1014. [PubMed: 16679401]

31. Bluhm WF, Martin JL, Mestril R, Dillmann WH. Specific heat shock proteins protect microtubules during simulated ischemia in cardiac myocytes. Am J Physiol. 1998; 275:H2243–H2249. [PubMed: 9843825]

32. Rybakin V, Clemen CS. Coronin proteins as multifunctional regulators of the cytoskeleton and membrane trafficking. Bioessays. 2005; 27:625–632. [PubMed: 15892111]

33. Sarkar S, Leaman DW, Gupta S, Sil P, et al. Cardiac overexpression of myotrophin triggers myocardial hypertrophy and heart failure in transgenic mice. J Biol Chem. 2004; 279:20422–20434. [PubMed: 14970239]

34. Devillard L, Vandroux D, Tissier C, Brochot A, et al. Tubulin ligands suggest a microtubule-NADPH oxidase relationship in postischemic cardiomyocytes. Eur J Pharmacol. 2006; 548:64–73. [PubMed: 16973157]

35. Dong X, Liu J, Zheng H, Glasford JW, et al. In situ dynamically monitoring the proteolytic function of the ubiquitin-proteasome system in cultured cardiac myocytes. Am J Physiol Heart Circ Physiol. 2004; 287:H1417–H1425. [PubMed: 15105173]

36. Bulteau AL, Lundberg KC, Humphries KM, Sadek HA, et al. Oxidative modification and inactivation of the proteasome during coronary occlusion/reperfusion. J Biol Chem. 2001; 276:30057–30063. [PubMed: 11375979]

37. Powell SR, Samuel SM, Wang P, Divald A, et al. Upregulation of myocardial 11S-activated proteasome in experimental hyperglycemia. J Mol Cell Cardiol. 2008; 44:618–621. [PubMed: 18308332]

38. Burniston JG, Hoffman EP. Proteomic responses of skeletal and cardiac muscle to exercise. Expert Rev Proteomics. 2011 In Press.

39. Turko IV, Murad F. Quantitative protein profiling in heart mitochondria from diabetic rats. J Biol Chem. 2003; 278:35844–35849. [PubMed: 12851409]

40. Faber MJ, Dalinghaus M, Lankhuizen IM, Bezstarosti K, et al. Proteomic changes in the pressure overloaded right ventricle after 6 weeks in young rats: correlations with the degree of hypertrophy. Proteomics. 2005; 5:2519–2530. [PubMed: 15912512]

41. Bugger H, Schwarzer M, Chen D, Schrepper A, et al. Proteomic remodelling of mitochondrial oxidative pathways in pressure overload-induced heart failure. Cardiovasc Res. 2010; 85:376–384. [PubMed: 19843514]

42. Meng C, Jin X, Xia L, Shen SM, et al. Alterations of mitochondrial enzymes contribute to cardiac hypertrophy before hypertension development in spontaneously hypertensive rats. J Proteome Res. 2009; 8:2463–2475. [PubMed: 19265432]

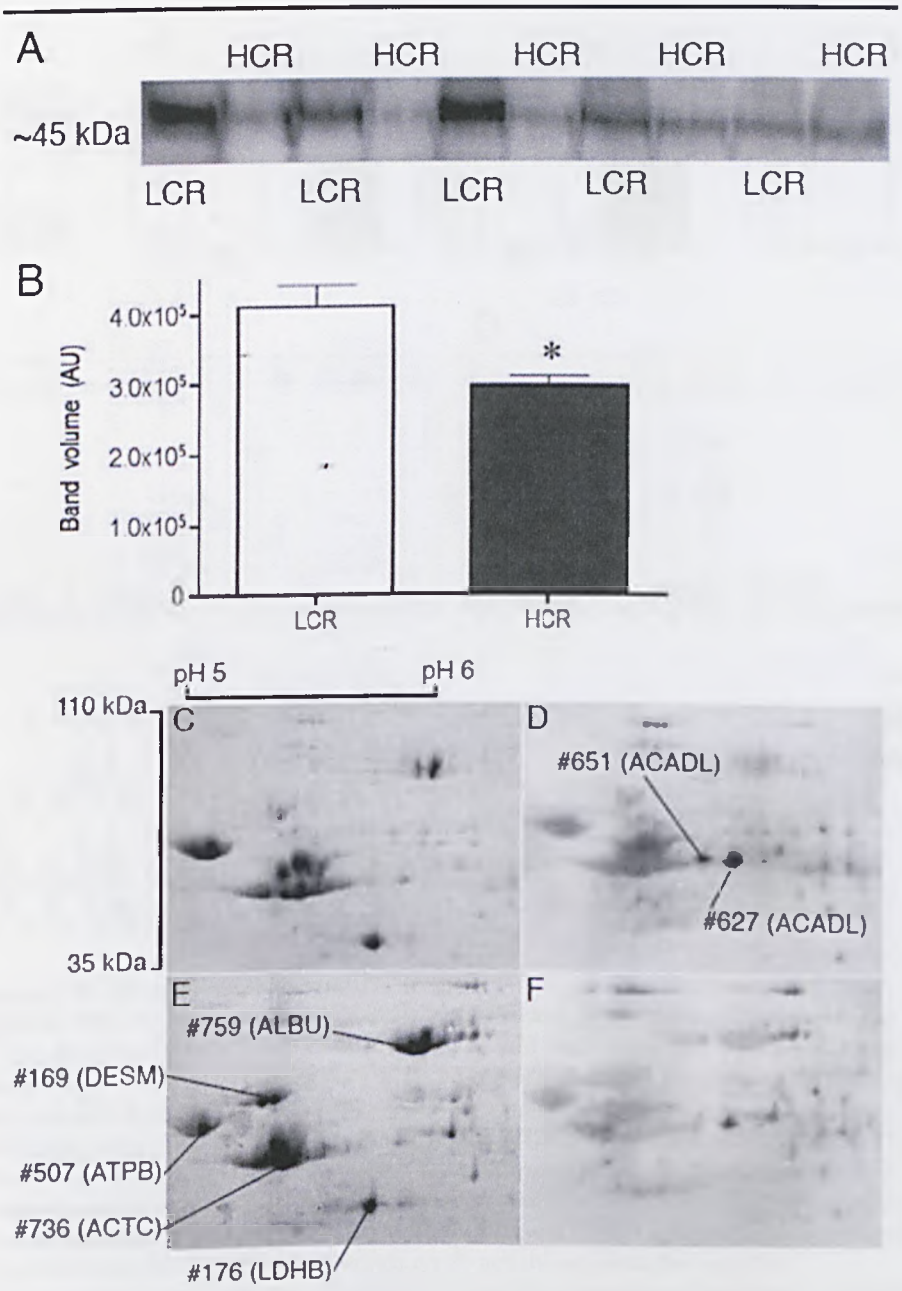



**Figure 1. 4-HNE modification of long-chain specific acyl-CoA dehydrogenase**

Western blot analysis (A) of 4-HNE modified proteins detected a prominent band at 45-50 kDa that was more abundant in LCR hearts (B). Homogenates of LCR (C and D) and HCR (E and F) hearts were separated by 2-DE and transferred to PVDF membranes. Proteins were visualised by Ponceau S staining (C and E) or immunoblot analysis of 4-HNE modified residues (D and F). 4-HNE immunoreactive spots and prominent spots stained by Direct Blue 71 were identified by MS analysis of on-membrane tryptic digests. Two 4-HNE immunoreactive spots, corresponding to #651 and #627 of the DIGE map, were more intensely labelled in LCR hearts (D) and were identified as long-chain specific acyl-CoA dehydrogenase (ACADL).

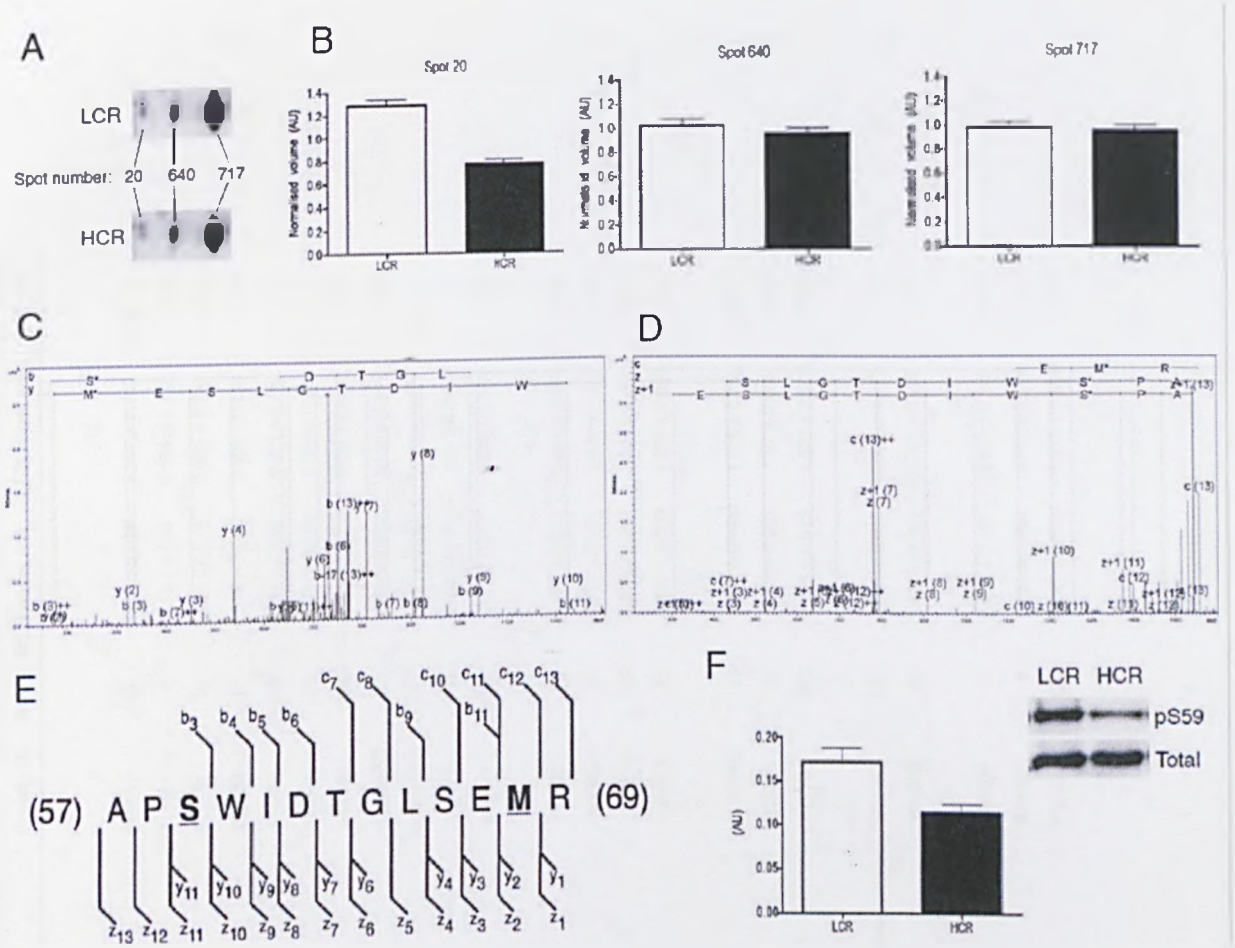



**Figure 2. Phosphorylation of α B-crystallin at S59**

Spots 20, 640 and 717 were unambiguously identified as α B-crystallin and the expression of spot 20 was 1.67-fold greater in LCR (A and B). nLC-MS/MS detected a doubly charged precursor (779.86 m/z) of peptide APSWIDTGLSEMR (residues 57-69) in spot 20 that was not present in spot 640 or spot 717. Site-specific mapping of serine phosphorylation was performed using (C) collision-induced dissociation (CID) and (D) electron-transfer dissociation (ETD) techniques. (E) Interpretation of fragment ion spectra confirmed oxidation of methionine 68 and phosphorylation of serine 59; bold underlined type denotes modified residues. MOWSE peptide scores from CID and ETD fragmentation were 55 and 80, respectively. (F) Western blot analysis confirmed the ratio of α B-crystallin phosphorylated serine 59 relative to total α B-crystallin was significantly greater (1.49-fold; P=0.02) in LCR hearts.

**Table 1**

DIGE expression data

| Description | Protein ID | Ref# | HCR norm. vol. | LCR norm. vol. | Fold diff. |
|---|---|---|---|---|---|
| **Energy metabolism** | | | | | |
| *Fatty acid metabolism* | | | | | |
| Short-chain specific acyl-CoA dehydrogenase | P15651 | 11 | 0.741 ± 0.018 | 1.517 ± 0.167 | +2.05 |
| Short-chain specific acyl-CoA dehydrogenase | P15651 | 4 | 1.432 ± 0.038 | 0.594 ± 0.232 | −2.41 |
| Short-chain specific acyl-CoA dehydrogenase | P15651 | 10 | 1.409 ± 0.055 | 0.665 ± 0.196 | −2.12 |
| Short-chain specific acyl-CoA dehydrogenase | P15651 | 14 | 1.198 ± 0.056 | 0.668 ± 0.118 | −1.79 |
| Enoyl-CoA hydratase, mitochondrial | P14604 | 105 | 1.138 ± 0.016 | 0.971 ± 0.01 | −1.17 |
| Enoyl-CoA hydratase, mitochondrial | P14604 | 127 | 1.021 ± 0.019 | 0.89 ± 0.021 | −1.15 |
| Enoyl-CoA hydratase, mitochondrial | P14604 | 174 | 0.963 ± 0.017 | 0.88 ± 0.013 | −1.09 |
| Hydroxyacyl-CoA dehydrogenase | Q9WVK7 | 80 | 1.378 ± 0.052 | 1.130 ± 0.048 | −1.22 |
| Hydroxyacyl-CoA dehydrogenase | Q9WVK7 | 143 | 1.175 ± 0.011 | 1.039 ± 0.036 | −1.13 |
| Acyl-coenzyme A thioesterase 2 | 055171 | 30 | 1.104 ± 0.056 | 0.7 ± 0.077 | −1.58 |
| Acyl-coenzyme A thioesterase 2 | 055171 | 61 | 1.113 ± 0.033 | 0.878 ± 0.023 | −1.27 |
| *Amino acid degradation* | | | | | |
| Propionyl-CoA carboxylase α chain | P14882 | 22 | 1.422 ± 0.052 | 0.861 ± 0.083 | −1.65 |
| Propionyl-CoA carboxylase α chain | P14882 | 23 | 0.873 ± 0.115 | 1.421 ± 0.043 | +1.63 |
| δ-1-Pyrroline-5-carboxylate dehydrogenase | P0C2X9 | 29 | 0.815 ± 0.049 | 1.293 ± 0.051 | +1.59 |
| Methylcrotonoyl-CoA carboxylase subunit α | Q5I0C3 | 53 | 0.841 ± 0.029 | 1.098 ± 0.067 | +1.31 |
| *Ketone body metabolism* | | | | | |
| Succinyl-CoA:3-ketoacid-coenzyme A transferase 1 | B2GV06 | 17 | 0.941 ± 0.088 | 1.628 ± 0.215 | +1.73 |
| Acetyl-CoA acetyltransferase | P17764 | 41 | 1.33 ± 0.065 | 0.92 ± 0.066 | −1.45 |
| Acetyl-CoA acetyltransferase | P17764 | 152 | 0.965 ± 0.015 | 1.082 ± 0.024 | +1.21 |
| *Glycolytic metabolism* | | | | | |
| Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex | P08461 | 147 | 1.053 ± 0.018 | 0.935 ± 0.028 | −1.13 |
| Glycogen phosphorylase, muscle form | P09812 | 47 | 1.117 ± 0.085 | 0.805 ± 0.051 | −1.39 |
| Tricarboxylic acid cycle | | | | | |
| Fumarate hydratase, mitochondrial | P14408 | 5 | 1.478 ± 0.067 | 0.620 ± 0.147 | −2.38 |
| Fumarate hydratase, mitochondrial | P14408 | 6 | 1.516 ± 0.062 | 0.649 ± 0.146 | −2.33 |
| Isocitrate dehydrogenase (NADP) | P56574 | 129 | 0.987 ± 0.033 | 1.131 ± 0.018 | +1.15 |

| Description | Protein ID | Ref# | HCR norm. vol. | LCR norm. vol. | Fold diff. |
|---|---|---|---|---|---|
| 2-Oxoglutarate dehydrogenase E1 component | Q5XI78 | 70 | 0.977 ± 0.043 | 1.218 ± 0.041 | +1.25 |
| *Electron transport chain* | | | | | |
| Succinate dehydrogenase [ubiquinone] iron-sulfur subunit | P21913 | 182 | 1.097 ± 0.005 | 1.073 ± 0.006 | −1.02 |
| Electron transfer flavoprotein subunit α | P13803 | 121 | 0.975 ± 0.021 | 0.843 ± 0.032 | −1.16 |
| Electron transfer flavoprotein-ubiquinone oxidoreductase | Q6UPE1 | 91 | 0.994 ± 0.041 | 0.835 ± 0.028 | −1.19 |
| Cytochrome b-c1 complex subunit 2 | P32551 | 77 | 1.197 ± 0.039 | 0.979 ± 0.041 | −1.22 |
| ATP synthase subunit α | P15999 | 164 | 0.994 ± 0.009 | 1.099 ± 0.027 | +1.11 |
| Voltage-dependent anion-selective channel protein 1 | Q9Z2L0 | 137 | 0.996 ± 0.026 | 0.876 ± 0.023 | −1.14 |
| **Antioxidant/stress response** | | | | | |
| Catalase | P04762 | 19 | 0.713 ± 0.049 | 1.212 ± 0.053 | +1.70 |
| Catalase | P04762 | 42 | 0.932 ± 0.072 | 1.329 ± 0.073 | +1.43 |
| Aldehyde dehydrogenase 2 | P11884 | 124 | 0.92 ± 0.028 | 1.058 ± 0.033 | +1.15 |
| Glutathione *S*-transferase Mu 2 | P08010 | 2 | 1.618 ± 0.189 | 0.542 ± 0.167 | −2.99 |
| Protein disulfide-isomerase A3 | P11598 | 13 | 1.472 ± 0.062 | 0.799 ± 0.15 | −1.84 |
| Protein disulfide-isomerase A3 | P11598 | 37 | 1.222 ± 0.054 | 0.809 ± 0.096 | −1.51 |
| Protein disulfide-isomerase A3 | P11598 | 45 | 1.262 ± 0.069 | 0.909 ± 0.055 | −1.39 |
| Heat shock protein β-7 (Fragment) | Q9QUK5 | 7 | 1.419 ± 0.259 | 0.61 ± 0.057 | −2.33 |
| α-Crystallin B chain | P23928 | 20 | 0.768 ± 0.094 | 1.282 ± 0.133 | +1.67 |
| **Myofibrillar and cytoskeletal** | | | | | |
| Myosin regulatory light chain 2, ventricular/cardiac muscle | P08733 | 66 | 0.988 ± 0.031 | 1.245 ± 0.038 | +1.26 |
| Myosin regulatory light chain 2, ventricular/cardiac muscle | P08733 | 83 | 0.950 ± 0.041 | 1.15 ± 0.043 | +1.21 |
| Desmin | P48675 | 60 | 0.88 ± 0.03 | 1.118 ± 0.049 | +1.27 |
| Desmin | P48675 | 169 | 1.022 ± 0.02 | 1.125 ± 0.023 | +1.10 |
| Vimentin | P31000 | 15 | 1.373 ± 0.222 | 0.771 ± 0.044 | −1.78 |
| EH domain-containing protein 1 | Q641Z6 | 25 | 0.755 ± 0.044 | 1.213 ± 0.029 | +1.61 |
| EH domain-containing protein 1 | Q641Z6 | 50 | 0.861 ± 0.036 | 1.147 ± 0.025 | +1.33 |
| Coronin-6 | Q920J3 | 1 | 0.446 ± 0.046 | 1.972 ± 0.485 | +4.42 |
| Rab GDP dissociation inhibitor α | P50398 | 56 | 1.149 ± 0.054 | 0.898 ± 0.047 | −1.28 |
| PDZ and LIM domain protein 1 | P52944 | 64 | 1.054 ± 0.036 | 0.835 ± 0.049 | −1.26 |
| **Miscellaneous** | | | | | |
| Annexin A3 | P14669 | 150 | 1.053 ± 0.029 | 0.937 ± 0.022 | −1.12 |

| Description | Protein ID | Ref# | HCR norm. vol. | LCR norm. vol. | Fold diff. |
|---|---|---|---|---|---|
| Apolipoprotein E | P02650 | 89 | 0.903 ± 0.041 | 1.076 ± 0.019 | +1.19 |
| Prostaglandin reductase 2 | Q5BK81 | 9 | 1.673 ± 0.221 | 0.733 ± 0.188 | −2.28 |
| Proteasome activator complex subunit 1 | Q63797 | 133 | 0.859 ± 0.01 | 0.98 ± 0.033 | +1.14 |
| Pyridoxine-5′-phosphate oxidase | O88794 | 39 | 1.443 ± 0.043 | 0.99 ± 0.14 | −1.46 |

Protein description and Protein ID relate to the Swiss-Prot database (57.1) entry identified from MASCOT searches of MS and MS/MS spectra. HCR and LCR expression ratios are average ± SD relative to the pooled internal standard. Fold difference relative to HCR values are reported for spots exhibiting significant ($p < 0.05$) differences in expression at an FDR of <10%.

**Table 2**

iTRAQ analysis of insoluble protein Fraction

| Description | Swiss-Prot ID | MOWSE | HCR | LCR | Difference (*p*-value) |
|---|---|---|---|---|---|
| 3-Ketoacyl-CoA thiolase, mitochondrial | P13437 | 70 ± 4 | 1.690 ± 0.665 | 0.903 ± 0.165 | −1.872 (0.01) |
| ADP/ATP translocase 1 | Q05962 | 119 ± 12 | 1.802 ± 0.856 | 0.72 ± 0.253 | −2.503 (0.025) |
| Cytochrome b-c1 complex subunit 1, mitochondrial | Q68FY0 | 79 ± 24 | 1.053 ± 0.088 | 0.553 ± 0.229 | −1.904 (0.003) |
| Cytochrome *c* oxidase subunit 4 isoform 1, mitochondrial | P10888 | 70 ± 4 | 1.258 ± 0.164 | 0.748 ± 0.374 | −1.682 (0.024) |
| Cytochrome *c* oxidase subunit 5A, mitochondrial | P11240 | 69 ± 5 | 1.170 ± 0.198 | 0.702 ± 0.313 | −1.667 (0.023) |
| Tropomyosin α-1 chain | P04692 | 159 ± 49 | 0.540 ± 0.277 | 1.060 ± 0.153 | +1.963 (0.01) |

Protein description, database name and ID relate to the Swiss-Prot database (57.1) entry returned using the MASCOT search engine. A MOWSE score greater than 51 denotes a confident ($p < 0.05$) protein identification. HCR and LCR expression ratios are the average ± SD of peptides quantified in each sample expressed relative to the pooled internal standard. Fold difference relative to HCR values is reported and *p*-values were determined from log-transformed data using Student's independent *t*-tests. The complete list of proteins identified by iTRAQ and the average ± SD number of MS/MS ions analysed and the average ± SD peptides that met the inclusion criteria for iTRAQ analysis (including amino acid sequences identified) are reported in Supporting Information Table 3.

OMICS A Journal of Integrative Biology
Volume 16, Number 9, 2012

DOI: 10.1089/omi.2012.0042

# A Software Toolkit and Interface for Performing Stable Isotope Labeling and Top3 Quantification Using Progenesis LC-MS

Da Qi,[1] Philip Brownridge,[1] Dong Xia,[1,2] Katherine Mackay,[1] Faviel F. Gonzalez-Galarza,[1] Jenna Kenyani,[2] Victoria Harman,[1] Robert J. Beynon,[1] and Andrew R. Jones[1]

## Abstract

Numerous software packages exist to provide support for quantifying peptides and proteins from mass spectrometry (MS) data. However, many support only a subset of experimental methods or instrument types, meaning that laboratories often have to use multiple software packages. The Progenesis LC-MS software package from Nonlinear Dynamics is a software solution for label-free quantitation. However, many laboratories using Progenesis also wish to employ stable isotope-based methods that are not natively supported in Progenesis. We have developed a Java programming interface that can use the output files produced by Progenesis, allowing the basic MS features quantified across replicates to be used in a range of different experimental methods. We have developed post-processing software (the Progenesis Post-Processor) to embed Progenesis in the analysis of stable isotope labeling data and top3 pseudo-absolute quantitation. We have also created export ability to the new data standard, mzQuantML, produced by the Proteomics Standards Initiative to facilitate the development and standardization process. The software is provided to users with a simple graphical user interface for accessing the different features. The underlying programming interface may also be used by Java developers to develop other routines for analyzing data produced by Progenesis.

## Introduction

A VARIETY OF EXPERIMENTAL METHODS have been described for quantitative MS-based proteomics research, in which mass spectrometry (MS) is used to quantify peptides and proteins. These can be categorized into two general principles: label-based methods and label-free methods. Label-based methods commonly involve either differential stable isotope labeling through metabolic labeling *in vivo*, or by chemical modification *in vitro*. In the former approach, intact proteins are labeled *in vivo* through the incorporation of stable isotopes into the media on which cells are growing, for example incorporation of $^{13}C$ into specific amino acids, such as stable isotope labeling of amino acids in cell culture (SILAC; Ong et al., 2002). Chemical labeling techniques include protein-level tagging such as with isotope-coded affinity tags (ICAT) applied to cysteine residues (Gygi et al., 1999), and isobaric tag for relative and absolute quantitation (iTRAQ), which uses a multiplexed set of isobaric reagents that yield amine-derivatized peptides for quantitation (Ross et al., 2004). Label-free quantitation methods that directly use raw spectral data from parallel MS runs to determine relative protein abundance are increasing in popularity. Spectral counting and intensity-based methods are among the most commonly used approaches. Spectral counting relates protein abundance to the number of peptide-spectrum matches (PSMs) in a given run for each protein. Intensity-based methods align precursor ion spectra of the same peptide from parallel runs according to their retention times (RT); protein quantitation is obtained by summation of ion intensities that have been matched to peptides for a given protein.

The Progenesis LC-MS software from Nonlinear Dynamics (Newcastle upon Tyne, U.K.) has been designed to perform label-free quantitation. A variety of other software packages exist, as described in another article in this issue (Gonzalez-Galarza et al., 2012). However, Progenesis is capable of processing a large number of replicates, and has an accessible graphical user interface allowing users to view their MS data in two- or three-dimensional (2D or 3D) maps to verify if features have been quantified accurately; a "feature" is defined as a peptide signal in 2D (retention time versus mass/charge space). The basic steps the software performs are as follows.

1. Map alignment to warp the 2D maps of different replicate runs to remove effects derived from peptides eluting at different times from the liquid chromatography (LC)

[1]Institute of Integrative Biology and [2]Institute of Infection and Global Health, University of Liverpool, Merseyside, U.K.

stage in different runs, ensuring that like-for-like peptide signals are compared;
2. Feature detection, comprising the identification of the isotope pattern of the peptide, including the separation of overlapping peptides, and quantitation using an area-under-the-curve method;
3. Peptide identification using an external search engine; and
4. Peptide and protein quantitation using the abundance signals derived from the features to which identifications have been matched.

Data are output for the user as tables of values for raw features, peptides, and proteins. As such, since quantitation is performed by simple analysis of the signal present in a given feature, the basic feature-level quantitation performed by the software could be used for label-based or other types of experimental workflows in which quantitation is performed at the $MS^1$ level.

At present, it is not possible to use Progenesis LC-MS for techniques other than label-free. We have developed a Java programming interface to enable the feature-level quantitation performed by Progenesis to be used for some other techniques, such as SILAC. The software searches the table of features to find pairs of features nearby in the RT dimension of the 2D map (since heavy:light peptide pairs tend to co-elute), separated by the mass shift in the m/z space due to the heavy label being incorporated into lysine or arginine. For example, typical labels incorporate $^{13}C$ and/or $^{15}N$ into lysine and arginine to produce known mass shifts of +6 Da, +8 Da, +10 Da, and so on. There are no built-in limits on the mass shift that is permissible, for example overlapping features (say +2 Da) are in theory identifiable, so long as Progenesis LC-MS has correctly separated features in the 2D map, although this feature has not yet been comprehensively tested. We have also produced a simple graphical interface for capturing the parameters required for such an analysis (the mass shift on a particular residue), such as the number of missed cleavages, mass/charge tolerance for the peptide pair match, and RT tolerance for the pair matching. Minor updates will be made in future releases to capture triple SILAC and labeling on any amino acid.

A similar method of label-free analysis performed by Progenesis is the top3 protocol, developed by Waters for the SYNAPT range of instruments (Silva et al., 2006). The basic premise of this protocol is to find and sum the signal of the three highest-intensity peptides matched to each protein in a given run. A correlation has been observed between the sum of the three most intense peptide ions and the absolute abundance of a given protein. The summed (or average) intensity over the top3 can be converted into a value of absolute protein abundance by spiking in one or more proteins of known abundance and scaling protein abundance values accordingly. Top3 quantitation is not natively supported by Progenesis, and as such the Progenesis Post-Processor incorporates a module for top3 quantitation (in fact topX, where X is any integer).

Finally, we are contributing to the development of a new data standard from the Proteomics Standards Initiative (PSI), called mzQuantML, which captures the output of quantitation software used in proteomics. It captures values about MS features, peptides, and proteins or groups of proteins, when there is ambiguity in peptide to protein inference (http://code.google.com/p/mzquantml/). This standard is complementary to the recently reported PSI standards for raw or processed MS data, mzML (Martens et al., 2011), and for peptide and protein identification data, mzIdentML (Jones et al., 2012). During the development of a data standard, it is important to develop software for creating genuine example files from software routinely used in laboratories. This toolkit was initially conceived to enable results from Progenesis to be converted into mzQuantML for label-free methodologies, and subsequently extended to cover label-based and top3 quantitation.

We demonstrate that the software is effective for SILAC and top3 label-free quantitation of a dataset created in-house (Supplementary Tables S2 and S3; see online supplementary material at http://www.liebertonline.com). In the first experiment, a SILAC test dataset, created by mixing digests of unlabeled and labeled yeast lysates in known ratios, was prepared and analyzed by ion mobility-coupled data-independent LCMS. The processed data was input into the Progenesis Post-Processor and the reported ratios compared to the actual ratios. We have also used a standard data set created by CPTAC (Paulovich et al., 2009), in which 48 human proteins (SIGMA UPS1 standard) were spiked into an equimolar yeast background in five different concentrations, which can be used for label-free and top3 benchmarking. We have performed various comparisons and benchmarks with these data sets to demonstrate the effectiveness of the software. The software toolkit and data sets are available in the public domain under a permissive license at http://code.google.com/p/progenesis-post-processor.

## Materials and Methods

### *CPTAC study 6*

Five sample conditions were created by splitting a single yeast culture and spiking it with different levels of a standard protein mix: Sigma48 UPS proteins at 0.25, 0.74, 2.2, 6.7, and 20 fmol/$\mu$L (Paulovich et al., 2009). These samples were designated A–E and run in triplicate on various MS platforms. It was reported in the original study that the UPS proteins could not be detected in the A sample (hence only samples B–E are used in our analysis). For this study, the dataset resulting from analysis using a Thermo LTQ Orbitrap mass spectrometer (LTQ-OrbitrapO@65 dataset) was used, as this was seen to have the highest number of peptide-spectrum matches for yeast proteins, and as this is a popular instrument platform within the field. The Thermo raw files were loaded into the Progenesis LC-MS software (version 1.0.0.1), aligned using automatic alignment, and the resulting aggregate spectrum filtered to include +1, +2, and +3 charge states only. The samples were grouped according to experimental conditions (B–E). Features without MS2 spectra were deleted from the analysis, as this information is necessary for peptide/protein identification. An .mgf file representing the aggregate spectrum was exported and searched using Mascot (1 missed cleavage, fixed modification: carbamidomethyl [C]; variable modifications: oxidation [M]; peptide tolerance: ± 10 ppm; MS/MS tolerance: ± 0.6 Da), and the resulting .xml file was re-imported to assign peptides to features, using the following thresholds: Mascot score > 40, hits > 2, to ensure that only high-quality identifications were included. The spectra were searched against the reference strain yeast

database, downloaded from the Saccharomyces Genome Database (orf_trans.fasta, version 03-Feb-2011), concatenated with the Sigma UPS 48 proteins sequences, and a full reverse database (Yeast and UPS sequences reversed) for decoy searching.

### *SILAC dataset*

To create a SILAC test sample, *Saccharomyces cerevisiae* (EUROSCARF accession number Y11335 BY4742; *MATα; his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0; YJL088w::Kan*MX4) was grown in minimal media (Supplementary Table S1; see online supplementary material at http://www.liebertonline.com) using 10 g/L of glucose as the sole carbon source. The minimal media was supplemented with 0.5 mM arginine (unlabeled or $^{13}[C]_6$ arginine), and 1 mM lysine (unlabeled or $^{13}[C]_6$ lysine), to meet the auxotrophic requirements of the strain. Cultures of 200 mL were grown in flasks, and aliquots (15 mL) of the culture were centrifuged (4000 rpm at 4°C for 10 min). The supernatant was discarded and the pellet was immediately frozen at −80°C and stored at that temperature for subsequent protein extraction. Proteins were extracted according to a previously described method (Carroll et al., 2011). An amount of unlabeled or labeled lysate representing 100 μg of protein was dispensed into low-protein-binding microcentrifuge tubes and digested according to the previously described method. A 1:1 mixture of unlabeled and labeled digest was analyzed by LC-MS using a nanoACQUITY UPLC™ system (Waters MS Technologies, Manchester, U.K.), coupled to a Synapt G2 mass spectrometer (Waters MS Technologies). The sample (2 μL corresponding to approximately 1 μg of protein) was loaded onto the trapping column (Waters MS Technologies C18; 180 μm × 20 mm) using partial loop injection for 3 min at a flow rate of 5 μL/min with 0.1% (v/v) TFA. The sample was resolved on an analytical column (nanoACQUITY UPLC HSS T3 C18; 75 μm × 150 mm × 1.7 μm column), using a gradient of 97% A (0.1% formic acid), 3% B (99.9% ACN and 0.1% formic acid), to 60% A 40% B over 120 min at a flow rate of 300 nL/min. The mass spectrometer acquired data using an ion-mobility-coupled data-independent program with 1 sec scan times, and a collision energy ramp of 15–40 eV for elevated energy scans. The mass spectrometer was calibrated before use with the fragment ions of glufibrinopeptide, and throughout the analytical run at 1-min intervals using the NanoLockSpray source with glufibrinopeptide. The data were processed and the database was searched using ProteinLynx Global Server v2.5 (hereafter referred to as PLGS; Waters MS Technologies). The data were processed using a low-energy threshold of 100 (counts), and an elevated energy threshold of 20, and the processed spectra were searched against the complete proteome set of *S. cerevisiae* from Uniprot (reference strain, downloaded June 2010, 6560 proteins). Two searches were applied. The first search was targeted to identify unlabeled peptides using the following parameters: fixed carbamidomethyl modification for cysteine and variable oxidation modification for methionine were specified; one trypsin miscleavage was allowed; and the automatic settings in PLGS for the precursor ion and fragment ion mass tolerance were used. The search thresholds used were: minimum fragment ion matches per peptide 3; minimum fragment ion matches per protein 7; minimum peptides per protein 1; and a false-positive value of 4. The threshold score/expectation value for accepting individual spectra was the default value in the program, such that the false-positive value was 4. The second search targeted labeled peptides using identical search parameters, but with additional fixed modifications of $^{13}C_6$ lysine and $^{13}C_6$ arginine. The average light:heavy ratio was determined from this single analytical run, and then further ratios of 100:1, 50:1, 25:1, 10:1, 5:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:5, 1:10, 1:25, 1:50, and 1:100 were prepared. These samples were analyzed by the same workflow, and technical variability was assessed by three replicate injections. Quantification was performed by Progenesis (version 4, Nonlinear Dynamics, Newcastle upon Tyne, U.K.). Samples were aligned according to RT using a combination of manual and automatic alignment. Maximum peak picking parameters were applied with identifications imported from PLGS, and features with charges from 1+ to 4+ featuring three or more isotope peaks were retained. Identifications were imported from PLGS, and peptides were filtered on a PLGS score of 7. The resulting feature set was exported to the Progenesis Post-Processor, which was used with the following settings: a mass difference of 6 Da (with a 0.5-Da window) for lysine and arginine, one miscleavage allowed, and a 0.8-min retention time window.

## Results

### *Programming and software interface*

The software interface (Fig. 1) is coded in Java and has two major parts: (1) Progenesis file reader classes; and (2) the mzQuantML application programming interface (API), called jmzQuantML (http://code.google.com/p/jmzquantml/). Progenesis output files, which are in .csv (comma delimited) format, report observed values and attributes from detected features, peptides, and proteins. The Progenesis file reader classes provide the functions to read the feature list and protein list files; to retrieve specific values for selected peptide, protein, or other attributes; and calculate summary attributes (e.g., replicate number and identified peptide number), and load these into Java objects that can be accessed by other methods. The mzQuantML API is generated using JAXB, a Java Architecture for XML Binding, which provides automated mapping between elements in the XML document and Java objects. In this context, jmzQuantML is used for loading objects from the Progenesis file reader and writing out valid mzQuantML files (marshalling).

### *SILAC analysis with the Progenesis Post-Processor*

A module has been created in the Progenesis Post-Processor for the analysis of SILAC-generated datasets. The module performs the following steps: (1) imports the quantified features from Progenesis in all replicates; (2) orders all results by retention time of the master run assigned by Progenesis (assuming that all replicates have been correctly aligned); (3) searches for peptide pairs in the two-dimensional space (matches are accepted as correct if the following criteria are met: the light and heavy feature both have at least one identification across all replicates of the same peptide sequence; the heavy feature has the correct mass shift [± the user-entered tolerance], and is within a given tolerance in the retention time axis); (4) the software calculates protein-level ratios, allowing the user control over the method used for

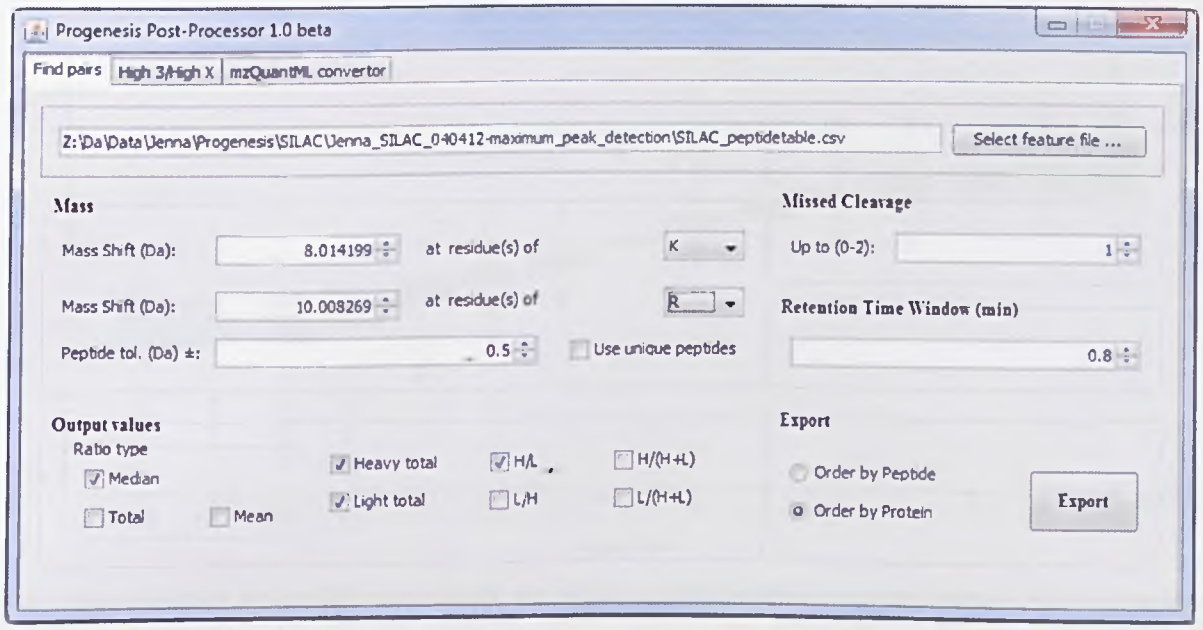



**FIG. 1.** Screenshot of Progenesis Post-Processor used for SILAC analysis.

calculation (discussed below); and (5) exports results for individual peptide pairs and proteins into .csv files for the user to process further. Ratios for proteins are created by one of three methods, offered as choices to the user: median peptide ratio, mean peptide ratio, or ratio of total peptide abundance. Software packages such as MaxQuant favor the median peptide ratio for calculating a protein ratio (Cox and Mann, 2008), although we are not aware of any systematic evaluation. By providing the user some control over this aspect, such comparisons can be made. The user can also select whether to use unique peptides only for quantitation (i.e., those that cannot be matched to other proteins). Users may wish to experiment with this parameter on a case-by-case basis, depending on the similarity between different proteins within their search database. We have experimented with relaxing the criteria for making peptide pair identification, for example not requiring that an identification has been made to both the light and heavy peptide (based on a database search); however, we believe this increases the number of incorrect ratios in the final result, without a great increase in sensitivity (data not shown). The outcome is that users can select a relatively relaxed tolerance for peptide pair matching (say 0.5 Da), without adversely affecting data quality (even if almost all matches are accurate to < 5 ppm). Missed cleavages are currently only handled for lysine and arginine (supporting protocols using trypsin, ArgC, or LysC). If the user selects that n missed cleavages are allowed, the module searches for the peptide pair with mass shifts of (1..n) * mass shift specified, according to the number of lysine or arginine amino acids observed in the peptide sequence (since the peptide sequence is known prior to matching). As such, if the user specifies to allow no missed cleavages (say for a +8 Da on K protocol), the pair matching module will not identify any peptide pairs for which the sequence contains two lysines, since it will not search for the heavy peptide at +16 Da. Such results are therefore automatically excluded from the calculation of protein ratios.

As detailed in the methods section, we have generated an example dataset for testing the performance of the software for which we have control over the expected ratios. The datasets were generated on the Waters SYNAPT HDMSe instrument, for which, to our knowledge, there is currently no software provision for this type of analysis. Samples were generated with expected ratios of 1:100, 1:50, 1:25, 1:10, 1:5, 1:3, 1:2, 1:1, 2:1, 3:1, 5:1, 10:1, 25:1, 50:1, and 100:1, by mixing peptides from parallel grown cultures in either unlabeled or labeled medium. To remove any variation due to the samples being grown in parallel plus incomplete heavy labeling, we have applied post-processing to our results, using the ratios obtained from the 1:1 mix to apply a consistent normalization factor to all other samples. The results from a single run are provided as box plots in Figure 2, showing the protein ratio observed plotted versus the expected protein ratios (Supplementary Figs. S1 and S2; see online supplementary material at http://www.liebertonline.com). Protein ratios were calculated by taking the median peptide ratio, using all peptide pairs matched to the protein as input. Two further replicates are provided as supplementary information, since all show the same general pattern. The results are also summarized in Table 1. In each replicate, it can be observed that broadly accurate ratios are observed for conditions for which there are up to fivefold changes in peptide abundance (i.e., the expected and observed values are close, particularly for ratios 1:2 and 1:3). Higher ratios appear to be flattened, such that a maximum of 1:10 ratios appear observable, even for the spike at 1:100. At this stage it is not simple to determine if this is a facet of this specific benchmarking dataset, or more broadly is due to the use of Progenesis LC-MS in this way, although it should be noted that similar effects have generally been observed for analyses with iTRAQ reagents (Ow et al., 2009). There have been few benchmarking studies on SILAC for which there is a known "ground truth," due to the difficulty of assembling peptides in known abundances in cell culture. We feel there is value at this stage in making the Post-Processor available,

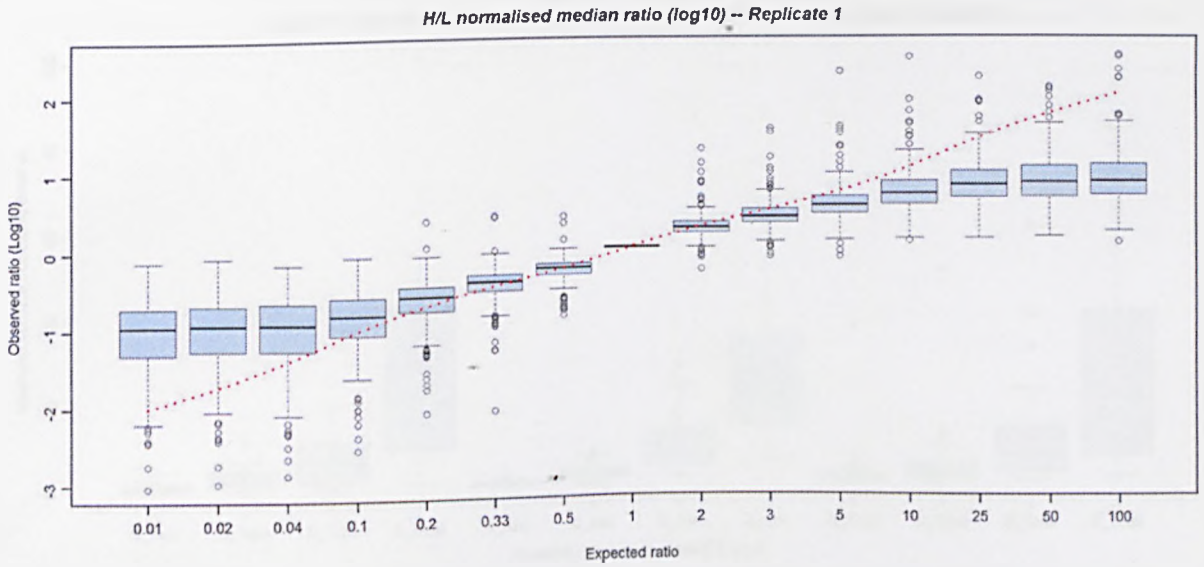



FIG. 2. A box plot of the first replicate of the SILAC data set. The x-axis displays the expected ratios of proteins in each run (calculated from the median peptide ratio, using all peptides matched to a given protein), and the y-axis displays the observed $\log_{10}$ ratios. The dotted line displays the expected values.

to enable other groups to benchmark the performance of Progenesis LC-MS for SILAC, under a wider range of test conditions.

### *Top3 results for label-free benchmarking*

The top3 module was developed to extract the highest three intensity peptides in each replicate and use these to compute a pseudo-absolute value of abundance. At present no additional threshold is applied to exclude peptides from the list, although in future releases we will provide options for excluding modified peptides or peptides resulting from missed cleavages. The software also calculates two other data types: the average of the total abundance, calculated by summing the abundance from all peptides divided by the number of peptides contributing, and the average abundance of the three most abundant peptides (top3). The user can then calculate estimated absolute abundance values via knowledge of proteins that were spiked in known abundance. In the CPTAC dataset, 48 proteins were added in known absolute abundances so all should exhibit the same calculated absolute abundance. For the B–E conditions (0.74, 2.2, 6.7, and 20 fmol/$\mu$L spike), we analyzed the data initially with Progenesis, and then ran the post-processor to extract top3 and average total abundance. We have also compared these data types versus the total abundance calculated directly by Progenesis LC-MS. In each condition, 41 out of the 48 UPS proteins were detected, and we used the median estimate of absolute abundance (across the 41 proteins) to convert top3 or total abundance values into fmol/$\mu$L estimates. The results in the form of a box plot are shown in Figure 3 and a summary table (Table 2). The results demonstrate that all three methods of calculation produce relatively sensible values for absolute abundance (the expected and observed mean values are fairly close), although

TABLE 1. THE EXPECTED RATIOS FOR PROTEINS DETECTED IN THE THREE REPLICATE RUNS, ALONG WITH THE CALCULATED MEDIAN AND MEAN RATIOS, AND THE STANDARD DEVIATION OF THE VALUES OBSERVED

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Replicate 1** | | | | | | | | | | | | | | |
| Expected | 0.01 | 0.02 | 0.04 | 0.1 | 0.2 | 0.33 | 0.5 | 2 | 3 | 5 | 10 | 25 | 50 | 100 |
| Median | 0.106 | 0.111 | 0.113 | 0.151 | 0.249 | 0.378 | 0.558 | 1.747 | 2.371 | 3.291 | 4.580 | 5.991 | 6.462 | 6.685 |
| Std dev | 0.133 | 0.139 | 0.138 | 0.149 | 0.214 | 0.263 | 0.233 | 1.551 | 3.078 | 12.388 | 19.550 | 13.538 | 14.616 | 33.015 |
| Mean | 0.136 | 0.143 | 0.151 | 0.183 | 0.272 | 0.401 | 0.563 | 2.007 | 2.928 | 4.848 | 7.606 | 8.834 | 10.061 | 13.722 |
| **Replicate 2** | | | | | | | | | | | | | | |
| Expected | 0.01 | 0.02 | 0.04 | 0.1 | 0.2 | 0.33 | 0.5 | 2 | 3 | 5 | 10 | 25 | 50 | 100 |
| Median | 0.095 | 0.103 | 0.110 | 0.154 | 0.248 | 0.382 | 0.562 | 1.778 | 2.497 | 3.658 | 5.101 | 6.169 | 6.782 | 6.739 |
| Std dev | 0.124 | 0.110 | 0.139 | 0.140 | 0.176 | 0.168 | 0.101 | 0.436 | 1.049 | 2.477 | 7.616 | 9.238 | 14.255 | 19.823 |
| Mean | 0.123 | 0.129 | 0.140 | 0.180 | 0.269 | 0.393 | 0.566 | 1.834 | 2.698 | 4.127 | 6.509 | 8.653 | 10.108 | 11.503 |
| **Replicate 3** | | | | | | | | | | | | | | |
| Expected | 0.01 | 0.02 | 0.04 | 0.1 | 0.2 | 0.33 | 0.5 | 2 | 3 | 5 | 10 | 25 | 50 | 100 |
| Median | 0.100 | 0.108 | 0.121 | 0.157 | 0.242 | 0.377 | 0.562 | 1.749 | 2.463 | 3.475 | 4.797 | 6.005 | 6.502 | 6.724 |
| Std dev | 0.113 | 0.109 | 0.111 | 0.113 | 0.119 | 0.114 | 0.087 | 0.360 | 0.904 | 1.916 | 8.918 | 9.564 | 11.772 | 17.495 |
| Mean | 0.127 | 0.131 | 0.146 | 0.179 | 0.260 | 0.384 | 0.568 | 1.789 | 2.597 | 3.900 | 6.475 | 8.343 | 9.313 | 10.815 |

A total of 222 proteins were quantified in all three replicates, using two peptides or greater for quantification.

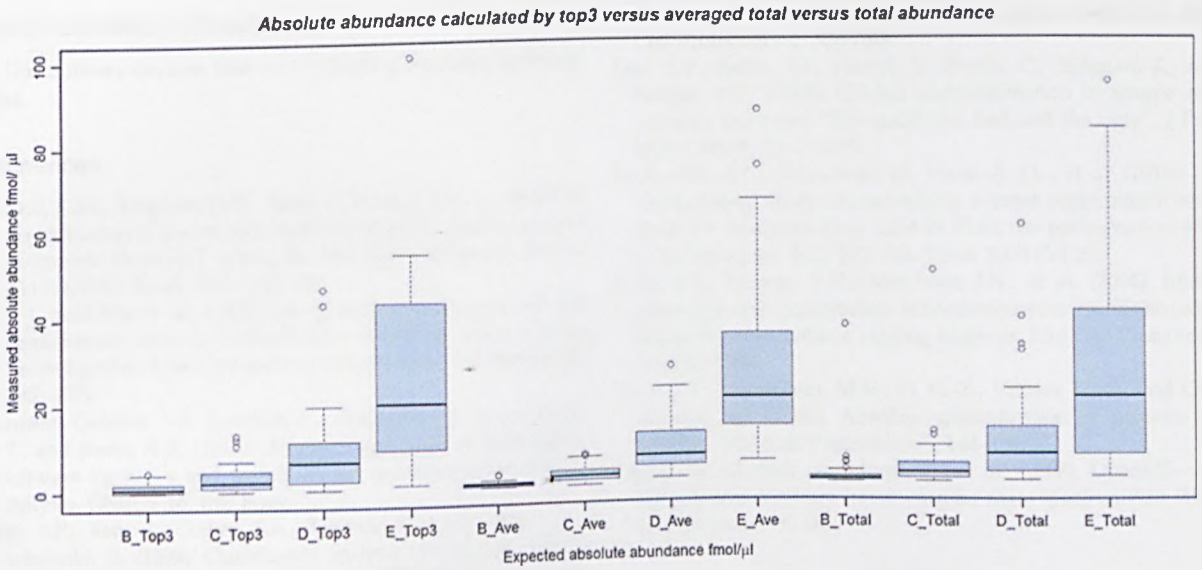



FIG. 3. A box plot displaying the observed values of absolute abundance (fmol/$\mu$L), calculated by the top3 protocol ("_Top3"), average of the total abundance ("_Ave"), and total abundance ("_Total"), in conditions B–E for 41 UPS proteins. Expected values and summary statistics are displayed in Table 2.

the average of the total appears to produce the most reliable estimates of absolute abundance (difference from the mean and standard deviation is lowest in all four comparisons).

It should be noted that the original protocol (Silva et al., 2006) was intended for use on the Waters SYNAPT $MS^E$ instruments. These data were generated on a Thermo Orbitrap, which may explain the difference, but clearly this is an interesting topic for further investigations, which will be facilitated by this tool. We also believe that this tool may be useful for processing datasets originating from Waters Instruments, since Progenesis LC-MS is compatible with Waters RAW files, and PLGS is limited in the number of replicates that can analyzed at once.

### *Export to mzQuantML*

We have built an exporter to write out a single mzQuantML file, capturing peptide and protein values, from the Progenesis output files (accepting feature, peptide, and protein files). As part of the formal standardization process for mzQuantML, several files generated by the software have been used to demonstrate that genuine label-free experiments can be encoded in the standard, thus fulfilling an important role in the standardization process (http://code.google.com/p/mzquantml/). A number of open-source software packages are currently implementing mzQuantML input, such as Proteosuite (http://www.proteosuite.org/), and OpenMS (Sturm et al., 2008), allowing for re-analysis and visualization of datasets generated by other software.

## Conclusions

We have generated a software application for post-processing results files generated by the Progenesis LC-MS software package. It employs two additional protocols beyond the native support of Progenesis, for SILAC and top3 quantitation, on which we have performed preliminary benchmarking to demonstrate that results are sensible and the software is functional. We expect that the software will continue to develop and will offer additional quantitation protocols as they are requested. We have also used the software to generate example files in the developing mzQuantML data standard from the PSI, to facilitate the standardization of this format.

## Acknowledgments

We would also like to gratefully acknowledge current funding sources that have supported this work from BBSRC (BB/I00095X/1 to A.R.J. and BB/G009112/1 to R.J.B.), and EU FP7 grant ProteomeXchange (grant no. 260558).

TABLE 2. SUMMARY RESULTS FOR CPTAC ANALYSIS BY TOP3

| *Protein (n = 41)* | *B_Top3* | *B_Ave* | *B_Total* | *C_Top3* | *C_Ave* | *C_Total* | *D_Top3* | *D_Ave* | *D_Total* | *E_Top3* | *E_Ave* | *E_Total* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Expected (fmol/$\mu$L) | 0.740 | 0.740 | 0.740 | 2.200 | 2.200 | 2.200 | 6.700 | 6.700 | 6.700 | 20.000 | 20.000 | 20.000 |
| Mean (fmol/$\mu$L) | 1.197 | 0.857 | 2.099 | 3.267 | 2.600 | 4.133 | 8.805 | 8.730 | 10.479 | 28.499 | 25.935 | 27.609 |
| Absolute difference from mean | 0.457 | 0.117 | 1.359 | 1.067 | 0.400 | 1.933 | 2.105 | 2.030 | 3.779 | 8.499 | 5.935 | 7.609 |
| Standard deviation | 1.255 | 0.745 | 5.741 | 3.247 | 1.849 | 7.829 | 9.777 | 6.329 | 11.988 | 28.235 | 19.015 | 26.586 |

Average total abundance and total abundance measurement, displaying the expected absolute abundance (fmol/$\mu$L), the mean observed abundance (in 41 out of 48 UPS proteins quantified with $\geq$3 peptides), absolute difference of the observed from the expected value, and the standard deviation.

### Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

### References

Carroll, K.M., Simpson, D.M., Eyers, C.E., et al. (2011). Absolute quantification of a metabolic pathway in yeast: deployment of a complete QconCAT approach. Mol Cell Proteomics 10(12), M111.007633. Epub 2011 Sept 19.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotech 26, 1367–1372.

Gonzalez-Galarza, F.F., Lawless, C., Hubbard, S.J., Hermjakob, H., and Jones, A.R. (2012). A critical appraisal of techniques, software packages and standards for quantitative proteomic analysis. OMICS 16, this issue.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotech 17, 994–999.

Jones, A.R., Eisenacher, M., Mayer, G., et al. (2012). The mzIdentML data standard for mass spectrometry-based proteomics results. Molecular & Cellular Proteomics, first published on Feb. 27, 2012; doi: 10.1074/mcp.M111.014381.

Martens, L., Chambers, M., Sturm, M., et al. (2011). mzML—a community standard for mass spectrometry data. Mol Cell Proteomics 10(1), R110.000133. Epub 2010 Aug 17.

Ong, S.-E., Blagoev, B., Kratchmarova, I., et al. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1, 376–386.

Ow, S.Y., Salim, M., Noirel, J., Evans, C., Rehman, I., and Wright, P.C. (2009). iTRAQ underestimation in simple and complex mixtures: "The good, the bad and the ugly". J Proteome Res 8, 5347–5355.

Paulovich, A.G., Billheimer, D., Ham, A.-J.L., et al. (2010). Interlaboratory study characterizing a yeast performance standard for benchmarking LC-MS Platform performance. Mol Cell Proteomics 9(2), 242–254. Epub 2009 Oct 26.

Ross, P.L., Huang, Y.N., Marchese, J.N., et al. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3, 1154–1169.

Silva, J.C., Gorenstein, M.V., Li, G.-Z., Vissers, J.P.C., and Geromanos, S.J. (2006). Absolute quantification of proteins by LCMSE. Mol Cell Proteomics 5, 144–156.

Sturm, M., Bertsch, A., Gropl, C., et al. (2008). OpenMS—An open-source software framework for mass spectrometry. BMC Bioinformatics 9, 163.

Address correspondence to:
*Da Qi*
*The Biosciences Building*
*Institute of Integrative Biology*
*University of Liverpool*
*Crown Street*
*Liverpool, L69 7ZB, U.K.*

*E-mail:* D.Qi@liv.ac.uk