A NETWORK INFERENCE APPROACH TO UNDERSTANDING MUSCULOSKELETAL DISORDERS

by

NIL TURAN

A thesis submitted to

The University of Birmingham

for the degree of

Doctor of Philosophy

College of Life and Environmental Sciences
School of Biosciences
The University of Birmingham
June 2013

ABSTRACT

Musculoskeletal disorders are among the most important health problem affecting the quality of life and contributing to a high burden on healthcare systems worldwide. Understanding the molecular mechanisms underlying these disorders is crucial for the development of efficient treatments.

In this thesis, musculoskeletal disorders including muscle wasting, bone loss and cartilage deformation have been studied using systems biology approaches. Muscle wasting occurring as a systemic effect in COPD patients has been investigated with an integrative network inference approach. This work has lead to a model describing the relationship between muscle molecular and physiological response to training and systemic inflammatory mediators. This model has shown for the first time that oxygen dependent changes in the expression of epigenetic modifiers and not chronic inflammation may be causally linked to muscle dysfunction.

Bone and cartilage deformation observed in ageing, arthritis and multiple myeloma (MM) patients have also been investigated by using a novel modularization approach developed within this thesis. This methodology allows integration of multi-level dataset with large interaction networks. It aims to identify sub-networks with genes differentially expressed between experimental conditions that are co-regulated across samples in different biological systems. This study has identified several potential key players such as Myc, DUSP6 and components of Notch that could enhance osteogenic differentiation in MM patients.

In conclusion, this thesis present the effectiveness of systems biology approaches in understanding complex diseases and these approaches could be applied for studying other systems and datasets.

ACKNOWLEDGMENTS

I would like to thank people and organisations that have contributed to my scientific journey. First of all, I am grateful to my supervisor Francesco Falciani for giving me the opportunity to work in his group, his efforts to secure funding, for all of his guidance, scientific expertise, and support. I thank to my second supervisor Neil Hotchin for his support and for introducing me to the world of experimental biology. Although I have never held a pipette previously, my experience in the lab has been great fun thanks to him.

Speaking about fun, I also have to mention Shabane's company in the lab. My colleagues from the 5th floor provided laboratory facilities and help. We shared long office hours and we had delightful chats and collaborations with former and present members of our group and other colleagues from IBR and the 8th floor. Collaborating with the members of Biobridge project, Erasys project and the members of the University of Birmingham has been enriching and their input to this thesis has been invaluable; I thank Dieter Maier, Josep Roca, Anna Sotoca, Michael Weber, Joop Von Zoelen, Katrin Sameith, Philipp Antczak, Kim Clarke, Sarah Durant, Anna Stincone, Sarah Essex, Tim Williams and Kevin Chipman.

I am also grateful to University of Birmingham, Biobridge and Erasys projects for providing the necessary funding, which made it possible to carry out my PhD.

Special thanks go to Julie Mazzoloni, Rita Gupta and Harriet Davies for their delightful company, Elsa Boudadi for being my funniest friend, Anna Stincone for the memorable

garlic juice, Philipp Antczak for being my “scientific bro”, Wazeer Varsally for his listening skills during the writing-up time, and Peter for giving me access to his drawer full of snacks.

I thank my dear family: my parents Meral and Faruk, my cousin Cigdem, and my brother’s family for their love and their constant support. My brother has taught me survival skills at a very early age, which turns out to be extremely useful in my grown-up life. My love Marcin has been my inspiration, motivation and my joy. Finally, I would like to thank our Trus for keeping me warm during the writing-up time.

LIST OF PUBLICATIONS

[1] **Nil Turan**, Susana Kalko, Anna Stincone, Kim Clarke, Ayesha Sabah, Katherine Howlett, S. John Curnow, Diego A Rodriguez, Marta Cascante, Laura O'Neill, Stuart Egginton, Josep Roca, and Francesco Falciani. A systems biology approach identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease. *PLoS Comput Biol*, 7(9):e1002129, Sep 2011.

[2] Dieter Maier, Wenzel Kalus, Martin Wolff, Susana G Kalko, Josep Roca, Igor Marin de Mas, **Nil Turan**, Marta Cascante, Francesco Falciani, Miguel Hernandez, Jordi Villà-Freixa, and Sascha Losko. Knowledge management for systems biology a general and visually driven framework applied to translational medicine.*BMC Syst Biol*, 5:38, 2011.

[3] Tim D Williams, **Nil Turan**, Amer M Diab, Huifeng Wu, Carolynn Mackenzie, Katie L Bartie, Olga Hrydziuszko, Brett P Lyons, Grant D Stentiford, John M Herbert, Joseph K Abraham, Ioanna Katsiadaki, Michael J Leaver, John B Taggart, Stephen G George, Mark R Viant, Kevin J Chipman, and Francesco Falciani. Towards a system level understanding of non-model organisms sampled from the environment: a network biology approach. *PLoS Comput Biol*, 7(8):e1002126, Aug 2011.

[4] Sabrina Moro, J. Kevin Chipman, Philipp Antczak, **Nil Turan**, Wolfgang Dekant, Francesco Falciani, and Angela Mally. Identification and pathway mapping of furan target proteins reveal mitochondrial energy production and redox regulation as critical targets of furan toxicity. *Toxicol Sci*, 126(2):336–352, Apr 2012.

CONTENTS

TABLE OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION AND BACKGROUND

1.1 Human musculoskeletal system

The functional role of the musculoskeletal system is to allow support, protection and movement of the body [1]. All these are achieved via a complex organisation of muscles, bones, joints, tendons, cartilage, and ligaments [2]. With the contraction of a muscle, the force generated is transmitted via tendons to bones allowing them to move (**Figure 1.1**) [3]. Bones are connected to each other via ligaments and protected from rubbing each other via cartilage that functions as cushion in the joints [1]. They are also a source of mineral homeostasis such as calcium, magnesium and phosphorus. Bone marrow, which fills the internal cavity of the bones, is responsible for the production of blood cells.

In the next sections, the components of the musculoskeletal system that are affected during the ageing process and in various diseases including COPD, multiple myeloma and arthritis will be reviewed in detail. In section 1.1.1, skeletal muscle structure is described and known molecular mechanisms for muscle homeostasis are reviewed. In section 1.1.2 bone and cartilage formation and maintenance are described.

1.1.1 Skeletal muscle

Skeletal muscles are composed of numerous bundles of fibres (**Figure 1.1**), which are responsible for force and movement production. Muscle fibres are multinucleated cells

formed via the fusion of myotubes. Muscle fibres are surrounded by basal lamina. Beneath the muscle fibre basal lamina, there are satellite cells that play a crucial role for muscle repair as these have the ability to differentiate into myoblasts to form new muscle fibres [4, 5], [6, 2].

Proportions of muscle fibre types determine the physiological, metabolic and functional properties of the muscle [6, 2]. Individual fibre phenotypes are well conserved among mammals. However, the composition of a muscle can differ between mammals [7]. For instance, soleus muscle consists of 100% of slow muscle fibre types in guinea pig [7], whereas human soleus fibre composition can contain between 60% and 100% slow fibres [8]. Depending on the fibre type composition, muscles have different contractile speed, strength and resistance to fatigue. However, the muscle properties can change to adapt to various conditions such as disuse and exercise training.

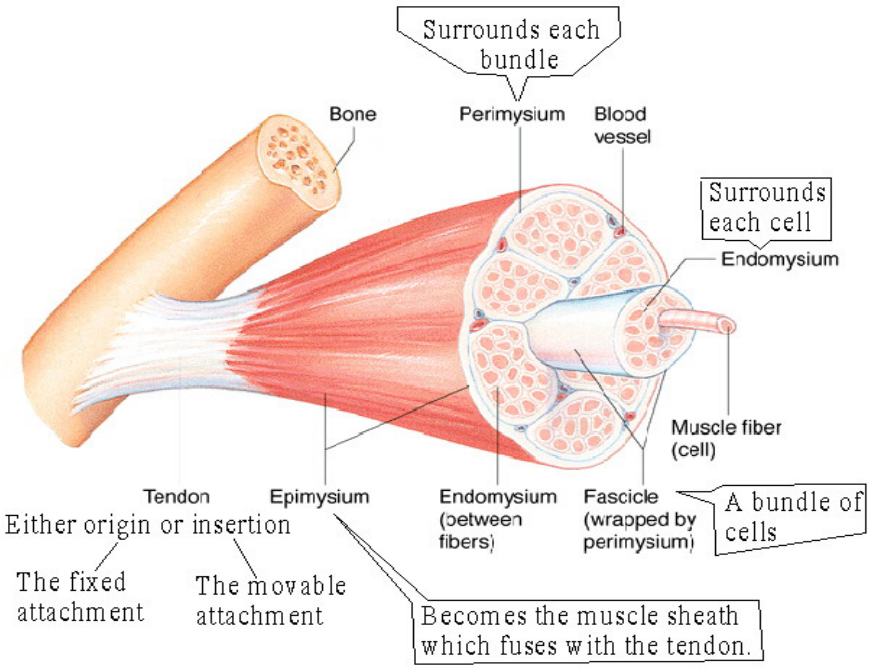


Figure 1-1. Structure of skeletal muscle published in [4]

This figure shows the structure of the skeletal muscle, which is connected via tendons to bones. Skeletal muscles are composed of numerous bundles of fibres, called fascicle. The diameter and length of the fibres determine the strength of the muscle and contraction

velocity. The basal lamina surrounds the fibres and it plays a crucial role in fibre recovery from injuries.

Muscle fibre types and properties

Classification of muscle fibres is achieved using various techniques including histochemical, biochemical, morphological and physiological characteristics [6]. Histochemical ATPase assay classification can differentiate between slow (Type I) and fast fibres (Type IIa and IIb). The characteristics of the fibres are summarized in Table 1.1. Type I fibres, also called slow oxidative, are characterized by high density of mitochondria, which are responsible for the generation of energy aerobically (oxidative phosphorylation). As oxidative phosphorylation is very effective for ATP generation, type I muscles are resistant to fatigue. Hence, the muscle fibres' endurance property is dependent on the mitochondria density; for instance in highly oxidative muscle 20% of the total cell volume consists of mitochondria [2]. Type IIa fibres (fast oxidative) are rich in michochondria and generate ATP mainly through oxidative phosphorylation. This type of muscle has faster contractile speed than type I and is less resistant to fatigue. Both of these types require adequate supply of oxygen for energy generation, these fibres are rich with capillaries and with high myoglobin content. On the other hand, Type IIb fibres (fast glycolytic) are low in mitochondrial density; therefore energy generation is mainly achieved in the cytoplasm in anaerobic conditions. The low oxidative capacity of type IIb fibres leads to increased susceptibility to fatigue. These fibres have very fast contractile speed as the tension activation system occupy a large volume, almost two to three times more than slow fibres [2].

	Type I (slow oxidative)	**Type IIa (fast oxidative)**	**Type IIb (fast glycolytic)**
Contractile speed	Slow	Intermediate	Fast

Mitochondrial density	High	High	Low
Capillaries	Many	Many	Few
Enzymes for anaerobic glycolysis	Low	Intermediate	High
Resistance to fatigue	Resistant to fatigue	Intermediate resistance to fatigue	Fatigue easily

Table 1.1. Properties of muscle fibre type I and II, selected from the book of Richard L.Lieber [2].

Skeletal muscle tissue remodeling

The maintenance of muscle homeostasis and adaptation to various conditions occurs via a remodeling process. Several studies have been carried out for the identification of molecular mechanisms underlying the signaling pathways regulating muscle remodeling, for a recent review see [9]. The key players include growth factors, cytokines, receptors, extracellular proteins that control protein synthesis and degradation via various mechanisms. For instance, it has been reported that insulin-like growth factor 1 (IGF1) activation induces muscle hypertrophy *in vitro* and *in vivo* through the activation of the mammalian target of rapamycin (Akt/mTOR) pathway [10, 11]. The role of pro-inflammatory cytokines such as tumor necrosis factor alpha (TNFa) and interleukin 1 (IL1) in skeletal muscle remodeling is more contradictory in the literature. It has been reported that high level of TNFa regulates myogenesis and muscle regeneration [12], but it has also been reported to induce muscle atrophy [13]. Over-expression of IL1B has been reported to reduce protein synthesis, muscle weight and protein content in rats gastrocnemius and soleus musles [14] and impair myogenesis [15, 16]. On the other hand, in response to exercise training IL1 activity is increased in skeletal muscle [17]. The opposite role of TNFa and IL1 in skeletal muscle is thought to be a result of duration and magnitude of the inflammatory signals [18].

Regulation of skeletal muscle metabolism

A number of regulators of skeletal muscle energy metabolism that is crucial for the maintenance and growth of skeletal muscle have been reported. One of the most studied metabolic regulators is peroxisome proliferator-activated receptor c coactivator 1 alpha (PGC-1a), which induces mitochondrial biogenesis and oxidative metabolism and controls muscle growth [19]. PGC-1a over-expression in skeletal muscle of mice increases type I fibres, muscle performance and resistance to muscle fatigue [20]. Regulation of PGC-1a and its targets has been reported to be mediated by chromatin modifier the silent information regulator T1 (SIRT1) in myoblasts [21]. Several other chromatin modifiers such as the silent information regulator T3 (SIRT3), histone deacetylase 4 (HDAC4) and histone deacetylase 5 (HDAC5) have also been shown to be important regulators of skeletal muscle bioenergetics expression [22]. For instance, in human myotubes, HDAC5 was shown to regulate GLUT4, a glucose transporter [23].

1.1.2 Bone and cartilage

Bone remodeling is controlled by bone forming cells osteoblasts and bone-resorbing osteoclasts [24]. The tight balance between osteoclasts and osteoblasts allows the degradation of old bone and synthesis of new bone. The osteoblasts are derived from human mesenchymal stem cells (hMSCs) differentiation.

Defining hMSCs

hMSCs are adherent stromal cells, which were first identified in the bone marrow [25]. Later, they were also extracted from other tissues such as fat [26], skin [27], blood [28], umbilical cord blood [29], teeth [30], and pancreas [31]. These cells are self-renewal and have the ability to differentiate into various lineages including bone, cartilage and fat cells (**Figure 1.2**) [32]. It should be noted that the usage of the definition MSC has been criticised lately

and three criteria have been proposed to define MSC by the Mesenchymal and Tissue Stem Cell Committee [33]. These criteria are derived from the data currently available. The first one is their adherence property to plastic when maintained in standard culture conditions. They should express specific surface antigen CD105, CD73 and CD90 and lack the expression of CD45, CD34, CD14, CD11b or CD19 and HLA-DR. The last criterion is based on their differentiation capacity *in vitro* towards the three lineages. In this thesis, hMSCs was used for referring to the stromal cells that are derived from bone marrow. It should be noted that with this current definition, it is not possible to distinguish MSCs from other stromal cells such as fibroblasts [34]. Although some efforts have been made to distinguish differences between those cells, more studies are required for redefining the complex family of stromal cells [34].

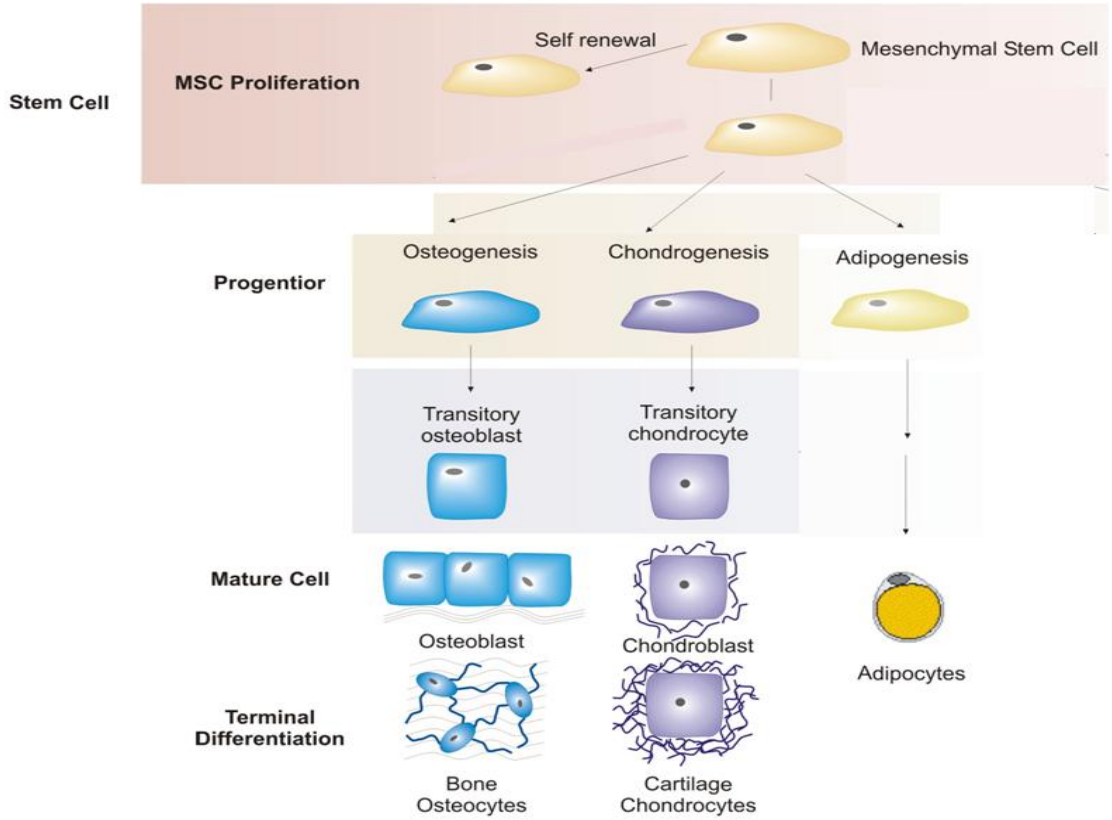


Figure 1-2. hMSC differentiation towards bone, fat, and cartilage cells.
This figure is a selection from [35], which was adapted from [36].

Known molecular mechanisms during the differentiation of hMSCs towards bone forming cells

Understanding molecular mechanisms underlying normal osteogenic and chondrogenic differentiation from hMSCs is crucial for the identification of potential regulators that are abnormal in conditions leading to bone loss and cartilage deformation.

hMSCs differentiation towards osteoblasts can be triggered *in vitro* with dexamethasone treatment, which is a glucocorticoid class of steroid drugs. Moreover, addition of other factors such as bone morphogenetic protein 2 (BMP2) and cholecalciferol (Vitamin D_3) to dexamethasone treatment has been shown to enhance the differentiation capacity of hMSCs towards bone as compared to only dexamethasone treatment [37, 38]. A number of factors including cytokines, growth factors, and transcription factors have been shown to be involved during osteogenesis differentiation. Wnt signaling, Notch signaling, and BMP signaling pathways play a role in activating runt-related transcription factor 2 (RUNX2), which has been shown to regulate early stages of osteogenesis and RUNX2 null mice exhibits endochondral ossification (for recent reviews [39, 40]). However, the role of Wnt signaling pathway in osteogenesis is conflicting in the literature. Several components of Wnt activation have been shown to enhance the osteogenesis differentiation *in vitro* [41, 42] and mice lacking the Wnt antagonist sFRP1 display a phenotype with increased trabecular bone mineral density and volume [43]. On the other hand, it has been shown that Wnt3 activation suppresses dexamethasone-induced osteogenesis in hMSCs [44]. Other osteogenic enhancer factors such as dual specificity phosphatase 6 (DUSP6), guanine nucleotide-binding protein (GNAS) and purinergic receptor P2Y, G-protein coupled, 11 (P2RY11) were identified in a recent study by performing a siRNA library screening [45].

Description of key players during chondrocytes differentiation from hMSCs

The differentiation of hMSCs towards chondrocytes can be achieved via simultaneous treatment with dexamethasone and TGFB. Furthermore, addition of BMP6 or IGF1 to this treatment has been shown to enhance chondrocytes differentiation [46, 47]. The previous studies from hMSCs differentiation towards chondrocytes allowed the identification of several key players of this process. These include the transcription factor SOX9 [48], growth factors such as fibroblast growth factor 2 (FGF2) [49], BMPs [50], IGF1 [47], and transforming growth factor (TGF) [51], components of SMAD signaling [52], Notch signaling [53] and the Wnt signaling pathway [54]. Activation of SOX9 has been shown to be essential for chondrocytes differentiation and cartilage formation [48]. Inducing chondrogenic differentiation from hMSCs leads to activation of SOX9 that in turn activates the chondrocyte markers collagen type II alpha 1 (COL2A1), cartilage oligomeric matrix protein (COMP) and Aggrecan (ACAN) [48, 55]. For the activation of SOX9, suppression of Notch signaling pathway is required in TGFB induced chondrogenesis from hMSCs [56]. The importance of notch inhibition has also been reported in mice, which lead to induced chondrocyte activity [57].

1.1.3 Diseases affecting the musculoskeletal system

In this thesis, a number of diseases affecting various components of the musculoskeletal system including muscle, bone and cartilage have been studied. In the context of muscle wasting, muscle transcriptomics and physiological differences between COPD patients and aged-matched healthy individuals before and after exercise training have been studied. In the next section (section 1.2), the background and state of the art in muscle wasting in COPD patients are described. Then, the second biological system related to bone loss and cartilage deformation observed in ageing, in arthritis, and in MM will be introduced (section 1.3). The abnormal bone marrow microenvironment in these patients might be partially responsible for the aberrant hMSCs differentiation towards fat, bone and cartilage cells. In this thesis, the

aim has been to study the osteogenesis, adipogenesis and chondrogenesis differentiation process in ageing, in arthritis and in MM patients, as the defective differentiation process might result in bone and cartilage deformation.

1.2 Muscle wasting in Chronic Obstructive Pulmonary disease (COPD) patients

COPD is a progressive disease leading to accelerated decline in lung function. COPD patients suffer from limited airflow, which is induced by closure of small airways and loss of lung elasticity. The disease is the result of the interaction between genetic predisposition factors, smoking and exposure to fine particulates. A significant percentage of COPD patients suffer from severe extra-pulmonary effects such as weight loss, muscle wasting, osteoporosis, and cardiovascular diseases [58]. These contribute to poor quality of life and mortality in COPD patients. The World Health Organization has predicted COPD to be the third cause of death by 2030 [59].

1.2.1 Muscle wasting in COPD

COPD patients suffer from shortness of breath caused by impaired lung function that leads to a decrease in exercise tolerance and physical activity. Skeletal muscle is a very plastic tissue with the ability to adapt in response to various levels of activity and disuse [2]. The loss of physical activity of COPD patients contributes to progressive muscle wasting, which is estimated to be observed in 40% of the patients [60]. Muscle wasting determines the course of the COPD [61] and predicts survival [62]. As the exercise training in COPD patients has limited recovery on skeletal muscle mass and strength compare to healthy individuals, other factors are likely to contribute to muscle wasting. Indeed, chronic hypoxia, inflammation, ROS imbalance has been shown to be key factors driving skeletal muscle dysfunction in COPD patients and these would be discussed in great details in the subsequent sections.

1.2.2 Muscle characteristics of COPD patients

COPD muscle is characterized by muscle mass loss, reduced muscle strength, fibre type shift, reduced oxidative metabolic capacity and loss of capillary density. Loss of muscle bulk occurs early in COPD patients even in the absence of severe airflow limitation [63]. Reduced thigh muscle cross-sectional area has been observed in COPD patients compared to normal individuals [64]. Muscle biopsy samples taken from vastus lateralis muscles of COPD patients at rest have shown atrophy of type I fibres and an increase in type IIb fibres compared with healthy controls [65, 66, 67]. This shift from fibre type I to fibre type IIb has a major consequence in increasing susceptibility to fatigue and contributing to exercise limitation in COPD, as type IIb fibres are characterised by fast contraction velocity and low oxidative capacity (Table 1.1).

1.2.3 Skeletal muscle energy metabolism in COPD

In accordance with the shift of fibre type composition observed in COPD patients, skeletal muscle energy metabolism is altered towards glycolytic metabolism and reduced oxidative capacity. Reduced concentrations of mitochondrial oxidative enzyme in COPD skeletal muscle was reported compared to aged-matched healthy muscle [68]. The loss of mitochondrial density in COPD skeletal muscle has been recently reported, which contributes to the impaired oxidative capacity [69]. The impaired oxidative capacity and increase in glycolytic metabolism in COPD can cause reduced ATP synthesis due to the inefficient mechanism of ATP generation in anaerobic conditions. In support, several studies have reported a reduction in resting ATP concentration, in the quadriceps muscle of COPD patients [70, 71]. It has been also shown that required ATP demand during low exercise workload is not generated in COPD patients [71].

1.2.4 Potential players in skeletal muscle dysfunction and wasting in COPD patients

The cause of the skeletal muscle disorder observed in COPD may be due to a number of factors including an imbalance between protein degradation and synthesis, systemic and local inflammation, oxidative stress and hypoxia. In the next paragraphs, these factors and the possible molecular mechanisms underlying the muscle degeneration in COPD patients will be elaborated.

1.2.4.1 The role of systemic inflammation in muscle wasting

High level inflammation in the lung of COPD patients leads to systemic inflammation, refered also as 'overspill' in the community [72]. COPD is associated with evidence of increased plasma levels of proinflammatory cytokines. A systematic review and meta-analysis supported this and concluded that individuals with COPD had significantly raised levels of several markers of inflammation including C-reactive protein (CRP), IL-6, fibrinogen, and TNFa, confirming the presence of systemic inflammation [73]. Until recently, the systemic inflammation was thought to be a major driver of muscle wasting in COPD patients by inducing local inflammation and causing higher protein degradation.

1.2.4.2 The role of local inflammation in muscle wasting

The earlier common view was that local inflammation caused via systemic inflammation in COPD patients contributes to decline in muscle function. Although this hypothesis is supported by animal models linking systemic inflammation with local inflammation and muscle atrophy [74], it is still unclear whether this mechanism may play a role in COPD. For example, the analysis of cytokine profiles in the quadriceps of patients with COPD shows that the level of TNFα in COPD individuals is lower than in control subjects [67]. Moreover, in this case TNFα protein levels correlated positively with muscle function, implying higher levels were associated with increased quadriceps strength. Other pro-inflammatory cytokines

(e.g. TNFα receptors) as well as the pro-angiogenic factor VEGF showed similar trends [67]. These observations are consistent with the role of local inflammatory signals, induced by muscle contraction and required for the activation of tissue remodeling pathways.

1.2.4.3 Molecular mechanisms underlying the imbalance between muscle protein synthesis and degradation

As the tight balance between the protein synthesis and degradation controls the muscle mass, higher protein degradation and/or reduced protein synthesis is responsible for muscle loss. However, molecular mechanisms responsible for an imbalance between protein synthesis and degradation in COPD are still unclear. There are several pathways that are hypothesized to induce muscle protein degradation or inhibit protein synthesis (**Figure 1.3**). These include inhibition of AKT/MT pathway via catabolic signals and activation of NF-KB pathway MAPK signaling pathway via inflammatory signals (**Figure 1.3**). Activation of AKT/MTOR pathway has been shown to induce muscle hypertrophy *in vivo* [11]. Inhibition of Akt has been shown to dephosphorylate FOXOs causing localization of the FOXOs in the nucleus, which then induces activity of muscle specific target genes ubiquitin ligases MURF1 and MAFBX leading to protein degradation [75, 11, 76]. Similarly activation of MAPK pathway through phosphorylation of p38MAPK activates MURF1 and MAFBX in the muscle atrophy in cell culture experiments in response to cytokine stimulation [77] and in animal models of immobilization induced atrophy [78]. The activation of the transcription factor NF-KB leads to inhibition of MyoD expression, which is essential for skeletal muscle differentiation and the repair of the damaged tissue [79]. However, the molecular mechanisms derived from the animal studies cannot be verified in COPD patients. Recent studies have reported that activation of MAPK and NF-KB pathways in COPD patients are lacking. In a large clinical study, consisting of 105 COPD patients, samples from quadriceps muscle showed no differences in p38 MAPK protein or p38 MAPK mRNA levels compared with healthy

muscle [80]. Similarly lack of activation of NF-KB pathway was shown by Mercken et al. [81]. On the other hand, the consequence of Akt/mTOR pathway inhibition in animal models appears to be consistent with COPD muscle [82]. In a recent study, Akt/mTOR pathway has been shown to be inhibited in hyoxemic COPD patient compared with nonhypoxemic COPD patients, but the study consists of only 8 patients in total [82]. Therefore larger studies are needed to elucidate this pathway.

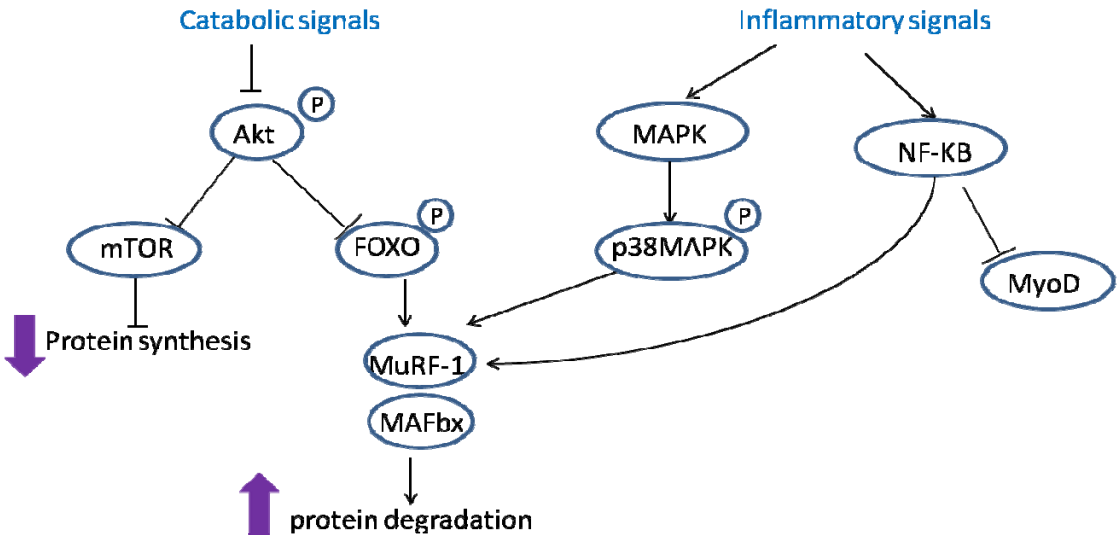


Figure 1-3. Schema of the molecular mechanisms leading to increased activity of protein degradation and decreased protein synthesis.

This figure represents the role of AKT signaling pathway, MAPK signaling pathway and NFKB in skeletal muscle atrophy. This figure has been summarized and adapted from [83, 75, 11, 76, 79]

1.2.4.4 Is muscle hypoxia a factor determining muscle wasting in COPD?

As COPD patients have difficulties breathing and suffer from impaired oxygen uptake, they have a deficient oxygenation of the blood, which results in muscle tissue hypoxia [60]. There are several pieces of evidence supporting the view that this may be an important trigger for muscle dysfunction in COPD. Chronic hypoxia has been proposed to explain skeletal muscle loss in COPD patients as well as in healthy individuals at high altitudes [84, 85, 86]. It has

been reported that individuals at high altitudes suffer from reduced muscle mass and peak power [84, 85, 86]. A recent small study based on four hypoxemic and four nonhypoxemic COPD patients has reported that hypoxia is responsible for inhibition of Akt/mTOR signaling pathway [82], which was previously shown in animal models to be a crucial regulator of skeletal muscle hypertrophy [11]. In this study, they have also shown that inhibition of Akt/mTOR pathway is due to the activation of REDD1 (Regulated in Development and DNA Damage Responses) [82]. This was also validated in an animal study exposed to severe hypoxia for 3 weeks and REDD1 overexpression *in vivo* caused skeletal muscle fibre atrophy [82].

1.2.4.5 Oxidative stress in COPD muscle

Oxidative stress has also been shown to contribute to muscle structural and functional damage [87]. It occurs as a result of an imbalance between Reactive Oxygen Species (ROS) and antioxidant systems. Inflammation, hypoxia and mitochondrial dysfunction have been linked with increased muscle ROS levels [88, 89]. In line with this, oxidative stress in muscle measured by protein carbonylation has been shown to be higher in COPD patients as compared to healthy individuals and negatively correlated with quadriceps muscle strength, which was measured by muscle voluntary contraction [67, 90]. Koehlin and collegues have reported that COPD patients with chronic hypoxemia have increased muscle oxidative stress at rest and after exercise as compared to non-hypoxemic COPD patients [91].

1.2.5 Current therapies for muscle wasting in COPD

Current treatments are still inefficient for increasing and preserving the muscle mass [92]. The known treatments are exercise training, oxygen therapy, nutritional interventions, steroid supplements and anti-inflammatory drugs [92]. The oxygen therapy has been shown to be beneficial in short term by enhancing exercise capacity [93]. Although exercise training represents one of the most effective treatments, the selection of the training regime is very

important as it has been shown that high intensity exercise training is causing oxidative stress [94].

Steroid supplementation combined with exercise training has been shown to be beneficial in COPD patients. A recent study has reported that COPD patients with low serum testosterone levels who were subjected to resistance training and testosterone supplementation for 10 weeks increased leg muscle strength as compared with patients subjected to resistance training and placebo [95]. However, dose optimization, establishing the safety and the benefits of long term anabolic steroids treatments are required [95].

Nutritional intervention did not lead to improvements to muscle function in COPD patients in several studies [96, 97, 98]. For example, creatine supplements have not yield to significant improvements on muscle strength and fatigue in COPD patients compared with exercise training [97]. Similarly, a recent study has reported no effect of glutamine ingestion of oxidative metabolism in stable COPD patients [98].

As the current treatments are still inefficient for restoring the muscle mass and strength in the COPD patients, it is crucial to understand molecular mechanisms underlying muscle wasting.

1.3 Bone and cartilage degradation in ageing, myeloma and arthritis patients

Having described various aspects of the muscle wasting in COPD patients, in this section the second biological system related to bone loss and cartilage deformation observed in ageing, in arthritis, and in MM will be introduced. Although this study represents a very heterogeneous scenario, all these conditions are associated with bone and cartilage deformation. In this thesis, the aim is to understand possible abnormalities that are deriving such deformation. First a background on the diseases will be provided. Then the role of abnormal microenvironment contributing to the abnormal bone and cartilage differentiation in ageing and in these conditions will be described.

1.3.1 Background on diseases associated with bone and cartilage degradation

Multiple Myeloma (MM) is a cancer that is initiated by proliferation of malignant plasma B-cells in the bone marrow. This disease has a median survival rate of 62 months for stage 1, 44 months for stage 2, and 29 months for stage 3 patients [99]. The categorization of patients in various stages is based on the degree of bone lesions, anaemia, hypercalcemia, levels of M protein in the serum and urine [100]. During the course of the disease, most of the patients develop bone fractures and bone pain. Severity of bone lesions is a predictor of survival duration [101].

Arthritis patients suffer from bone and cartilage deformation in the joints causing pain and mobility restriction. There are numerous types of arthritis but the most frequently observed type is osteoarthritis [102]. **Osteoarthritis** (OA) is progressive degenerative disease affecting the articular cartilage such as knees, hips, fingers and spine, causing bone deformation and mild synovial inflammation [102]. It is estimated to affect 9.6% of men and 18.0% of women aged over 60 years worldwide [102]. A less frequently observed type of arthritis is **rheumatoid arthritis** (RA), which is a systemic inflammatory disease affecting synovium in joints, leading to joint and bone destruction [102].

1.3.2 Bone and cartilage degradation as a result of abnormal hMSCs differentiation in ageing and in diseases

Changes in the bone marrow microenvironment that occurs as a result of ageing and disease can potentially influence hMSCs differentiation capacity to bone, cartilage and fat cells.

In ageing, it has been shown that there is an increase in the number of adipocytes, which results in a fatty bone marrow microenvironment [103]. The changes in the bone marrow microenvironment are favouring adipogenic and osteoclastic differentiation over osteogenic

differentiation [104], [105]. The increase activity of osteoclastogenesis and decreased osteoblast activity result in sustained bone loss in ageing.

Similarly, in MM patients a decreased osteoblast activity [106, 107] and an increased osteoclast activity [108] have been shown. Moreover, malignant plasma cells from MM patients cultured with human osteoblastic cells enhances osteoblast apoptosis [109]. Corre et al has reported that the osteogenesis differentiation from the MM patients is two-fold less effective than normal patients [106]. A novel study has confirmed the impaired osteogenic differentiation of hMSCs from Myeloma patients compare to normal patients [107]. They have also shown that the osteogenesis process is less effective in patients with bone lesions compared with the patients without bone lesions [107].

In arthritis patients, the inflammatory bone marrow microenvironment leads to joint and bone destruction that can occur via interfering with osteoclast, osteoblast and chondrocyte activities and differentiation of hMSCs towards these lineages. Several in-vitro [110, 111, 112] and animal models [112, 113] have shown that osteoclast activity is induced via cytokine activity, which is supporting that the abnormal bone remodeling in RA patients can be due to enhanced osteoclast activity. Indeed, a recent study has shown an increased osteoclast functional activity in RA patients [114]. The number of bone forming cells has also been shown to be decreased in inflammatory conditions. Osteoblasts collected from bone biopsies of RA patients treated with IL-1B and TNFa has been shown to induce apoptosis of osteoblasts via induction of FAS [115].

Only a small number of studies have investigated the hMSCs and their efficiency of differentiation from RA and OA patients. A particularly interesting result shows that *in vitro* chondrogenic and adipogenic differentiation of bone marrow derived hMSCs from OA patients is significantly reduced compared with healthy individuals, whereas osteogenic

differentiation was comparable [116]. Osteogenic differentiation of hMSCs from OA and RA patients has also been shown to be similar [117]. In another study, *in vitro* differentiation of hMSCs from RA patients towards the three lineages have been shown to be as efficient as from healthy donors [118]. Further studies are required to confirm the properties of hMSCs, differentiation efficiency in RA, OA and healthy individuals. Microarrays capturing the mRNA response during differentiation would be also very valuable for understanding key players in this process.

1.3.3 Potential role of hMSCs for treatments

Development of clinical applications for hMSCs is crucial for the treatment of degenerative diseases and many preclinical studies were reported, for recent reviews [119, 120, 121]. The potential clinical applications include autologous or allogeneic local implantation and systemic infusion of hMSCs, which might be also combined with gene therapy. Two clinical trials have been reported to be successful for bone regeneration. An autologous hMSCs treatment in three patients with defective fracture healing [122] and allogenic hMSCs treatment in three patients with osteogenesis imperfecta [123] has lead to bone formation. Autologous hMSCs treatment has also been demonstrated to contribute to cartilage repair in OA patients [124].

Although the reported encouraging outcomes of the clinical trials in the literature, larger randomized studies are required before hMSCs treatments can be routinely available. Better understanding of normal osteogenesis and chondrogenesis process and identification of potential players in disease states is crucial for enhanced bone formation combining hMSCs treatments with gene therapy.

1.4 Functional genomics and systems biology approaches

In this thesis, muscle wasting in COPD patients, deformation of bones and cartilage in ageing, arthritis patients and multiple myeloma patients was studied using transcriptomics and systems biology approaches. In the following sub-sections, the state of the art in systems biology will be provided.

1.4.1 Omics technology and challenges

Technology advances in the last 10-15 years have made possible measuring the expression of genes, proteins and metabolites at genome level in single experiments. This has contributed to generate an unprecedented amount of data. Integrating, analyzing and interpreting these data using advanced computational methodologies is one of the objectives of the emerging discipline of systems biology. Noisy data, technical variations between samples due to generation of data from different people or laboratories, patient heterogeneity, small number of samples and large number of variables to study makes this task very challenging. Furthermore, biological interpretation of this data is also difficult due to the large amount of data. In the section below, an overview of the computational approaches are described for the analysis of microarray based expression profiling, which is one of the most widespread functional genomics technology. Moreover, some of the methodologies described below allow also integration of multi-level datasets with classical physiology.

1.4.2 Microarrays and various types of technologies

DNA microarrays allow measuring expression levels of thousand of genes simultaneously. Microarrays are made of thousands of DNA probes immobilized into specific locations (spots) on a solid small support such as glass or silicon chips. There are two major types of platform to generate microarrays. In one platform, cDNA or long oligonucleotides (70-80mers probes) have been spotted on a glass slide. The second technique has been pioneered

by Affymetrix technology and it is produced by synthesizing short oligonucleotides (20-25mer probes) on silicon chips using photolithography. Another major provider is Agilent, which uses the first type of platform by printing long oligonucletodies on glass slides using an ink-jet based technology. In this thesis, microarrays generated with Affymetrix and Agilent technologies have been studied. Although the protocols of the two techniques differ, the principle behind microarray generation can be summarized in three major steps including sample preparation and labelling, array hybridization and scanning. Firstly mRNA extracted from a sample is reverse transcribed into complementary DNA strand (cDNA) by using an enzyme reverse transcriptase. The labelling protocol differs between the two technologies. For the Agilent arrays, the cDNA is labelled with fluorescent dye Cy3 for single colour arrays or Cy3 and Cy5 for two-channel arrays. The cDNA in the labelled samples hybridizes to specific spots on the microarray containing its complementary sequence. For the Affymetrix arrays, the cDNA is not labelled but it is used as a template in an in vitro transcription reaction to generate biotinylated cRNA stand. The cRNA is then hybridized to microarrays. Finally the hybridized arrays can be scanned with a laser. The binding between the fluorescently target sequence and probes on the spot generate a signal, which is directly proportional to the initial number of RNA molecules present for that gene in the sample.

1.4.3 Typical steps in microarray data analysis

There are a number of steps involved in microarray data analysis including data processing, quality control of data, identification of differential expression analysis, functional annotation, building networks and modularization approaches (**Figure 1.4**), which are described in the next paragraphs.

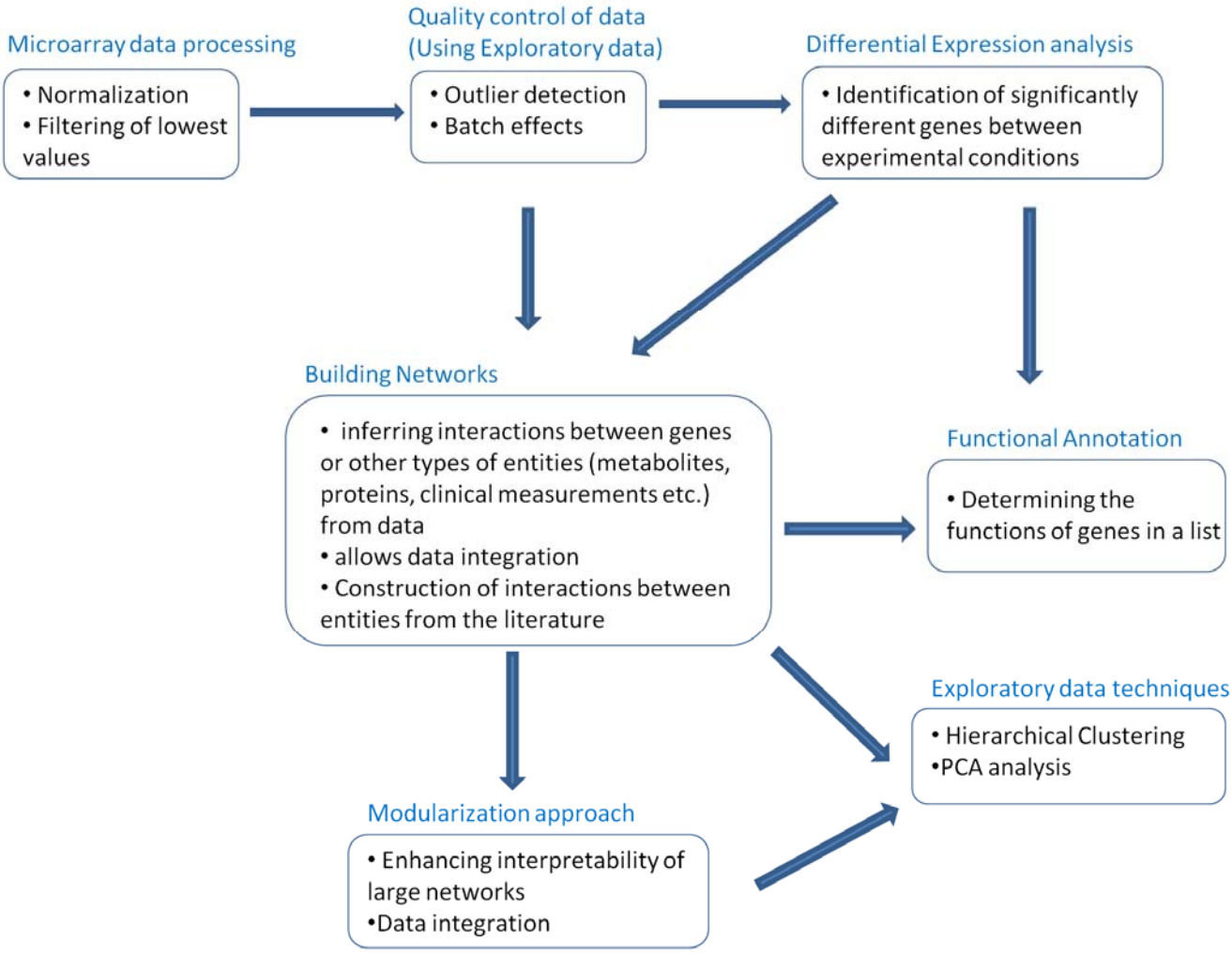


Figure 1-4. The major steps in microarray data analysis.

This figure shows the major steps for the analysis of microarray data. The normalization and filtering of the lowest values is essential for data processing. Quality control of the processed data could be achieved by checking for outliers and batch effects by performing exploratory data analysis. To identify gene expression differences between experimental conditions, there are numerous statistical tool. These genes could be then used for building gene regulatory networks, identifying modules, annotating their functional role, and identify similar expression patterns.

1.4.4 Microarray data processing and normalization

The first step in microarray data analysis is the normalization and quality control of the data. The aim of the normalization is to correct the systematic technical variation without losing the biological variation. The systematic technical variation might be caused by labelling, scanning, different labs, and different scientists generating the data. To correct these, several normalization methodologies have been developed in the field, for reviews [125, 126]. The

choices of normalization methods depend on the array type (single or two channel arrays) and the manufacturer. For single channel arrays produced with Affymetrix, the most popular methods are Robust Multi-array Analysis (RMA) [127], Guanine Cytosine Multi-array Analysis (GCRMA) [128] and Affymetrix Microarray Suite (MAS5) [129]. The RMA method provides background correction for the probes, quantile normalization and summarization into expression measurements. GCRMA is an improved version of RMA that is able to incorporate the sequence-specific probe affinities [128]. These methods are implemented in R in the package affy [130] and also several GUI interfaces exist such as babelomics [131] and Genespring GX. For the normalization of the Agilent arrays, limma package provides functions to perform loess normalization (Locally weighted polynomial regression), simple scaling or quantile.

It should be noted that before performing the normalization methods, a log2 transformation is applied to the data in order to reduce variation among measures with high magnitude, thus adjust the variance to be the same for all intensities [126]. Once the data is normalized, it is important to check for batch effects and outliers using exploratory analysis techniques.

1.4.5 Exploring the data

The exploratory data analysis techniques are crucial in many scenarios including for outlier detection, identification of sample and gene similarity/dissimilarity and dimension reduction. The commonly used techniques in the field are clustering algorithms and ordination analysis such as principal component analysis (PCA), independent component analysis (ICA). There have been numerous clustering algorithms proposed including self organizing tree algorithm (SOTA) or Hierarchical Ordered Partitioning and Collapsing Hybrid (HOPACH) [132]. In each of these algorithms, it is possible to choose the distance measure and the linkage type.

To choose between various clustering algorithms, it is possible to perform a close analysis of clusters via a visual inspection. There are also tools that allow assessing the quality of the clusters generated and comparing the results of various algorithms [133].

1.4.6 Identification of differentially expressed genes

An important question to address from microarray dataset is to identify genes that are differentially expressed between experimental conditions such as disease and healthy. To achieve this task several statistical analyses have been published and implemented. The appropriate methods can be chosen depending on the experimental setting, which might require comparison of two groups (t-test, SAM, ANOVA), more than two groups (SAM, ANOVA), or consideration of several factors such as treatment and disease types (2 Factor ANOVA), or time course data (SAM, ANOVA, BETR). These methodologies are summarized in Table 1.3. These tests can be performed in R on the command line [134] or in TMEV which provides a GUI [135]. For performance accuracy, a more recent review reporting the results of an extensive analysis comparing a wide range of these methodologies has recently been published using simulated and real datasets [136]. The authors have concluded that the selection of the significant genes heavily depends on the choice of the testing methods [136]. Another important factor in selection of significant genes relies on the threshold. Most of the statistical tests return a p-value for each gene representing the probability of observing the difference between experimental conditions by random chance. However, as the number of genes to test in a microarray dataset tends to be in the region of tens of thousands, the p-values should be corrected for the multiple testing. One of the most widely accepted approach to control false positive rate and false negative rate is False Discovery Rate (FDR) [137]. It is acceptable to choose any threshold that is lower than 10% FDR, which is equivalent to expecting 10% false positives and 90% true positive. However, the final cut off should be selected by studying the results of various thresholds. In doing so

the number of genes is important but also sample separation, reproducibility, functional annotation of these genes should be performed before deciding the threshold.

Test	Tests for comparing	Variance assumption between groups	Description and examples
t-test	One class Two class paired Two class unpaired	Unequal	Genes with low variance, which are not affected by heterogeneity, can be statistically significant despite very low fold change (called stability problem).
One way ANOVA	Two or more classes	Equal	Generalization of t-test Example of use: different treatments
Two way ANOVA	Two factors Two or more classes for each factor	Equal	Example of use: factor patient type and factor treatment
Three way ANOVA	Three factors Two or more classes for each factor	Equal	Example of use: factor patient type, factor treatment, factor training
BETR	Time-course		empirical bayes approach
SAM	One class Two class paired Two class unpaired Multiclass Time-course	Equal	Modification of t-test that removes the problem of stability encountered with t-test

Table 1.2. Summary of the most used statistical tests.

1.4.7 Functional annotation

The advantage of microarray technologies is to measure gene expression of the whole genome. However, interpreting the results from a large amount of genes selected for instance from differential expression analysis is challenging. It can also be biased as one would be

inclined to concentrate on the genes that one is familiar with. Functional annotation tools overcome this issue by performing statistical enrichment in databases such as KEGG pathways curated by experts in the fields, and GO terms. One of the most popular tools is DAVID (The Database for Annotation, Visualization and Integrated Discovery) and it provides a web interface and an R package (DAVIDQuery) to identify gene ontology terms and curated pathways that are significantly enriched in a list of genes [138, 139]. This is achieved by using a modified fisher test, called the EASE score. For only a few sets of genes the web interface is very practical, however the R package becomes the only option to avoid manual work for annotating large number of gene sets, which might be derived from clustering analysis for instance.

1.4.8 Methodologies for building networks from datasets

The aim of network inference methodologies is to capture interactions between entities such as genes, metabolites, proteins, and clinical measurements. Therefore, the network inference approaches allow data integration of various dataset and provide a framework to capture dependencies between entities. Constructing these interaction networks is crucial for the identification of disease responsible genes that can alter other entities and hence the function of important biological functions. These networks can enhance understanding of the mechanisms underlying diseases, which could lead to identification of biomarkers and development of potential drug targets.

Numerous methods have been developed for building networks (for recent reviews [140, 141, 142]). The choice of methods depends on the type of the data (time-course or static), number of samples, and number of genes. There are methodologies based on pairwise association of genes (correlation, mutual information), on probabilistic methods (Bayesian networks), and differential equations. The methods based on a pairwise score can be applied by using the

whole dataset as an input, whereas methods based on Bayesian networks and differential equations can only be applied to a small number of entities.

In the case of a study concentrating on a specific question with small number of genes to infer, the Bayesian networks and differential equations have the advantage of revealing the direction of the connection, unlike the methods based on pairwise association. There are several methodologies that have been published to infer networks using a Bayesian approach [143] and ordinary differential equations [144, 145, 146]. Systematic comparisons of these methodologies were performed in several reviews [140, 142].

To infer large networks involving more than hundred genes, several metrics can be computed between all gene pairs in the data, including correlation and mutual information. It is important to note that these relations are statistical dependencies and not causal relationship. With smaller number of samples, correlation such as Spearman and Pearson with the latest being more sensitive to nonlinear relationships can be used. ARACNE computes the mutual information between all pairs of genes and eliminates potentially indirect connections using the Data processing Inequality (DPI) principal [147]. Mutual information can capture linear and non-linear relationships but a large amount of samples (>40 samples) is required.

1.4.9 Constructing networks from the literature

Inferring accurate networks from datasets of small number of samples is not possible. Constructing networks using knowledge in the literature provides an approach to achieve this task. Cytoscape provides plugins with a user interface (BionetBuilder [148] and MIMI [149]) that allows integration of databases including DIP, BIND, Prolinks, KEGG, HPRD, The BioGrid and GO and construct networks for a given gene list [150].

Construction of networks from the literature is also very useful to define a priori knowledge in network inference approaches such as Bayesian networks to enhance the accuracy of these methodologies or for validation of the networks that are inferred from the dataset.

1.4.10 Modularization approaches

An immediate challenge rises after the inference of large networks for their interpretation and visualisation. To tackle this problem, several modularization approaches have been proposed to group entities based on certain properties including functional role, topology of the network, high co-regulation of nodes or nodes with high prediction of the phenotype. The first modularization approach was proposed by Ideker and colleagues with the aim of identifying sub-networks that consist of genes differentially expressed between experimental conditions in large protein-protein interaction networks [151]. The approaches based only on the network topology aims to identify sub-networks that are highly interconnected in a large network [152], for a recent review [153]. Other approaches include scoring of the edges using metrics such as correlation coefficient or mutual information with the aim to identify sub-networks consisting of genes that are highly co-regulated [154, 155].

1.5 Aims and outline of the thesis

The overarching goal of this thesis is to identify molecular networks linked to musculoskeletal disorders. With the development of genomics and functional genomics technologies, the ability to measure and characterise the molecular state of cells and tissues have reached an unprecedented level of sophistication. Unfortunately, this has not resulted in a comparable increase in knowledge. The main rate limiting step has been the difficulty in analysing and interpreting this large amount of data. This thesis directly address this important challenge by illustrating in several relevant biological and clinical scenarios the usefulness of advanced statistics and systems biology approaches in understanding musculoskeletal disorders.

More precisely, muscle wasting occurring in COPD patients has been studied using an integrative approach and a pathway level approach. In Chapter 2, the aim has been to identify

pathways that are co-ordinatly modulated in response to exercise training in muscle of COPD patients and aged matched healthy individuals. This work has identified a number of pathways that are modulated between COPD and healthy individuals after training (inflammatory pathways) or specifically modulated in healthy individuals in response to training (metabolism and tissue remodeling pathways). In Chapter 3, the abnormal energy metabolism, tissue remodeling pathways in relation to physiological outcome and systemic inflammation have been studied by constructing gene regulatory networks. This approach addressed the hypothesis that systemic inflammation might be a key player driving muscle wasting in these patients. This approach has outlined no link between systemic inflammation and muscle dysfunction. Instead it shows that oxygen dependent changes in the expression of epigenetic modifiers may be controlling bioenergetics imbalance in COPD.

The second aim of this thesis has been the identification of regulatory networks underlying bone and cartilage differentiation in healthy and diseased tissues (Chapter 5). First, high-level dynamical gene regulatory networks have been inferred describing the normal hMSCs differentiation to bone and cartilage cells. Then, I have aimed to identify molecular networks representing an alteration of the normal hMSCs differentiation in ageing, multiple myeloma and arthritis, a series of conditions that are associated to loss of bone and cartilage formation. For this, a novel modularization approach developed and described in Chapter 4 has been used. This methodology has been developed for integrating multi-level dataset. It allowed integration of biological knowledge, gene co-expression during normal hMSCs differentiation and differential expression analysis in disease patients. This work has lead to the identification of gene regulatory networks that consist of genes highly co-regulated in the hMSCs differentiation and at the same time differentially expressed in ageing or disease. Ultimately, the hypothesis generated using these models represent plausible targets, which

could enhance the differentiation in ageing and in disease and could be experimentally validated.

CHAPTER 2: PATHWAY ANALYSIS OF TRAINING-INDUCED EFFECTS ON SKELETAL MUSCLE TRANSCRIPTOME IN COPD PATIENTS

2.1 INTRODUCTION

In a significant percentage of patients, COPD is characterized by a number of systemic effects, such as skeletal muscle dysfunction and wasting. This leads to a disease phenotype associated with poor prognosis [156]. The precise mechanisms underlying muscle wasting in this COPD are still largely unknown.

Active life-style and skeletal muscle training are important factors improving the prognosis of COPD patients, including reduced systemic effects of the disease such as skeletal muscle dysfunction [157, 158, 159]. There is evidence suggesting that physical activity at early disease stages might have a beneficial impact on nitroso-redox balance and muscle remodeling [160], but the interplay between training and inflammatory phenomena at the systemic level is also still controversial.

Recently, mRNA expression profile in quadriceps of patients with COPD provided the first evidence for impaired gene expression in the muscle of COPD patients compared to healthy subjects [161]. These genes represented a wide spectrum of biological processes such as protein turnover, energy production, muscle regeneration, and transcription factors. Radom-Aizik and colleagues [162] performed a gene expression profiling analysis of the effects of

training on skeletal muscle biopsies of healthy and COPD individuals. The authors reported that training in healthy individuals induced modulation of a larger number of genes than in COPD [162]. However, even in healthy subjects, the reported fold changes in individual genes generated by training were too small to explain the prominent training-induced physiological effects. Hence, training-induced physiological effects are likely to be a consequence of a combination of small genomic changes within functional pathways. Consequently, in the current study, a pathway level analysis was applied using advanced bioinformatics methods that allow identification of multi-gene pathway, based molecular signatures associated with training-induced physiological changes. Differentially modulated functional pathways were identified between healthy subjects and COPD patients before and after training using the Radom-Aizik's dataset (12 weeks training) [162] and a new expression profiling study generated by our collaborators in the group of Roca (3 weeks training) [160]. Because of the differences among the endurance training programs, the two datasets were analyzed independently.

The pathway level analysis is consistent with the view that COPD muscles lose the ability to modulate key metabolic and tissue-remodeling pathways in response to training. Moreover, the results show that endurance training can induce opposite responses in inflammatory pathways, in healthy and diseased muscles.

2.2 METHODS

2.2.1 Analysis strategy

This study is based on a pathway-level analysis strategy developed to identifying molecular networks underlying a given biological response (see **Figure 2.1** for a schematic representation). In this specific project this has been applied to study muscle dysfunction in COPD. The initial step in the analysis consists of mapping genes represented in the expression profiling datasets to KEGG pathways (**Figure 2.1A**). Once genes have been mapped into pathways, each pathway activity can be represented by performing a principal component analysis (**Figure 2.1B**). The overall activity of a given pathway characterized by the expression of several genes is summarized by two variables (the first two principal components, PCs, of the PCA analysis), (**Figure 2.1C**). This approach has been previously demonstrated to be very effective [154]. Then, each pathway can be assessed for differential transcriptional activity (for example in respect to training or disease factors) as a whole (**Figure 2.1D**). The pathways can be further analyzed using the knowledge management system Ingenuity Pathway Analysis (IPA) (Palo Alto, http://www.ingenuity.com), which can help with the biological interpretation (**Figure 2.1E**).

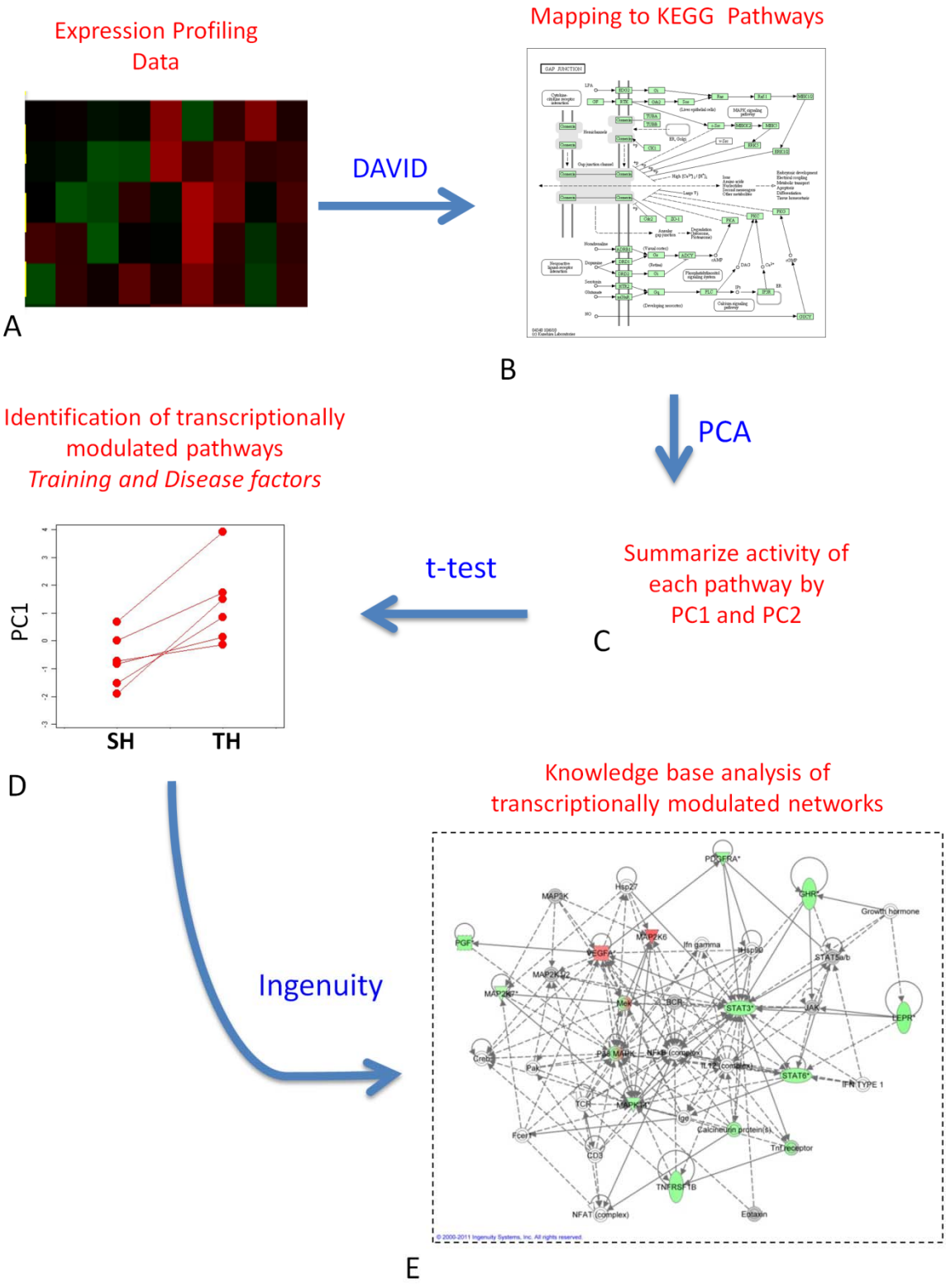


Figure 2-1. Analysis Strategy.

This figure shows in a diagrammatic format the analysis strategy employed in this study.

2.2.2 Study groups

The Radom-Aizik dataset [162] - These authors developed a gene expression profiling dataset based on Affymetrix microarray technology representing muscle biopsies derived from 6 healthy and 6 COPD individuals before and after ***13 weeks*** endurance exercise. Details of this study and of the training protocol are available in the original publication [162].

*The new dataset (3-weeks training study) – A*n additional transcriptional dataset with skeletal muscle biopsies of 5 COPD patients (aged 63 ± 2 year) and 3 age-matched healthy individuals studied before and after a much shorter (*3-weeks*) supervised endurance training program was generated in the group of Josep Roca. Details of the selection procedure, clinical assessment and measurements performed before and after training are described in [160]. This study was carried out in accordance with the ethical standards on human experimentation , which were approved by ethical committees at IMIM-Hospital del Mar (Barcelona, Catalonia, Spain), Maastricht University Hospital (Maastricht, The Netherlands), Royal Brompton Hospital (London, UK) and Cruces Hospital (Barakaldo, Basque Country, Spain). Open muscle biopsies of the *"vastus lateralis"* were obtained at rest, and after the training period. Muscle samples (~0.1 g) were placed in RNA stabilization reagent (RNAlater®, Ambion, Inc., 2130, Woodward street, Austin TX-USA) and stored at -20^{o}C for RNA extraction, which was performed by using the TRIzol Reagent (Invitrogen Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. Gene expression profiling was performed using oligonucleotide Affymetrix Human Genome Focus arrays following the manufacturer's instructions (Focus 2 arrays).

2.2.3 Mapping genes to pathways

Individual genes represented in the datasets have been mapped to the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway database using the web software package Database for

Annotation, Visualization and Integrated Discovery (DAVID) [163]. KEGG associates genes to a collection of expert defined pathways, representing the molecular interaction and reaction networks for metabolism, genetic information processing, environmental information processing, cellular processes, human diseases and drugs [164]. In this chapter, two independent analyses using two datasets published in [162] and [160] were analyzed.

2.2.4 Summarizing pathway activity

The analysis performed in this study is based on identifying KEGG pathways, which are differentially modulated between two groups, as described below. This strategy has great advantages compared to the analysis of individual genes because it simplifies the complexity of data analysis reducing the input data from tens of thousands of genes to a few hundreds of well characterized functional modules. It generally produces an increase in statistical power and improves biological interpretability [140].

One possible technique to summarize pathway activities is to use principal component analysis (PCA), which allows the representation of complex multi-dimensional datasets in a lower dimensional space by retaining the majority of the information [165]. Therefore, by using this technique the overall activity of a given pathway characterized by the expression of several genes is summarized by two variables (the first two principal components, PCs, of the PCA analysis), which describe up to 70% of the information present in the original data. These two indexes of pathway activity are then tested for their ability to separate the two sample groups. The procedure has been implemented in the statistical environment R [134]. The code is available for download on our group web site (http://biptemp.bham.ac.uk/software.html).

2.2.5 Pathway and gene level statistical analysis

Differential modulations of a pathway activity represented by the corresponding PCs have been tested by performing Student t-tests between groups (unpaired analysis) and within groups (paired analysis). First, the differences between COPD patients and healthy sedentary controls before training were compared. To assess training-induced effects, muscle biopsies before and after training within each group were compared. Finally, post-training data were also compared between the two groups.

A threshold for significance (FDR<10%) has been defined after correction for multiple testing using Benjamini-Hochberg false discovery rate (FDR) procedure [166]. Accordingly, functional pathway components associated with an FDR of 10% were considered differentially modulated and selected for further analysis and biological interpretation. For further analysis, heatmaps from standardized values of the PCs representing the pathway activity have been generated.

2.2.6 Ingenuity Pathway Analysis

Using the Ingenuity Pathway Analysis (IPA) software (Palo Alto, http://www.ingenuity.com,) biological networks linked to transcriptionally modulated pathways were identified. Once the gene lists were uploaded into the application, each gene identifier was mapped to its corresponding gene object in the Ingenuity Pathways Knowledge Base. These genes, called focus genes, were overlaid onto a global molecular network developed from information contained in the Ingenuity Pathways Knowledge Base. Networks of these focus genes were then algorithmically generated based on their connectivity according to the following procedure implemented in the IPA software application. The specificity of connection for each focus gene was calculated by the percentage of its connection to other focus genes. The initiation and the growth of pathways proceed from the gene with the highest specificity of connections. Each network had a maximum of 35 genes

for easier interpretation and visual inspection. Pathways of highly interconnected genes were identified by statistical likelihood. Networks with a Score greater than 20 and containing more than 60% of focus genes were selected for biological interpretation.

2.3 RESULTS

2.3.1 Mapping genes to pathways

Mapping individual genes to KEGG pathways lead to 166 pathways associated to the Radom-Aizik dataset [162] and 157 pathways associated to the dataset published in [160]. The two pathway sets are completely overlapping. These represent respectively 77% and 73% of the annotated pathways in the KEGG database and 33% and 38% of the annotated probes in the two datasets respectively [162, 160]. The high coverage both at functional and gene level in the derive datasets give us the confidence that the derived datasets are representative of the transcriptional state of muscle biopsies.

2.3.2 A comparison between Healthy subjects and COPD patients before endurance training

Functional module analysis of the data generated by Radom-Aizik and colleagues [162] did not reveal any significant difference between healthy subjects and COPD patients in overall activity of the pathways. However, the analysis of the new dataset revealed differences in the activity of 4 pathways (**Table 2.1**). Two of these are metabolic pathways (polyunsaturated fatty acid biosynthesis and glutathione pathways). The other two are (bladder cancer and colorectal cancer pathways) related to processes involved in angiogenesis (VEGF signaling component) and tissue remodeling (MAPK and TGFß signaling component).

PCs	KEGG pathways	Group
PC1	Glutathione metabolism	metabolism

PC1	Polyunsaturated fatty acid biosynthesis	Metabolism
PC2	Bladder cancer	Signaling
PC2	Colorectal cancer	Signaling

Table 2.1 This table shows all pathways which have a difference of >=1 between sedentary healthy and sedentary COPD patients (FDR<10%) using the new dataset.

2.3.3 Training effects on healthy subjects and COPD patients

In healthy subjects, training induced a massive response at pathway level in both datasets. Up to 79 differentially expressed pathways were seen in the dataset published in [162] and 52 pathways were detected in the new dataset [160], see appendix **Table A2.1** and **Table A2.2** respectively. These pathways belong to inflammation, metabolism and tissue remodeling.

In COPD patients, training had a smaller effect on the muscle transcriptome than in healthy individuals, confirming the results previously reported by [162]. Only 12 differentially active pathways were identified using the data published in [162] and 3 pathways in the new dataset (3), see appendix **Table A2.3** and **Table A2.4** respectively.

In the 12 weeks training study published by Radom-Aizik [162], healthy and COPD patients modulate the same inflammatory pathways in opposite directions (**Figures 2.2A and 2.2D and Figure 2.4**). In the shorter 3 weeks training study, only healthy individuals modulate inflammatory pathways (**Figure 2.3**).

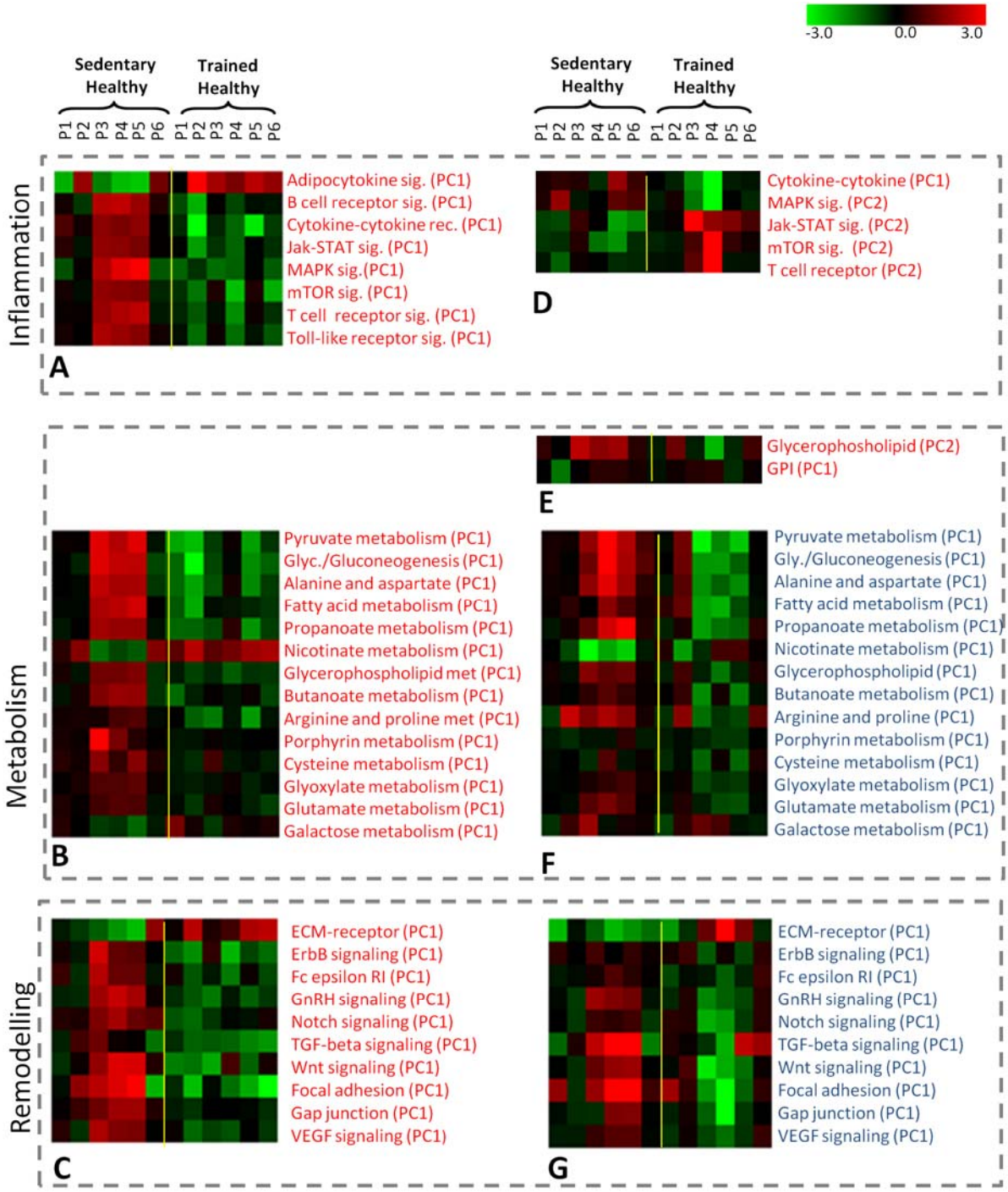


Figure 2-2. Pathway components modulated in response to training (RA study)

This figure shows the heatmaps generated from the standardized values of the PCs representing pathway activity. It shows significantly activated inflammatory related pathways between sedentary vs. trained healthy people (panel A) and sedentary vs. trained COPD patients (panel D) using Radom-Azik's dataset. The panel B and C show the heatmap of significantly activated (respectively) metabolism and tissue remodeling pathways between sedentary vs. trained healthy people. These pathways are not significantly activated between sedentary vs. trained COPD patients (panel E, panel F). However, the heatmaps shows that this is because of 3 patients which are not able to respond to training. Panel D shows the significantly activated metabolism pathways after training in COPD patients.

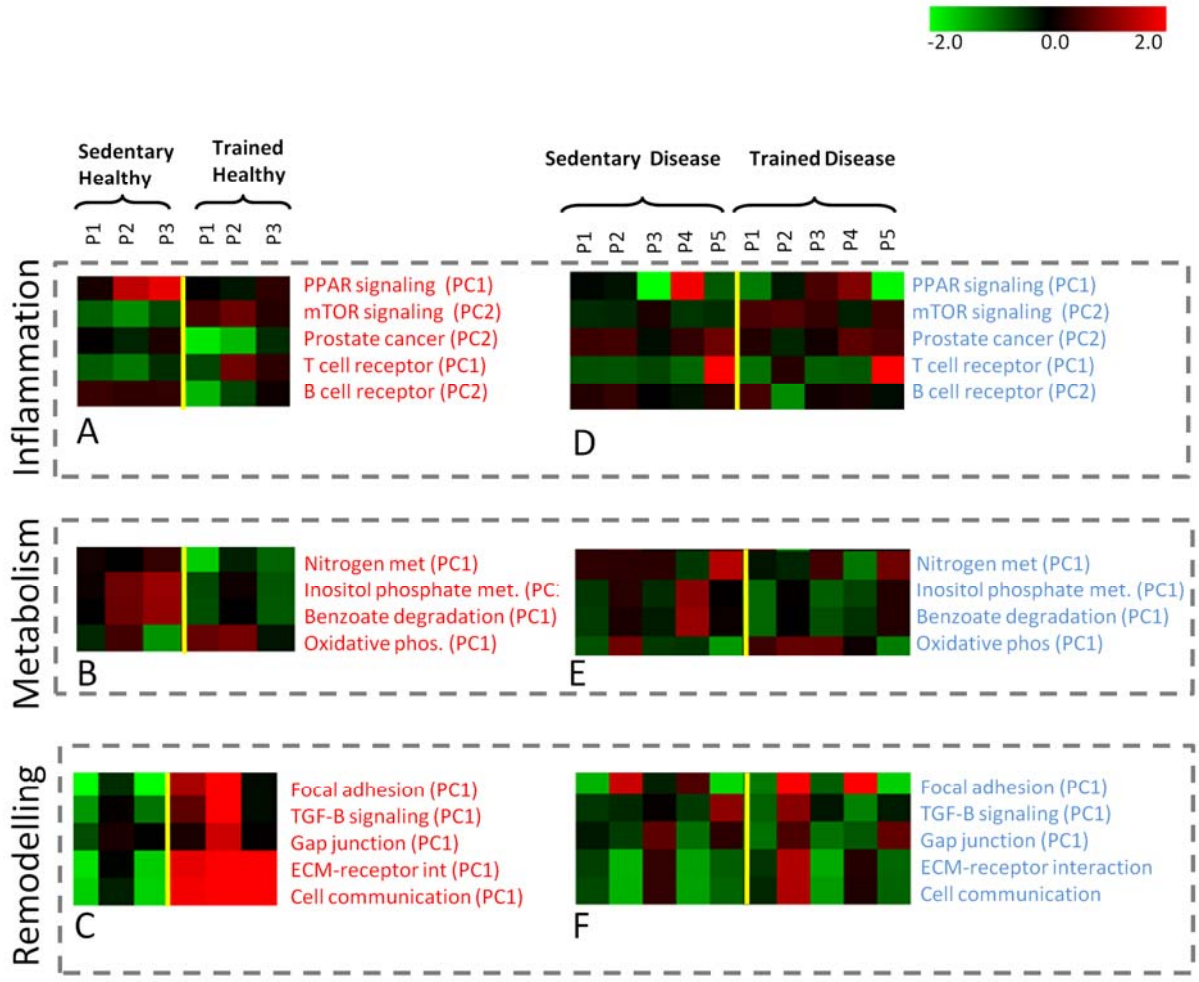


Figure 2-3. Pathway components modulated in response to training (new study)

This figure shows the heatmaps the standardized values of the PCs representing pathway activity. The heatmaps are generated from the significantly activated inflammatory and metabolism related pathways (represented in red) between sedentary vs. trained healthy people (panel A, panel B, panel C respectively) using the new dataset. The panel D, E, F shows the pathways between sedentary vs. trained COPD of the previously identified as significant between sedentary vs. trained healthy people. These pathways are not significantly different between sedentary disease and trained COPD patients (represented in blue).

My analysis also revealed that components of tissue remodeling pathways are activated in response to training in healthy individuals, but these pathways fail to be modulated in COPD individuals (**Figure 2.2** panels C and H, **Figure 2.3** panels C and F, and **Figure 2.4**). Training only effectively modulates the activity of tissue remodeling pathways such as gap junction and focal adhesion in healthy individuals. In contrast, COPD patients show either a very heterogeneous response or no response at all. Similarly, COPD patients fail to activate the

metabolic pathways and show a very heterogeneous response. Specifically, pathways belonging to lipid metabolism are activated in response to training in healthy and COPD patients whereas those belonging to carbohydrate metabolism, amino acid metabolism, and metabolism of cofactors and vitamins are only activated in response to training in healthy people. Analysis at the gene level for these pathways shows that the genes, which are significantly different in response to training in healthy patients are mostly up-regulated, whereas there are no significant changes in COPD patients (fold >1.5).

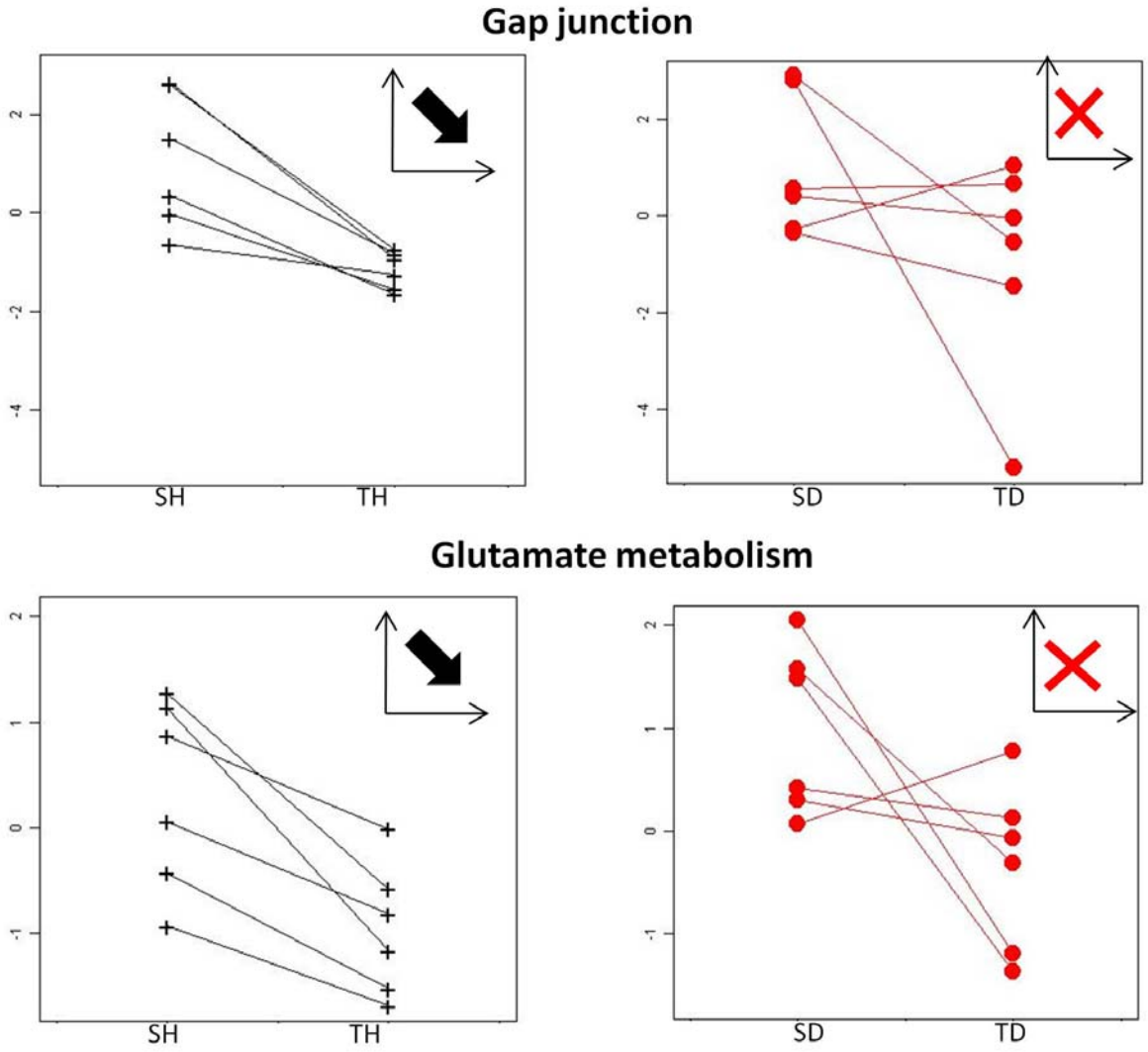


Figure 2-4.Tissue remodeling pathways and Glutamate metabolism pathway

This figure shows the response of training in Healthy (represented in black) and in COPD (represented in red) individuals tissue remodeling and glutamate metabolism pathways using

Radom-Aizik dataset. These pathways are all modulated in response to training in healthy people. However, in gap junction and glutamate metabolism only 3 out of 6 patients respond to training.

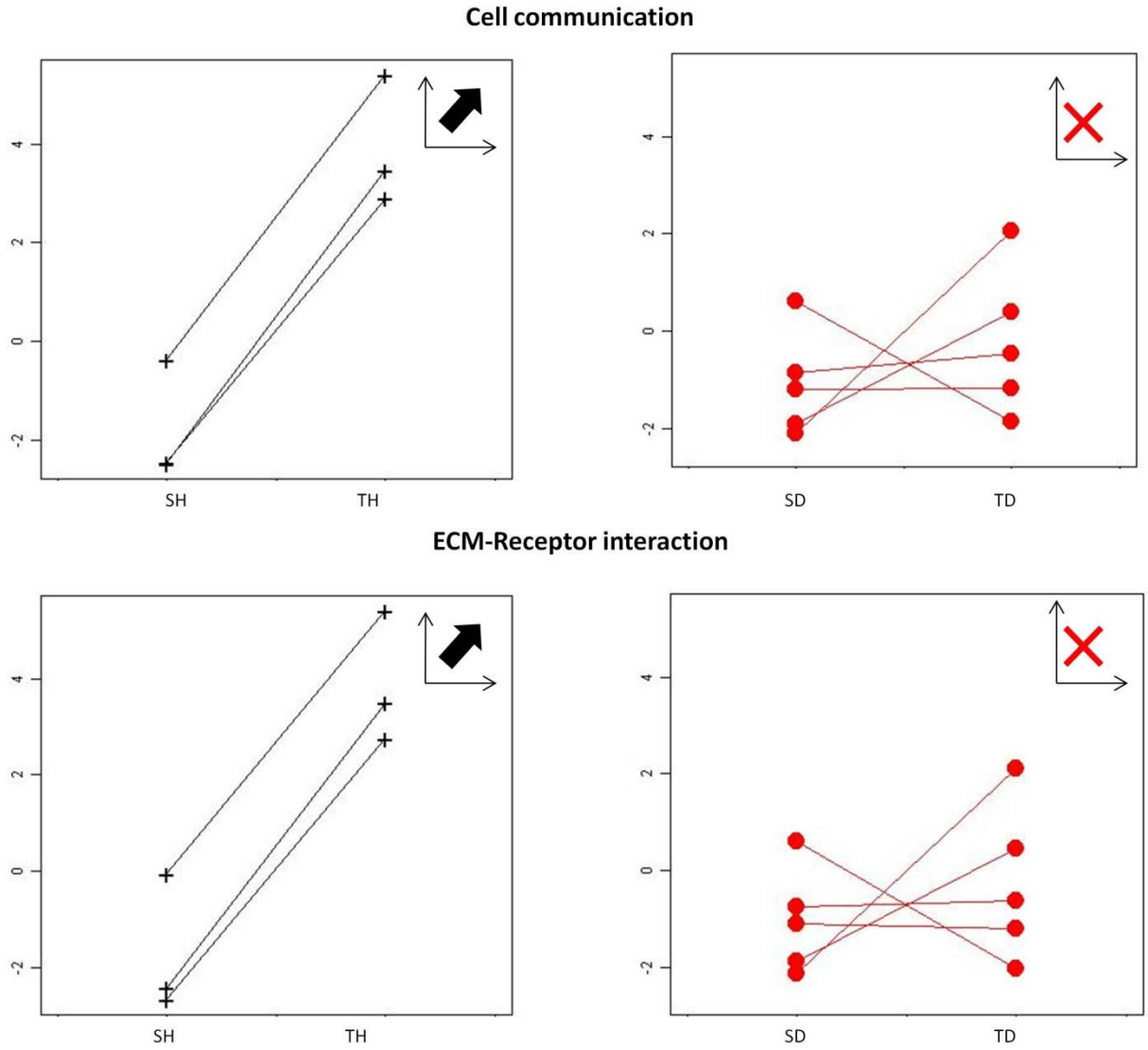


Figure 2-5. Tissue remodeling pathways

This figure shows the response of training in Healthy (represented in black) and in COPD (represented in red) individuals tissue remodeling and glutamate metabolism pathways using the new dataset.

2.3.4 Direct comparisons between healthy and diseased samples after training

There were 23 differentially active pathways between the two groups using the dataset published in [162], (**Tables 2.2**). These pathways represent metabolism and tissue remodeling

functions confirming the effects of exercise training separately in each group. Of these pathways obtained using Radom-Aizik dataset [162], 6 belong to metabolism (riboflavin metabolism, nicotinate and nicotinamide metabolism, nitrogen metabolism, fructose and mannose metabolism, pyrimidine metabolism), 2 belong to tissue remodeling (gap junction, focal adhesion) and 3 belong to inflammation (cytokine-cytokine receptor interaction, mTOR signaling pathway , ErbB signaling) (**Table 2.2**). The significantly differentially active pathways identified using the new dataset belong to metabolism (nicotinate and nicotinamide metabolism, glutamate metabolism, Ala13 and aspartate metabolism) and cell-to-cell communication (focal adhesion, cell communication, ECM-receptor interaction), (**Table 2.3).**

PC	Kegg pathway	Functional group
PC1	Melanoma	signaling pathway
PC1	ErbB signaling pathway	tissue remodeling
PC1	Cytokine-cytokine receptor interaction	inflammation
PC1	Nicotinate and nicotinamide metabolism	metabolism
PC1	mTOR signaling pathway	inflammation
PC1	Phosphatidylinositol signaling system	signaling pathway
PC1	Complement and coagulation cascades	signaling pathway
PC1	Neurodegenerative Diseases	signaling pathway
PC1	Nitrogen metabolism	metabolism
PC1	Riboflavin metabolism	metabolism
PC2	Focal adhesion	tissue remodeling
PC2	Adipocytokine signaling pathway	signaling pathway
PC2	Gap junction	tissue remodeling
PC2	Pancreatic cancer	signaling pathway
PC2	Pyrimidine metabolism	metabolism
PC2	Fructose and mannose metabolism	metabolism
PC2	Hedgehog signaling pathway	signaling pathway

Table 2.2. This table shows all pathways which has a difference of >=1 between trained healthy and trained COPD patients (FDR<10%) using the dataset generated by Radom-Aizik.

PCs	Kegg	Groups
PC1	Cell Communication	tissue remodeling
PC1	ECM-receptor interaction	tissue remodeling
PC2	Glutamate metabolism	metabolism
PC2	Alanine and aspartate metabolism	metabolism
PC2	Nicotinate and nicotinamide metabolism	metabolism
PC2	Alkaloid biosynthesis II	metabolism
PC2	Focal adhesion	tissue remodeling
PC2	Hematopoietic cell lineage	Metabolism

Table 2.3. This table shows all pathways which has a difference of >=1 between trained healthy and trained COPD individuals (FDR<10%) using the new dataset.

2.3.4 Analysis of the genes that are not mapped to the pathways

The analysis performed in this chapter focuses on detecting differences in the coordinate regulation of genes in a given pathway as opposed to identify individual gene expression. However, many genes could not be mapped into KEGG pathways. In order to not lose information from the discarded genes, it is important to carry out an analysis at the gene level. Here, the genes that are identified as significantly differentially expressed between COPD and healthy before and after exercise have been functionally annotated in DAVID. This analysis gives consistent results with the pathway analysis, especially for metabolic and tissue remodeling pathways (Table 2.5). The analysis derived from the new dataset shows only a few significantly enriched functions, but tissue remodeling functions are also enriched in response to training in healthy individuals but not in COPD individuals (Table 2.5). Results obtained from Radom-Aizik dataset show that the functional representation of the genes differentially expressed in response to training in healthy belong to bioenergetic pathways (mitochondrion, oxidative phosphorylation and mRNA metabolic process) and tissue remodeling (cell adhesion and ECM), (Table 2.4). The lack of response to training in

COPD patients is also observed in this analysis. The genes that did not map to KEGG pathways follow a similar functional trend. One major difference between this analysis and the pathway based analysis is the inflammatory pathways. Here, functional groups related to inflammation were not identified.

	All the genes	**Genes that are not mapped into KEGG pathways**
SH vs. TH	GOCC ribonucleoprotein complex [52] GOCC mitochondrion [45] GOBP mRNA metabolic process [21] GOBP oxidative phosphorylation [8] GOBP cell adhesion [5] GOCC ECM [4]	GOCC ribosomal protein complex [22] GO mitochondrion [21] mRNA metabolic process [19] GOBP oxidative phosphorylation [0]
SD vs. TD	GOCC Sarcomere [4] GOBP Oxidation reduction [7]	GOCC Sarcomere [2] GOBP Oxidation reduction [1]
TH vs. TD	GOBP regulation of translation [12] GOCC ribonucleoprotein complex [14] GOCC Sarcomere[5] GOBP mRNA metabolic process [12]	GOBP Translation [4] GOCC ribonucleoprotein complex [7] GOCC Sarcomere[3] GOBP mRNA metabolic process [5]

Table 2.4. Functional annotation of the genes significantly different between COPD and healthy before and after exercise (Radom-Aizik dataset).

This first column of this table shows the comparison between the experimental groups (SH: sedentary healthy, TH: trained healthy, SD: sedentary disease, TD: trained disease). The second column is the functional annotation of the genes that are significantly different between the groups indicated in column 1. The third column represents the functional annotation of the genes that were not mapped to the KEGG pathways. The functions in grey are not significantly enriched.

	All the genes	**Genes that are not mapped into KEGG pathways**
SH vs. TH	GOCC ECM-rec interaction[4] GOCC collagen [3]	GOCC ECM-rec interaction[0] GOCC collagen [0]
SD vs. TD	GOBP DNA metabolic process [10]	GOBP DNA metabolic process [7]
TH vs. TD	GOBP translation [4] GOCC ECM [4]	GOBP translation [4] GOCC ECM [3]

Table 2.5. Functional annotation of the genes significantly different between COPD and healthy before and after exercise (new dataset).

This first column of this table shows the comparison between the experimental groups (SH: sedentary healthy, TH: trained healthy, SD: sedentary disease, TD: trained disease). The second column is the functional annotation of the genes that are significantly different between the groups indicated in column 1. The third column represents the functional annotation of the genes that were not mapped to the KEGG pathways. The functions in grey are not significantly enriched.

2.3.5 Network Analysis

In order to further elucidate the detailed structure of molecular networks underlying differences in training response between healthy and COPD muscles, genes represented in the three groups of functionally related KEGG pathways (metabolism, inflammation and tissue remodeling) were used as input to the knowledge management system IPA.

The highest scoring network representing response to training in healthy individuals for genes linked to KEGG ***tissue remodeling*** pathways represent the interaction between components of the extracellular matrix (*COL3A1, COL4A1, COL6A3, COL5A2*), which are potentially induced by *TGFa* and components of the MAP Kinase signaling pathway (*MAP3K3, MAP2K7, MAPK14, MAPKAPK3*). The network also represents components of Notch signaling (the direct targets of notch *MAPK14, ID2* and the transcription factor *HES1*) (**Figure 2.6A**).

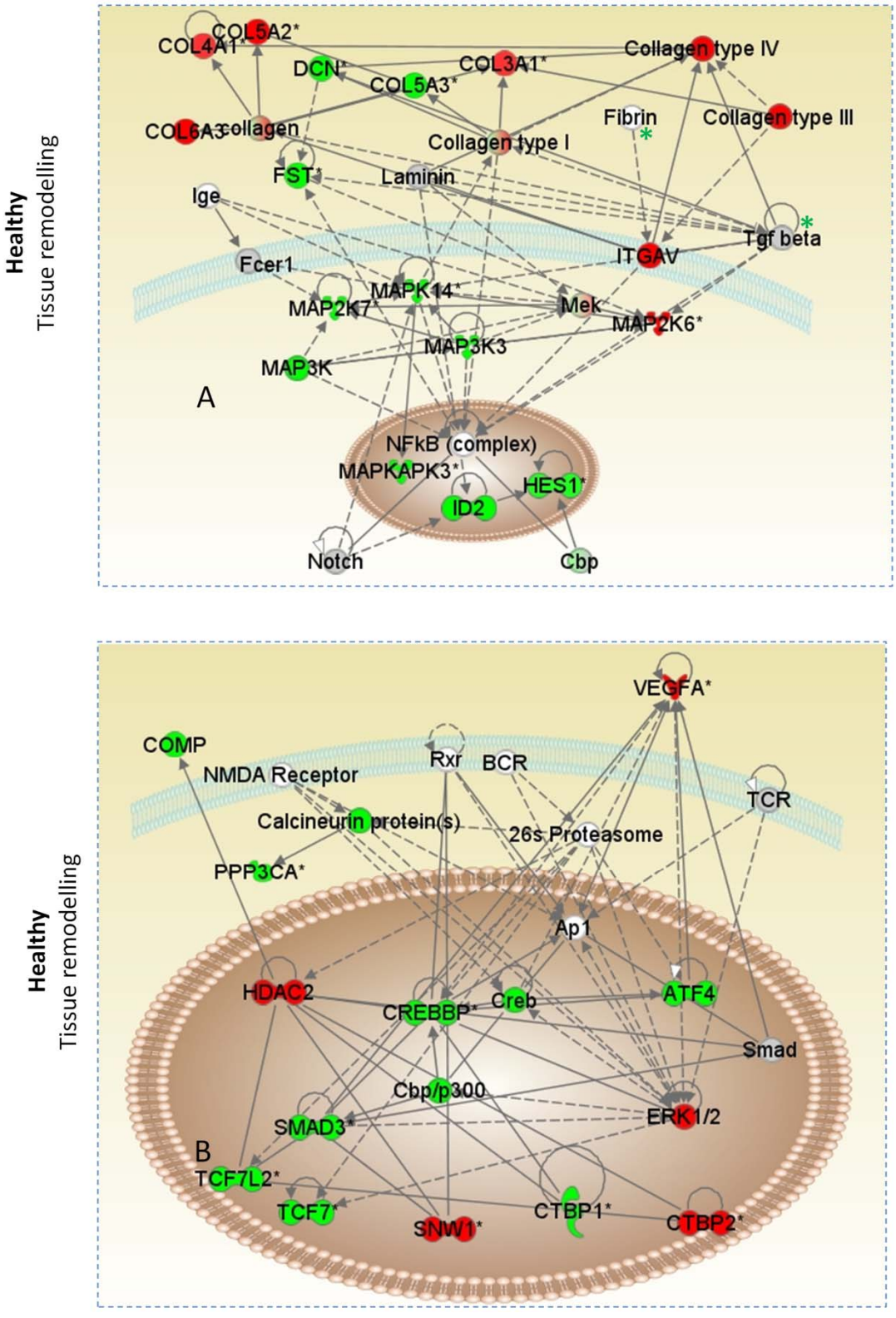


Figure 2-6. Networks derived from Ingenuity pathway analysis representing response to training in healthy patients.

This figure shows the top scored networks derived from the list of genes that are significantly different in response to training in healthy patients that are linked to tissue remodeling (Panel A and B). Genes that are upregulated are in response to training are represented in red,

whereas downregulated genes are represented in green. The genes marked with a green star (Fibrin and TGF-B) are not belonging to the tissue remodelling pathways but these are significantly downregulated in response to training in healthy patients. However, these genes fail to be modulated in COPD muscles.

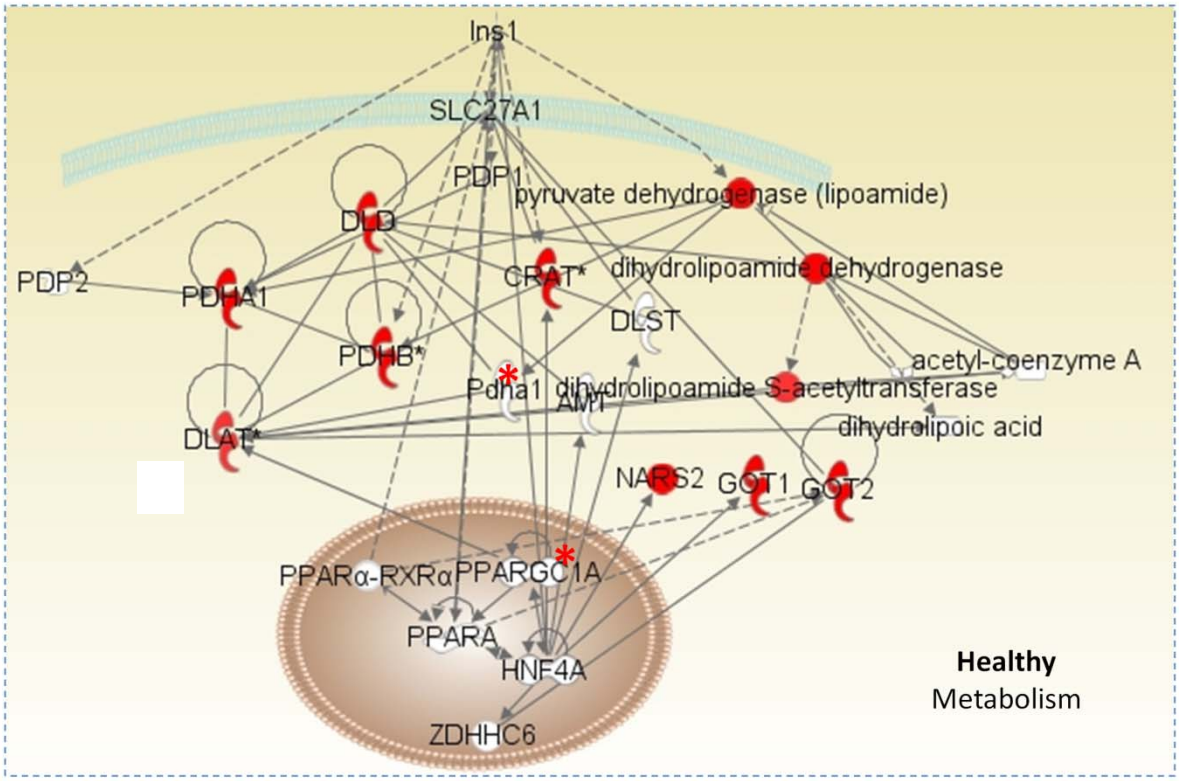


Figure 2-7. Networks derived from Ingenuity pathway analysis of genes involved in metabolism.

This figure shows the top scored networks derived from the list of genes that are significantly different in response to training in healthy patients that are linked to metabolism. The genes marked with a red star (Pdha1 and PPARGC1A) are not belonging to the metabolic pathways (the input to ingenuity analysis) but these are significantly upregulated in response to training in healthy patients However, these genes fail to be modulated in COPD muscles.

The second statistically significant network linked to KEGG ***tissue remodeling*** pathways represents a network linking the transcription regulators *SMAD3, TCFL2, TCF7, ATF4, SNW1,* and *CREBBP* to the epigenetic modifier *HDAC2* and the angiogenesis factor *VEGFA*.

Both networks show that effector molecules such as the angiogenesis factor *VEGFA* and collagen components are up-regulated in response to training. Both Notch signaling

components, which could potentially function as negative regulators of myogenesis [167] in sedentary state and effector genes fail to be modulated in COPD muscles.

The highest scoring network representing response to training in healthy individuals for genes linked to KEGG ***metabolism*** pathways shows the interaction between components of the bioenergetics pathways (*PDHA1, PDHB, DLAT, DLD* and *DLST*) and amino acid metabolism (*GOT1* and *GOT2*) (**Figure 2.7**). All these enzymes are up regulated in response to training in healthy but fail to do so in COPD muscles. Interestingly, the IPA software also identifies insulin as a key upstream regulator of this metabolic regulatory network.

2.4 DISCUSSION

By reducing the complexity of the transcriptomics data, a set of pathways that are modulated between COPD and healthy individuals after training (inflammatory pathways) or specifically modulated in healthy individuals in response to training (metabolism and tissue remodeling pathways) were identified. The separate analyses carried out in the two data sets showed consistent results, particularly in respect to tissue remodeling and metabolic pathways, supporting the hypothesis that diseased muscles are unable to mount an effective response to training, and in turn validating the approach. This is particularly important since the two datasets [162, 160] presented different training strategies both in terms of intensity and duration.

2.4.1 Tissue remodeling pathways are only modulated in response to training in healthy patients

The results derived from this study suggest that only healthy individuals are able to co-ordinately modulate tissue-remodeling pathways in response to training. The tissue-remodeling pathway derived from ingenuity has shown that *VEGF* is only up-regulated in response to training in healthy patients. Exercise induced angiogenesis in healthy patients is an adaptive response to training resulting in an increase capillaries number per muscle fibre. In the same pathway, *HDAC2* is up-regulated in response to training only in healthy muscles, which is known to be a critical player during the myogenic program for modulation and deacytylation of MyoD [168]. The current model for muscle wasting involves the inactivation of the myogenesis pathway driven by the expression of the *MyoD* gene.

2.4.2 Differential modulation of metabolic pathways between healthy and COPD muscles

The significantly modulated pathways suggest that oxidative stress and cell hypoxia are possible factors contributing to differentiate normal and diseased muscle in the pre-training state. For example, the glutathione pathway is differentially modulated between healthy and diseased muscles before training (the new dataset). The glutathione pathway is involved in protecting cells from oxidant-induced lung injury and inflammation [169]. In this pathway, the expression level of glutathione transferase is reduced in COPD patients compared to healthy subjects, which is inversely associated to glutamate-cysteine ligase level. Glutathione is quantitatively the most important intracellular antioxidant and serves multiple cellular and metabolic functions [170]. Even a relatively moderate depletion of intracellular GSH causes oxidative damage and impairs the structural and functional integrity of mitochondria [171]. Several studies have demonstrated that increased protein oxidation in the muscle of COPD patients plays a major role in the muscle dysfunction of these patients. However, normal [94] or increased level [172] of glutathione has been reported in COPD compared to controls. My analysis supports an alteration on the glutathione metabolism, which might explain the activation of the oxidative stress. The IPA analysis has revealed additional information on the transcriptional response of enzymes involved in bioenergetics and amino acid metabolism. These results reinforce the concept that diseased muscles are unable to activate the expression of enzymes involved in bioenergetics and on interconnected amino acid metabolism pathways (Figure 6B). It is reasonable to hypothesize that failure to transcriptionally activate these genes may contribute to muscle wasting in COPD.

2.4.3 Study limitations

The pathway analysis has shown that the lack of response to exercise training in metabolic and tissue remodelling pathways in COPD patients is due to patient heterogeneity. Hence, a

larger of number of samples is needed to better understand this lack of response to training in these patients. Merging the two datasets presented in this Chapter represent a potential way to increase the number of samples to study. However, this is not possible because of the different training intensity and duration carried out in the two datasets.

2.5 Conclusions

From a more methodological prospective, it is worth noticing that the analysis approach adopted in this study was integrated into the knowledge management system BioXM and published in [173]. The availability of such tools would indeed make available this analytical approach to a broader set of users.

This study has contributed to characterize at the pathway level the transcriptional programs in response to training, in healthy and COPD individuals. The picture that emerges from this analysis is consistent with the view that diseased muscles are unable to activate metabolism and tissue remodeling pathways in the two training regimes.

CHAPTER 3: A SYSTEMS BIOLOGY APPROACH IDENTIFIES MOLECULAR NETWORKS DEFINING SKELETAL MUSCLE ABNORMALITIES IN COPD

3.1 INTRODUCTION

The previous chapter describes the pathway level response to training in skeletal of COPD patients with muscle wasting and healthy aged matched individuals. This approach has shown that COPD muscles are unable to activate metabolism and tissue remodeling pathways in the two training regimes. Moreover, inflammatory pathways are either modulated in the opposite direction or lack the response to training. An interesting observation that emerged from the pathway analysis is that there is considerable molecular heterogeneity in the response of COPD muscles to training. Although the idea that COPD is a highly heterogeneous disease is well accepted, a molecular characterization of such diversity is going to require profiling a much larger population of individuals than the current studies. In this chapter, a larger multi-level clinical dataset is studied in order to address the molecular mechanisms underlying the abnormal energy metabolism, tissue remodeling pathway and the role of systemic inflammation (as published in [174]). The complex interplay of molecular pathways that are potentially involved in regulating muscle functionality makes systems biology approaches a desirable option. Such an approach aims to model the relationship between key molecular and

physiological variables in healthy and diseased individuals to derive a testable hypothesis on the disease mechanism.

In this study, skeletal muscle abnormalities in COPD were hypothesized to be the result of an imbalance in the physiological regulation of normal muscle homeostasis induced by systemic inflammatory mediators and chronic tissue hypoxia. To test this, a reverse engineering approach capturing relations between variables defining whole body physiology, systemic inflammation and muscle transcriptional state, with particular reference to cell bioenergetics and tissue remodeling functions. This analysis is based on a clinical study representing 12 healthy subjects and 18 age and sex-matched COPD patients, before and after undergoing an 8 weeks training program. In the latter category, patients with preserved muscle mass ($COPD_N$) and patients showing muscle wasting ($COPD_L$) were included.

In this study, COPD muscles are characterized by a lack of correlation in expression of bioenergetic and tissue remodeling pathways, which included genes specifically involved in myogenesis. My analysis suggests that failure to activate and coordinate these functions in response to training is associated with a general lack of activation of *NF-kB* targets, including many pro-inflammatory signals (e.g. *IL-1β*). This study lead to discovery that expression of chromatin modification enzymes, known to control muscle differentiation and energy balance in other biological systems, is abnormal in COPD muscles and correlated with oxygen availability. This finding raises the possibility that an epigenetic mechanism triggered by tissue hypoxia may be the basis of skeletal muscle wasting in COPD.

3.2 METHODS

3.2.1 Ethics Statement

All animal work has been conducted according to relevant national and international guidelines and approved by the University of Birmingham, Medical School ethics committee.

3.2.2 Study groups and design

The 8 weeks training project (TP) was a clinical investigation in which eighteen COPD patients (68±7 yrs, 17 men, FEV_1 46±12% predicted and PaO_2 75±0.7 mmHg) with a wide spectrum of body mass composition and twelve age-matched healthy sedentary controls (65 yrs, 10 men, FEV_1 107±14% predicted and PaO_2 93±0.7 mmHg) underwent a protocol of supervised endurance exercise. The inclusion criteria were: 1) diagnosis of COPD according to GOLD criteria, 2) a stable clinical condition using standard treatment with bronchodilators and inhaled corticosteroids, 3) absence of episodes of exacerbation or oral steroid treatment in the previous 4 months, 4) absence of significant co-morbidities. All procedures were performed in the Pulmonary Function Laboratory or the Rehabilitation Unit at the Hospital Clinic-IDIPAPS. The TP included twelve patients with stable COPD and normal fat free mass index (FFMI, 21Kg/m^2) ($COPD_N$), six COPD patients with low FFMI (16Kg/m^2) (COP_L), and twelve healthy sedentary subjects (FFMI 21Kg/m^2). The study was approved by the Ethics Committee of the Hospital Clinic (Barcelona, Spain) and all patients gave written informed consent. Subjects completed all phases of the protocol as well as fully contributing to the sampling regime.

Significant physiological training effects were obtained using a standard supervised interval training program. **Table 3.1** and **Figure 3.1A** summarize the characteristics of the study groups and the training-induced effects on physiological variables. Constant-work rate exercise at 70% of pre-training Watts peak (Wpeak) ($CardiO_2$ cycle Medical Graphics

Corporation, USA) was carried out before and after 8-weeks training with cycloergometer, until pre-training endurance time exhaustion. Measurements before and after training were obtained at isowork-rate and iso-time. This dataset and **Table 3.1** have been generated by Josep Roca's group.

		COPD patients	Healthy subjects
VO_2 peak, ml/min^{-1}	**Pre**	896±250	1566±410
	Post	1076±381	1819±533
	Δ post-pre	180±183†	253±204†
VO_2 peak, %pred	**Pre**	69±20	88±15
	Post	84±25	102±19
	Δ post-pre	15±11‡	14±11†
Wpeak, watts	**Pre**	56±21	111±27
	Post	73±29	135±37
	Δ post-pre	17±11‡	24±13‡
6MWT, meters	**Pre**	443±123	585±87
	Post	472±132	633±68
	Δ post-pre	29±25‡	48±51*

Table 3.1. Physiology measurements showing the effect of endurance training in healthy and COPD individuals, generated in the group of Josep Roca.

3.2.3 Measurement of inflammatory mediators in serum

Serum samples obtained at rest before training (Basal-BT) and at rest after the eight weeks training program (Basal-AT) were measured by Luminex xMAP technology, according to the manufacturer's instructions BIO-RAD [175]. This experiment was conducted by Anna Stincone. A number of cytokines/growth factors were selected for analysis based on the results of a previous proteomic study carried out in COPD patients and controls [176]. In that study [176] a total of 142 cytokines/growth factors were analysed and 42 of them were

differentially expressed in COPD. Our selection includes 30 analytes covering ~50% of the cytokines/growth factors previously identified [176], which demonstrated association to clinical parameters in the same study. The complete list of proteins measured and the corresponding results of the analysis is shown in **Table 3.3 and Figure 3.1B**. This dataset has been generated by Anna Stincone (University of Birmingham).

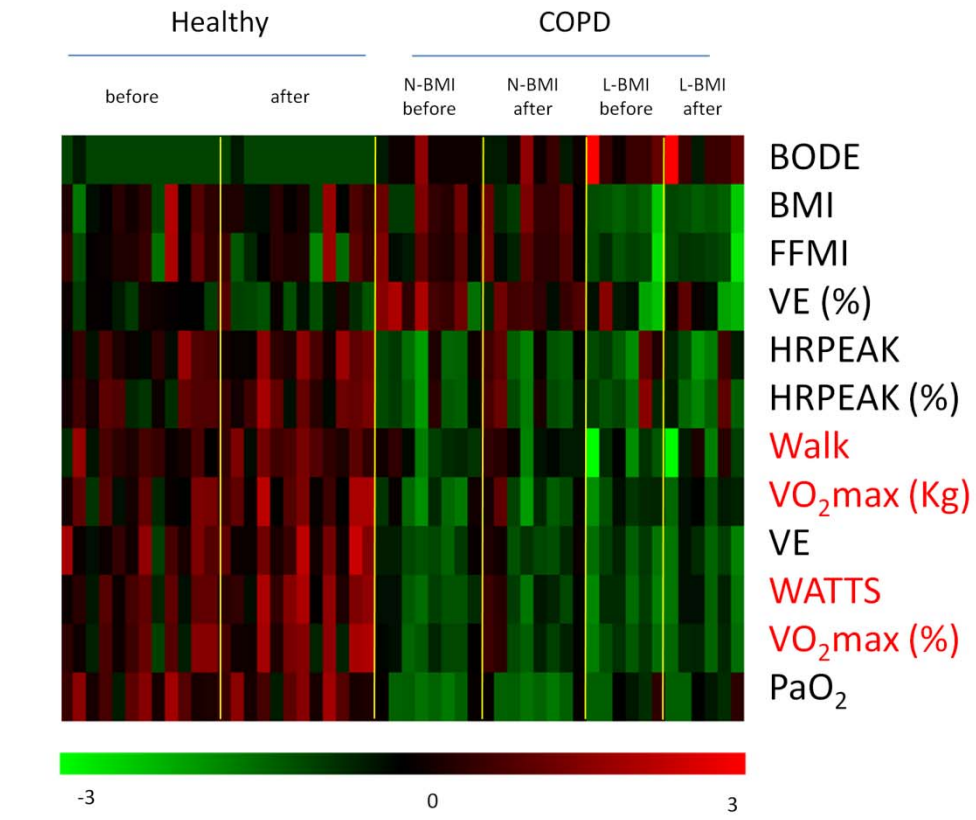


Figure 3-1. 8-weeks training study: Disease associated cytokines and Physiological measurements.

The figure shows two heatmaps representing respectively the standardized values for the physiology and serum cytokine measurements. Data are standardized to represent above average values in red and below average values in green. Panel A represents significantly different physiological measurements in the disease groups (Healthy, L-BMI and L-BMI). Those variables that are also significantly changing in response to training are labelled in red. Panel B represent serum cytokine measurements significantly different in the disease groups

(Healthy, N-BMI and L-BMI). No cytokines were significantly different in response to training.

3.2.4 Expression profiling of muscle biopsies

Skeletal muscle transcriptomics was performed on open biopsies from the *m. vastus lateralis* (quadriceps). In all participants these were obtained at rest, before and after training. RNA was isolated using RNeasy extraction kits (QUIAGEN, USA) according to the manufacturer instructions. Microarray gene expression analysis employing Affymetrix ® GeneChip technology was performed using Human U133 Plus2 Gene Chips according to the manufacturer's suggested protocols. Data were subsequently subject to quality control to assess integrity of the RNA using an Agilent Bioanalyzer (Agilent). The microarrays were generated in the group of Josep Roca.

These datasets were normalised using the R library gcrma, which converts CEL files using a robust multi-array average (RMA) expression measure with the help of probe sequences. Data were quantile normalised [177].

3.2.5 Identification of statistically significant differences in physiology, serum protein and muscle gene expression measurements.

In order to identify differential changes in physiological and protein measurements, two-factor ANOVA was used with disease and training as factors. Significantly different variables were selected using a threshold of $P<0.01$. Similarly, genes differentially expressed between sedentary and trained subjects in the three populations (healthy, $COPD_N$ and $COPD_L$ patients) were identified by t-test followed by Benjamimi-Hochberg multiple correction [137] using a false discovery rate (FDR) threshold of $q<10\%$.

3.2.6 Network inference

Building a model including mRNA, protein and physiology measurements representing pre and post-training samples in the 8 weeks study

In order to develop molecular networks representing the interaction between systemic and local signals, physiology, blood cytokine levels and muscle mRNA expression measurements were integrated by standardizing each variable, across all samples to have mean of 0 and standard deviation of 1 using the R package QuantPsyc [178]. After this procedure, measurements, in different units become directly comparable. A global interaction network was built as the union of individual network modules, each one of them defined as a cluster of variables linked by a significant correlation to the profile of key biomarkers. Such biomarkers, which can be thought of as network hubs, were chosen to represent biological factors as these could be important players in the context of the clinical problem. These can be summarized in four distinct groups (for a detailed list see **Table 3.2**):

1) Indicators of whole body physiology.
2) Serum protein levels representing systemic inflammatory signals.
3) Muscle mRNA expression profiles of genes encoding for receptors of cytokines and growth factors, known to play a role in muscle physiology.
4) Muscle mRNA expression profiles of genes encoding for enzymes in the glycolysis/gluconeogenesis pathway.

The ability to infer relationships between measured variables from observational data is at the core of this study. For this reason, a validated network inference technique ARACNE was used. This tool infers the interaction between pairs of variables using a measure of correlation called mutual information (MI) [179, 180]. Sparse networks are generated by eliminating non-significant connections according to a permutation-based statistical test. Although ARACNE has been used mainly in association with gene expression data it can be applied to

infer interactions between any variable that have quantitative measurements. ARACNE offers numerous advantages over more traditional measures of correlation, including the ability to spot non-linear correlations, which are very effective in identifying biologically relevant connections [179].

In order to infer networks integrating multi-level measurements from the eight weeks training study, ARACNE was used to reconstruct the network neighbourhoods of muscle-expressed intermediate metabolism genes, immune system and growth factor receptors genes (measured in muscle biopsies, using microarray technology), systemic cytokines (using Luminex technology, measured in serum), and physiological measures (**Table 3.2**). Statistically significant network edges have been selected using a threshold $p<10^{-7}$ and no edges have been eliminated using DPI. Significant connections were identified by applying a *P*-value threshold for significant MI values that ensure less than 5% false positives [147]. Individual networks generated with ARACNE were merged and the resulting network visualised using the software application Cytoscape [181], a force-driven layout where the strength of association in the graph is determined by the value of mutual information associated with the gene-to-gene connection. This layout allows the representation of a network structure where the position of the nodes is determined by the relative strength of connections. Genes that are strongly linked (i.e. by a high MI value) will appear closely associated in the graph, whereas genes with a weak association will be far apart. This procedure was applied to infer all networks presented in this paper.

gene symbol	gene name	function
ALDOA	ALDOLASE A	glycolysis
GAPDH	GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE	glycolysis
PFKM	PHOSPHOFRUCTOKINASE, MUSCLE	glycolysis
TPI1	TRIOSEPHOSPHATE ISOMERASE 1	glycolysis
PGM1	PHOSPHOGLUCOMUTASE 1	glycolysis
PDHAI	PYRUVATE DEHYDROGENASE	glycolysis
ALDH2	ALDEHYDE DEHYDROGENASE 2 FAMILY (MITOCHONDRIAL)	glycolysis
HK1	HEXOKINASE 1	glycolysis
DLAT	DIHYDROLIPOAMIDE S-ACETYLTRANSFERASE	glycolysis
ACYP1	ACYLPHOSPHATASE 1, ERYTHROCYTE (COMMON) TYPE	glycolysis
PGK1	PHOSPHOGLYCERATE KINASE 1	gluconeogenesis
DLD	DIHYDROLIPOAMIDE DEHYDROGENASE	gluconeogenesis
AKR1A1	ALDO-KETO REDUCTASE FAMILY 1, MEMBER A1 (ALDEHYDE REDUCTASE)	gluconeogenesis
ENO1	ENOLASE 1	gluconeogenesis
FBP2	FRUCTOSE-1,6-BISPHOSPHATASE 2	gluconeogenesis
CCR5	CHEMOKINE (C-C MOTIF) RECEPTOR 5	receptor
CSF1R	COLONY STIMULATING FACTOR 1 RECEPTOR	receptor
EGFR	EPIDERMAL GROWTH FACTOR RECEPTOR	receptor
FGFR1	FIBROBLAST GROWTH FACTOR RECEPTOR 1	receptor
IL-17RA	INTERLEUKIN 17 RECEPTOR	receptor
IL1RN	INTERLEUKIN 1 RECEPTOR ANTAGONIST	receptor
IL8RB	INTERLEUKIN 8 RECEPTOR	receptor
TNFR1	TUMOR NECROSIS FACTOR RECEPTOR TYPE 1	receptor
TNFR2	TUMOR NECROSIS FACTOR RECEPTOR TYPE 2	receptor
TrkB	TYROSINE KINASE, RECEPTOR, TYPE 2	receptor
VEGFR2	VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTOR 2	receptor
CD44R not sure (HA)	HERMES ANTIGEN GP90 HOMING RECEPTOR	receptor
IL-1r	INTERLEUKIN 1 RECEPTOR, TYPE I	receptor
IL1RB	INTERLEUKIN 1 RECEPTOR, TYPE II	receptor
IL13RA1	INTERLEUKIN 13 RECEPTOR, ALPHA 1	receptor
IFNAR	INTERFERON RECEPTOR 1	receptor
Mig-R	CHEMOKINE (C-X-C MOTIF) RECEPTOR 3	receptor
CCR3	CHEMOKINE (C-C MOTIF) RECEPTOR 3	receptor
HEGFR (MET)	HEPATOCYTE GROWTH FACTOR RECEPTOR	receptor
IFNGR2	INTERFERON GAMMA RECEPTOR 2	receptor
IFNGR1	INTERFERON GAMMA RECEPTOR 1	receptor
BMI	BODY MASS INDEX	physiological variable
FFMI	FAT FREE MASS INDEX	physiological variable
Walk	WALK	physiological variable
HRPEAK	HEART RATE PEAK	physiological variable
HRPEAKper	HEART RATE PEAK PERCENTAGE	physiological variable
VO2maxkg	WHOLE OXYGEN COMSUPTION AT PEAK OF EXERCISE	physiological variable
VO2_maxper	WHOLE OXYGEN COMSUPTION AT PEAK OF EXERCISE (PERCENTAGE)	physiological variable
WATTS	WATTS	physiological variable
VE	MINUTE VENTILATION	physiological variable
VE_per	MINUTE VENTILATION (PERCENTAGE)	physiological variable
VO2_max	WHOLE OXYGEN COMSUPTION AT PEAK OF EXERCISE	physiological variable
BODE	BODY-MASS INDEX, AIRFLOW OBSTRUCTION, DYSPNEA, AND EXERCISE CAPACITY INDEX	physiological variable
age	AGE	physiological variable
PaO2	PRESSURE OF OXYGEN IN ARTERIAL BLOOD	physiological variable
Eotaxin-exp2	SMALL-INDUCIBLE CYTOKINE A11	serum cytokine
IP10-exp2	INTERFERON-INDUCIBLE CYTOKINE	serum cytokine
MIP-1b-exp2	MAJOR INTRINSIC PROTEIN OF LENS FIBER	serum cytokine
Rantes-exp2	RANTES	serum cytokine

Table 3.2. List of physiology, serum cytokines and immune system/growth factors receptors gene expression measurements used as network hubs in the ARACNE procedure.

IL-1beta	IL-6	IL-15	MIP 1alpha	MCP-1
IL1RA	IL-7	IL-17	MIP 1beta	VEGF
IL-2	IL-8	TNF alpha	IP-10	G-CSF
IL-2R	IL-10	IFN alpha	MIG	EGF
IL-4	IL-12	IFN gamma	Eotaxin	FGF-basic
IL-5	IL-13	GM-CSF	RANTES	HGF

Table 3.3. Serum and cytokine growth factors measured using Luminex technology.

Moreover, gene lists derived from statistical analysis of microarray data and from the analysis of interaction networks have been functionally annotated using a Fisher test as implemented in the web-based application DAVID [163]. A Benjamini-Hochberg FDR correction for multiple testing was applied and a threshold of FDR <0.01 chosen for selecting significant enrichment of specific Gene Ontology (GO) or Kyoto Encyclopedia of Genes and Genomes (KEGG) terms in the different network modules.

Individual network models representing healthy and COPD muscles

In order to address whether the transcriptional uncoupling between genes involved in tissue remodeling and bioenergetics was a specific feature of COPD, the procedure described above was also applied to healthy and COPD muscle datasets. In order to increase the resolution, 8w training dataset with a similar microarray study recently performed by Radom-Aizik *et al*. [162] were integrated. This dataset provided microarray data representing an additional 12 healthy and 12 age-matched COPD patients pre and post endurance training (12 weeks). This dataset was quantile normalized using the R library gcrma and microarray batch effects were corrected using an Empirical Bayes framework available in R [182].

Network models representing diabetes and muscle dystrophy muscle biopsies

In this chapter, two additional datasets were also analyzed for studying muscle biopsies derived from patients affected by diabetes and dystrophy. The first microarray dataset consists of skeletal muscle biopsy samples from 43 age-matched males with different degrees of glucose intolerance and including Type 2 diabetes mellitus (available at: http://expression.gnf.org/) [183]. The dystrophy dataset consist of 84 samples from patients representing several types of muscle dystrophy (Facioscapulohumeral muscular dystrophy samples, Becker muscular dystrophy, Emery Dreifuss muscular dystrophy, from Duchenne muscular dystrophy) (GEO accession: GSE3307) [184].

3.2.7 Mouse *in vivo* experiments

To assess the effects of pro-inflammatory cytokines on skeletal muscle, recombinant mouse IL-1β (10mg/ml) was injected in a single dose into the tail vein of C57BL10 mice at a loading of 100 ng/mouse (4 animals per group) and samples taken from lateral gastrocnemius (glycolytic) and soleus (oxidative) muscles 24h later, with saline-injected animals acting as controls. Samples were stored at –80^{o}C until use. RNA was extracted using an RNeasy RNA extraction kit (QUIAGEN). Microarray expression profiling of four independent biological replicates was performed using full genome oligonucleotide arrays (OPERON) after labelling the mRNA with the Cy-Scribe post labelling kit (Amersham) according to the manufacturer's instructions. Genes differentially expressed were identified by a t-test followed by multiple testing correction (FDR<10%). Significant genes from the mouse experiment were mapped on the human COPD network by converting mouse gene identities into their human homologues using the database Homologene (data was normalised using the same procedures described for the COPD dataset).

3.2.8 Ingenuity Pathway Analysis

The gene set represented by those directly connected to growth factor and cytokine receptor components of the network and differentially expressed in response to training in healthy individuals were analyzed using the Ingenuity Pathway Analysis (IPA) application (Palo Alto, http://www.ingenuity.com). Networks with a score greater than 20 and containing more than 60% of the genes differentially expressed in response to training were selected for biological interpretation.

3.2.9 Molecular response of mouse quadriceps muscles to chronic hypoxia conditions

In order to assess whether genes correlated to VO2max in the clinical study are indeed modulated in response to hypoxia, a publicly available microarray dataset has been analyzed in this chapter. This dataset represents the transcriptional response of C57Bl/10 mice to 2 weeks of chronic hypoxia and compared those with matching controls kept in normoxic conditions. The dataset was developed by Budak et al. [185], and it is available in the Gene Expression Omnibus database under the accession number GSE9400. The details of the overall design and the dataset can be found in GEO.

For clarity, here a concise summary of the experimental design and data generation is reported. Adult male C57Bl/10 mice were divided into two groups, control (room condition) and hypoxic, over a period of 2 weeks. The hypoxic group was gradually exposed to lower levels of hypoxia in a specially designed and hermetically closed hypoxic chamber. A Pegas 4000 MF (Columbus Instruments) gas blending system was used. The oxygen level was gradually decreased from 21% to 8% over one week and animals were kept at 8% oxygen for another 7 days. After two weeks animals were euthanized using CO_2 and total RNA was isolated from each quadriceps femoris (QF) muscle by Tri-Reagent (Ambion, Austin, TX). Microarray analysis was then performed using an Affymetrix Mouse Genome 430 2.0 Array

according to the standard Affymetrix protocol (Affymetrix, Santa Clara, CA). Data were then quantile normalized using the R library gcrma [177]. Differentially expressed genes between normoxic and hypoxic mice were identified using SAM analysis [186]. Genes that have an FDR smaller than 10% were used to map to the neighborhood of the VO2max in the COPD network. Genes linked to the GO term chromatin remodeling were identified using the web-based tool DAVID [163].

3.3 RESULTS

3.3.1 Integration of physiological and molecular responses acquired in the 8 weeks training study within a network inference framework

With the aim of developing an initial high level model representing the relationships between systemic and local signals, all available measurements were integrated (physiology, blood cytokine levels and muscle gene expression measurements) for the individuals recruited in the 8 weeks training study. These samples included both pre and post-training blood and biopsy samples. The individual network modules generated using the hubs in **Table 3.2** were then tested for enrichment in GO and KEGG functional terms, which revealed a remarkably coherent association between target genes and the biological functions of network hubs.

Neighbourhoods of physiological variables, such as VO_2 peak, were mainly enriched in glycolysis/gluconeogenesis, mitochondria respiratory chain, oxidative phosphorylation (all positively correlated to VO_2 peak), and components of the ribonucleoprotein complex and RNA processing (negatively correlated to VO_2 peak, **Figure 3.2** and **Table 3.4**). Genes associated with mRNA expression of cytokine and growth factor receptor hubs were enriched in functions related to tissue remodeling and immune response (generally positively correlated with the expression of receptors for growth factors and cytokines, **Table 3.5**), while genes associated with expression of glycolysis/gluconeogenesis enzymes were mainly enriched in energy metabolism-related functions (**Table 3.6**). Systemic inflammatory cytokines were weakly connected to the network (only three significant correlations with genes linked to physiological measurements).

The union of the individual modules and consequent visualization with a force driven layout revealed a clear separation between two main sub-networks. First, a large sub-network,

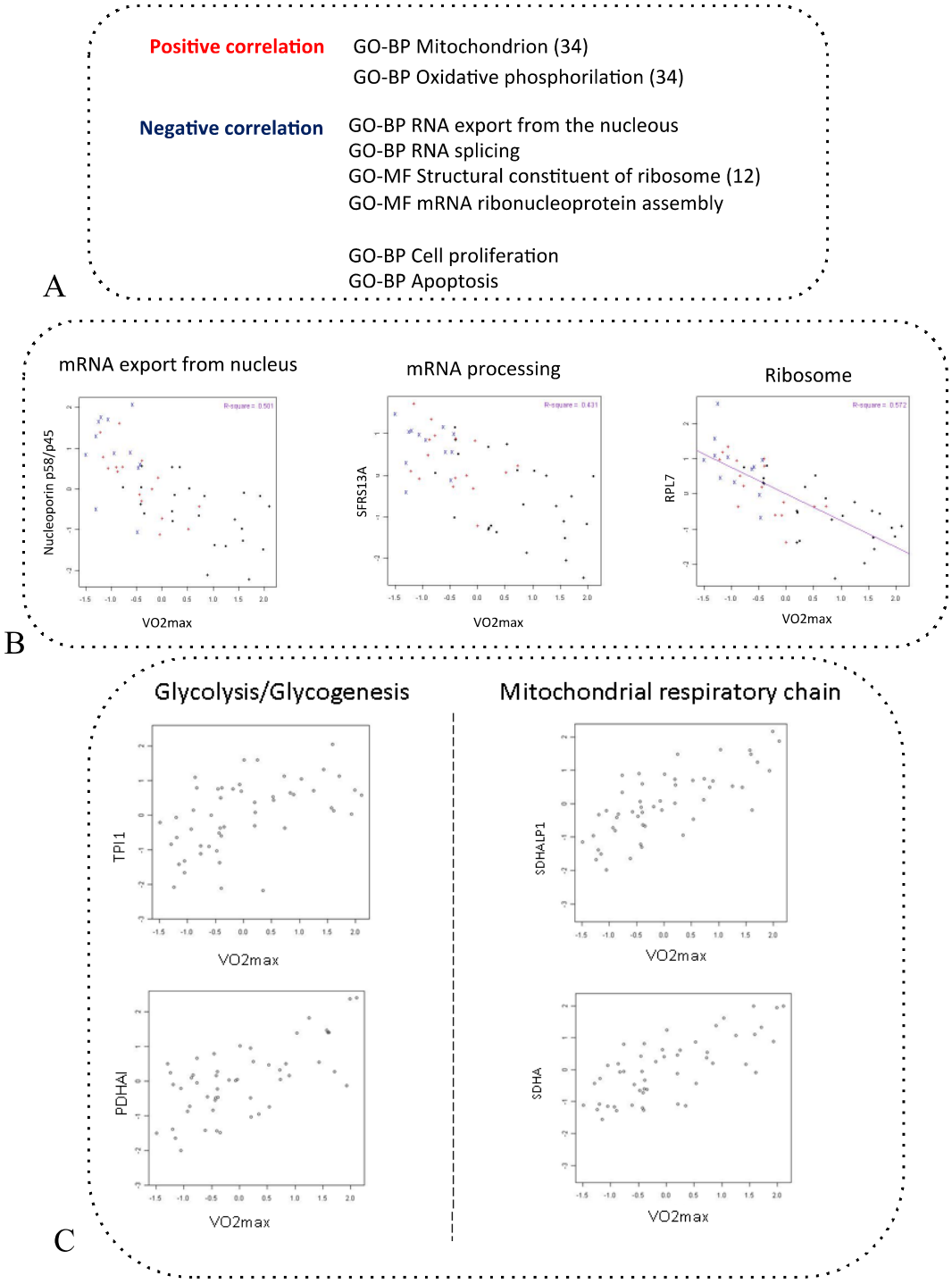


Figure 3-2. Correlation between VO2max and its neighbourhood as inferred by ARACNE.

Panel A represents the functional annotation of the genes, which are positively and negatively correlated with VO2max. Panel B shows instances of genes belonging to mRNA export from nucleus, mRNA processing, and Ribosome functional terms. Panel C shows instances of genes that belong to Glycolysis/Glycogenesis and to mitochondrion respiratory chain functional terms.

primarily representing the association of modules built from cytokines (both plasma and skeletal muscle) and growth factor receptor hubs and enriched by tissue remodeling functions. Second, a sub-network representing the association of modules built with physiology and skeletal muscle glycolysis/gluconeogenesis hubs and enriched in energy metabolism genes **(Figure 3.4)**. Within this sub-network, many of the physiological variables (PaO2, HRPEAK, HRPEAKper, WATTS, VE, VO2max, VO2maxper, VO2maxkg, BODE) were highly correlated (**Figure 3.3**) and consequently clustered together (**Figure 3.4**) to form a very compact network module. The other physiological variables (VEper, FFMI, BMI and age) were separated from this module and were characterized by a limited number of connections with gene expression variables (30 genes).

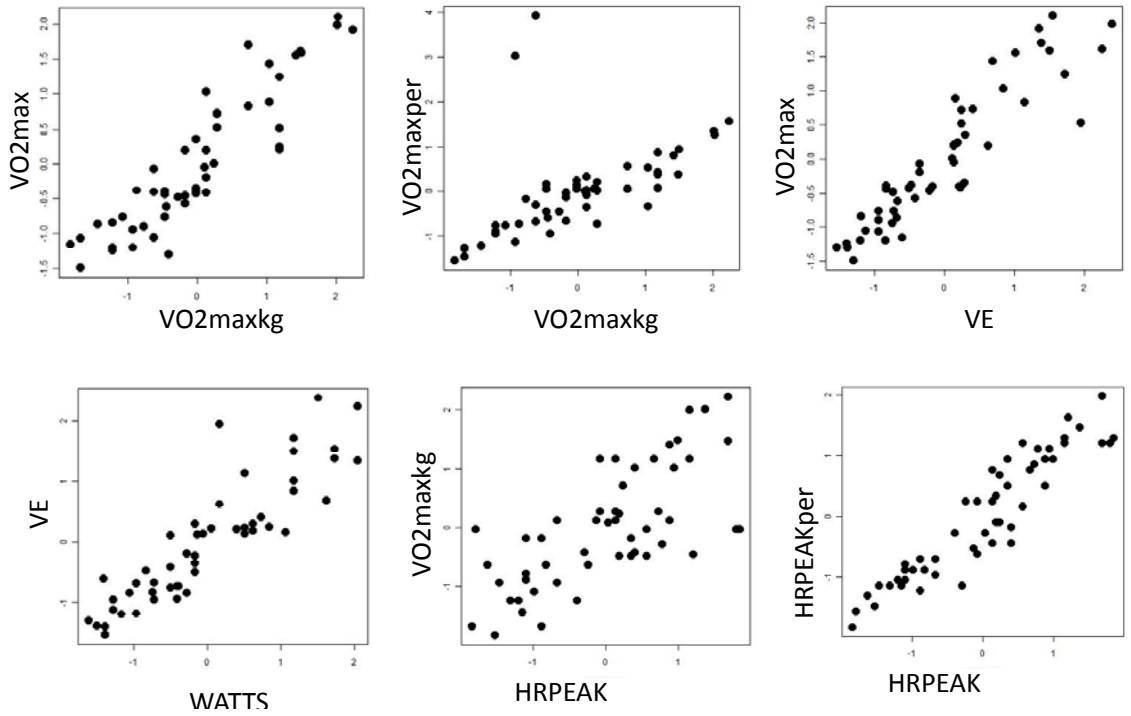


Figure 3-3. Correlation between physiological variables.

This figure shows the scatterplots between physiological variables that are highly co-regulated and form a compact subnetwork.

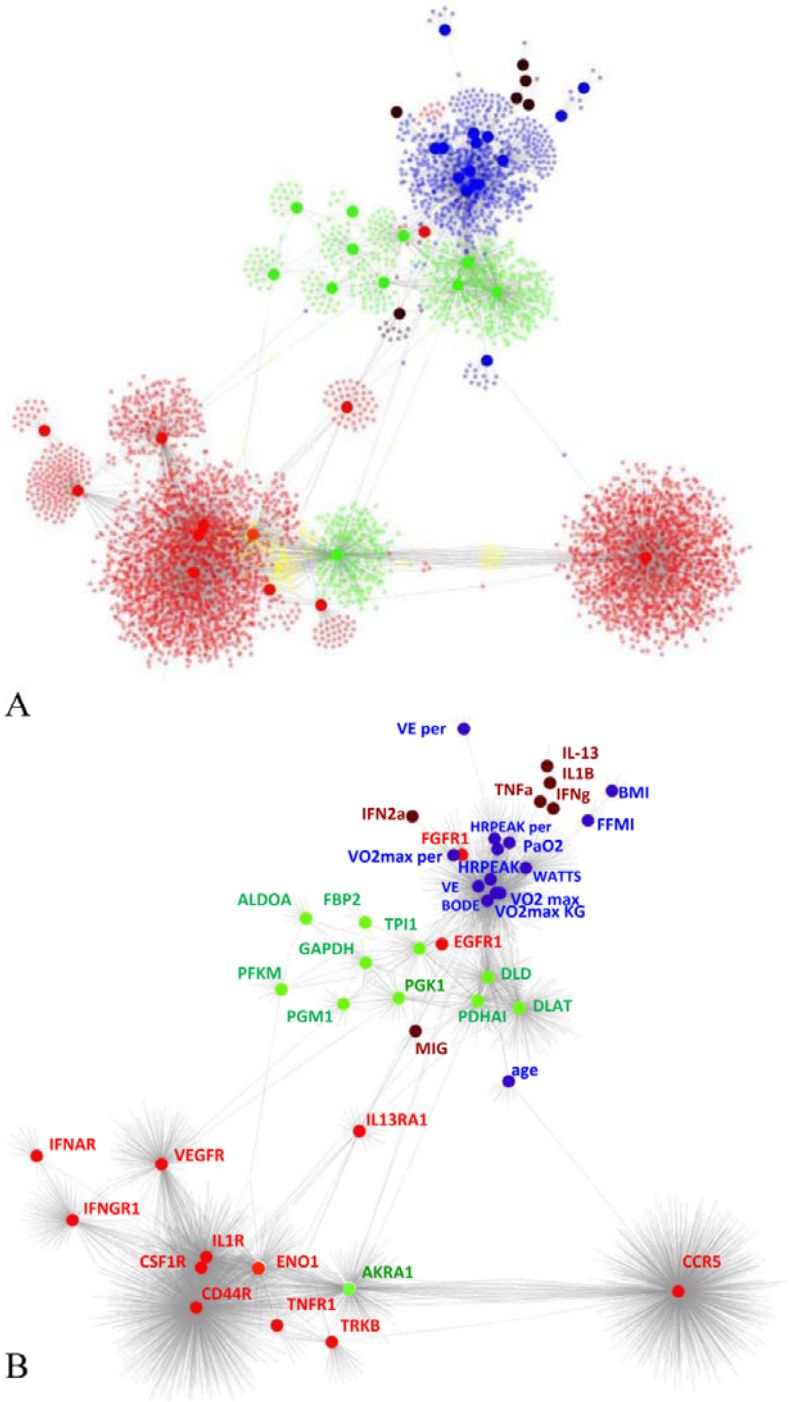


Figure 3-4. Network integrating multi-level measurements in the 8 weeks training study.

The figure shows the network representing the relationship between physiological measurements, serum cytokine levels and mRNA expression profiles in the muscle, following ARACNE analysis of the 8 weeks training dataset. Panel A shows an outline of the network topology where individual network hubs have been annotated. Red nodes represent genes coding for receptors for growth factors and immune system mediators, blue nodes represent physiological measurements, brown nodes represent serum cytokines, and green nodes represent gene expression of enzymes in the glycolysis and gluconeogenesis pathways. Panel B shows all nodes in the network (representing 16350 genes or measurements). The colour of nodes follows the same code as panel A. In addition yellow nodes represent genes in common between cytokine and growth factor and glycolysis/gluconeogenesis network hubs.

Hubs	No. genes	Representative Functional Terms
VE	149	[1] GOCC cytoplasm (79), intracellular part (101). [3] GOCC cytosolic large ribosomal subunit. [4] GOCC mitochondrial part.
VO2-max	259	[1] GOCC cytoplasm (116),GOBP cellular metabolic process (126). [2] GOCC cytoplasmic part (79). [3] GOCC mitochondrion (34). [4] KEGG Oxidative phospohorylation (8), GOCC mitochondrial respiratory chain (7).
VO2-maxkg	301	[1] GOCC cytoplasm (127). [2] GOCC intracellular organnele part (70). [6] GOCC organelle envelope (22), GOCC mitochondrion (28).
Walk	144	[1] GOCC cytoplasm (78). [2] GOCC nucleolus (8).
WATT	249	[1] GOCC cytoplasm (124). [2] GOCC organelle part (63). [3] GOCC mitochondrial membrane (16). [4] GOBP Cellular metabolic process (119). [6] GOMF RNA bindgin (23). [9] GOBP macromolecule localization (25). [13] GOCC ribosomal subunit (6). [14] GOCC mitochondrial respiratory chain (6).

Table 3.4. COPD interaction network.

Representative functional terms enriched in the neighbourhood of Physiological variables.

Hubs	No. of genes	Representative Functional Terms
CCR3	1815	[1] GOBP negative reguation of biological process (107). [2] GOBP anatomical structure development (174), [3] GOBP cell proliferation (73). [4] GOCC plasma membrane (230). [5] GOBP cell migration (32). [6] BIOCARTA Oxidative stress induced gene expression via Nrf2 (12). [7] GOBP positive regulation of biological process (106).
CCR5	1350	[1] GOCC extracellular region part (65). [2] GOBP multicellular organismal development (142).
CD44R	1368	[1] GOBP biological adhesion (107). [2] GOCC extracellular matrix (65). [3] cytoskeletal protein bindgin (55). [4] KEGG Focal adhesion (37), KEGG ECM (23). [5] GOMF glycosaminoglycan bindgin (20). [6] GOCC collagen (13). [7] GOBP development process (248). [8] GOBP cell motility (46). [9] GOBP ECM organization and biogenesis (13). [10] GOBP negative regulation of cellular process (101). [11] GOMF calcium ion binding (111). [12] GOBP cell development (100). [13] GOCC cytoskeleton (105). [14] GOBP positive regulation of immune response (17)
CSF1R	410	[1] GOCC extracellular matrix (43). [2] GOBP response to external stimulus (48), GOBP inflammatory respinse (29). [4] GOMF cytoskeletal protein binding (38), [5] GOBP cell motility (30). [6] GOMF actin binding (30). [7] GOBP cell commnication (127). [8] GOCC collagen (12). [10] GOBP activation of autoimmune response (13). [11] GOCC plasma membrane (109). [14] protein kinase cascade (22). [17] cell proliferation (33). [19] GOBP cell death (33)
IGNGR1	153	[1] GOCC cytoplasmic vesicle (14). [2] GOCC intracellular part (84).
IL1R1	355	[1] GOCC extracellular matrix (30). [2] GOCC actin cytoskeleton (26), GOBP development process (98). [4] KEGG Focal adhesion (21), [5] GOBP actin filamanent-based process (16). [6] GOBP localization (87). [7] GOBP negative regulation of cellular process (41). [8] GOMF lipase inhibitor activity (4). [14] GOMF enzyme inhibitor activity (17).
VEGFR	231	[1] GOBP cell adhesion (30). [2] GOBP multicellular organismal development (53). [3] GOBP blood vessel development (12). [4] GOBP intracellular signaling cascade (34). [5] GOCC intercellular junction (10), KEGG Tight junction (8). [6] GOBP cell motility (16). [7] GOCC plasma membrane (65)

Table 3.5. Representative functional terms enriched in the neighbourhood of immune system and growth factor receptor hubs in the COPD interaction network.

Hubs	No. of genes	Representative Functional Terms
ALDOA	17	[1] GOBP Glycolysis (4), GOBP glucose catabolic process (4). [2] GOBP muscle system process (4).
DLAT	109	[1] GOCC mitochondrion (57). [2] GOBP generation of precursor metabolites and energy (31). [3] GOBP coenzyme metabolic process (20). [4] GOCC mitochondiral inner membrane (30). [5] GOBP oxidative phospohorylation (13). [6] GOCC mitochondrial matrix (11). [8] GOMF catalytic activity (52).
DLD	233	[1] GOCC mitochondrion (98). [3] GOCC mitochondrial matrix (23). [4] KEGG oxidative phosphorylation (26). [5] GOBP cellular respiration (14), KEGG TCA cycle (8). [6] GOMF carbon-oxygen lyase activity (8). [7] GOMF oxidative phosphorylation (20). [8] GOMFoxidoreductase activity (4), [9] GOMF iron-sulfur cluster binding (6). [10] GOBP carboxylic acid metabolic process (21). [11] GOMF lipoic acid binding (4). [12] GOBP carboxylic acid metabolic process (21). [13] GOBP metabolic process (139). [14] GOBP cellular biosynthetic process (29). [17] KEGG Alanine and aspartate metabolism.
ENO1	217	[1] GOCC extracellular matrix (16). [3] GOBP development process (57). [4] GOMF phospholipase inhibitor activity (4). [5] KEGG EHEC and EPEC (9).
GAPDH	31	[1] GOBP cellular carbohydrate catabolic process (7),KEGG Glycolysis/ Gluconeogenesis (6).
PDHAI	171	[1] GOCC mitochondrion (100). [2] GOBP metabolic process (111). [3] GOBP generation of precursor metabolites and energy (52), KEGG oxidative phosphorylation (34), GOBP ATP synthesis (15). [4] GOBP coenzyme metabolic process (26), KEGG GOBP cellular respiration (16), TCA (8). [5] GOCC mitochondrial matrix (19). [6] GOMF ubiquinol-cytochrome-creductase activity (6). [7] KEGG oxidative phosphorylation (34). [8] GOMF metal cluster binding (8). [9] GOBP metabolic process (111). [10] GOMF hydro-lyase activity (6). [11] GOBP cofactor biosynthetic process (11).
PFKM	21	[1] GOCC contractile muscle thin filament (4), GOBP muscle contraction (4), GOCC cytoplasm (12)
PGK1	52	[1] GOBP cellular carbohydrate catabolic process (7),KEGG Glycolysis/ Gluconeogenesis (6). [3] GOCC mitochondiral membrane (8).
TPI1	67	[1] GOBP alcohol catabolic process (9), GOBP glycolysis (8). [2] GOCC mitochondrion (17). [4] GOBP metabolic process (44). [6] GOBP cellular biosynthetic process (13).

Table 3.6. Representative functional terms enriched in the neighbourhood of central metabolism hubs in the COPD interaction network.

3.3.2 Uncoupling between expression of tissue remodeling and bioenergetics modules is a specific feature of skeletal muscles of COPD patients

In order to assess whether the separation between tissue remodeling and bioenergetics sub-networks observed in the network constructed by integrating all measurements in the training study may be a specific feature of COPD, it was necessary to developed two separate networks representing healthy and diseased muscles (**Figure 3.5**).

Our analysis, further supported by the integration of an additional dataset representing an independent and comparable clinical study (see Methods section for details), confirmed that the separation between cytokine/growth factor receptors and bioenergetics modules is a specific feature of COPD networks (**Figure 3.5A, 3.5B**) (**Figure 3.7**). In contrast, there was extensive overlap between the neighbourhood of receptors and intermediate metabolism in the network constructed from healthy muscle biopsies (**Figure 3.2C, 3.2D and Figure 3.3**).

In order to assess whether uncoupling was a general feature of pathological muscles, network models representing the transcriptional state of dystrophy (**Figure 3.8A and 3.8B**) and diabetic muscles (**Figure 3.8C and 3.8D**) were also developed. In both cases the cytokine/growth factor receptor and bioenergetics sub-networks were localized in close proximity. It was remarkable that even in a situation where muscle functionality is severely impaired by the lack of important structural components, a condition typical of muscle dystrophy, energy and tissue remodeling functions still remain co-ordinately regulated. This observation further strengthens the concept that uncoupling between muscle remodeling pathways and bioenergetics is a specific feature of COPD muscles.

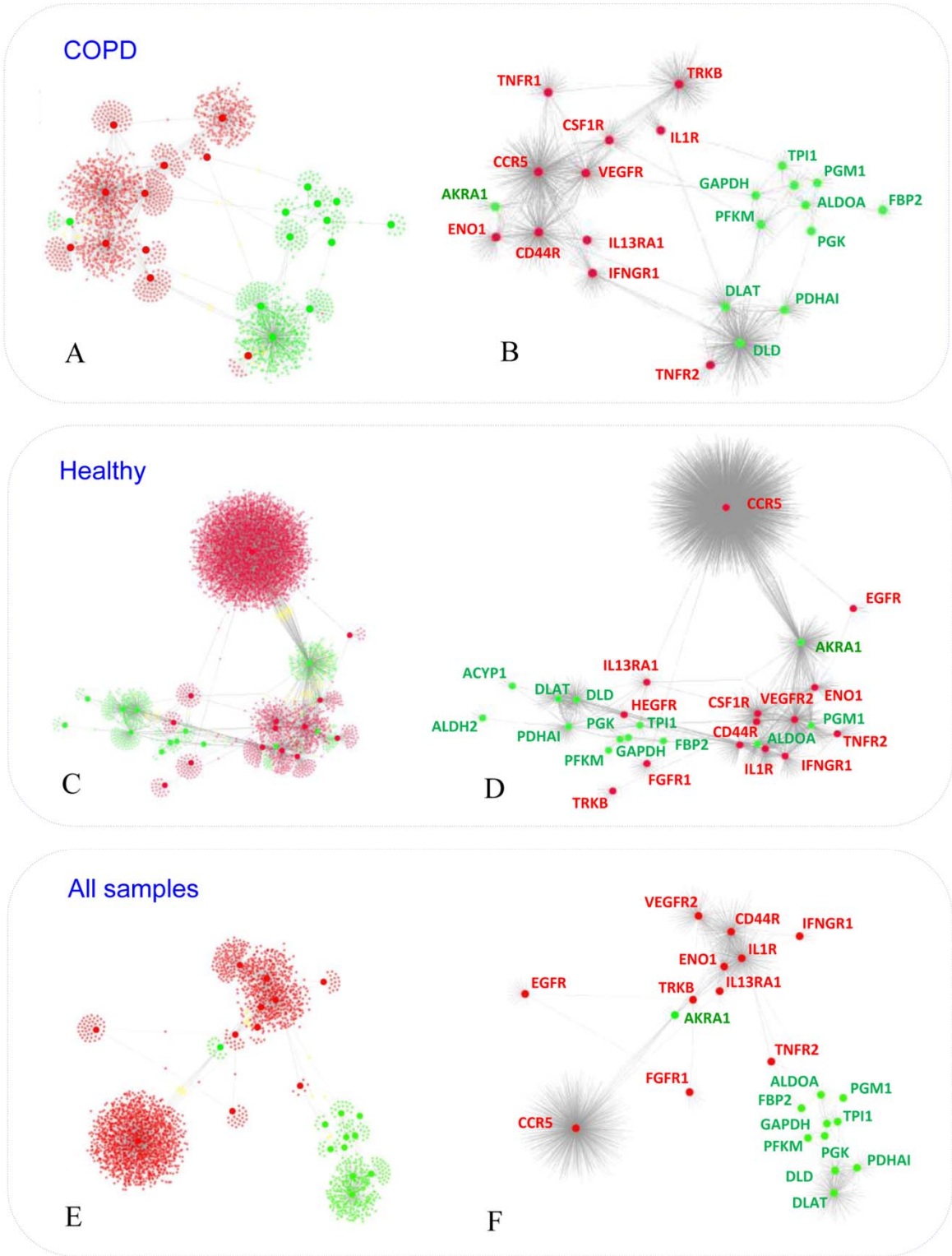


Figure 3-5. Healthy and COPD specific networks.

The figure shows interaction networks built using solely expression profiling data representing healthy (panels A and B) or COPD (panels C and D) muscle biopsy mRNA profiles. For comparison with Figure 3.4, the network representing integrated healthy and COPD samples has been represented in Panels E and F. Figure layout and the colour coding are identical to Figure 3.4.

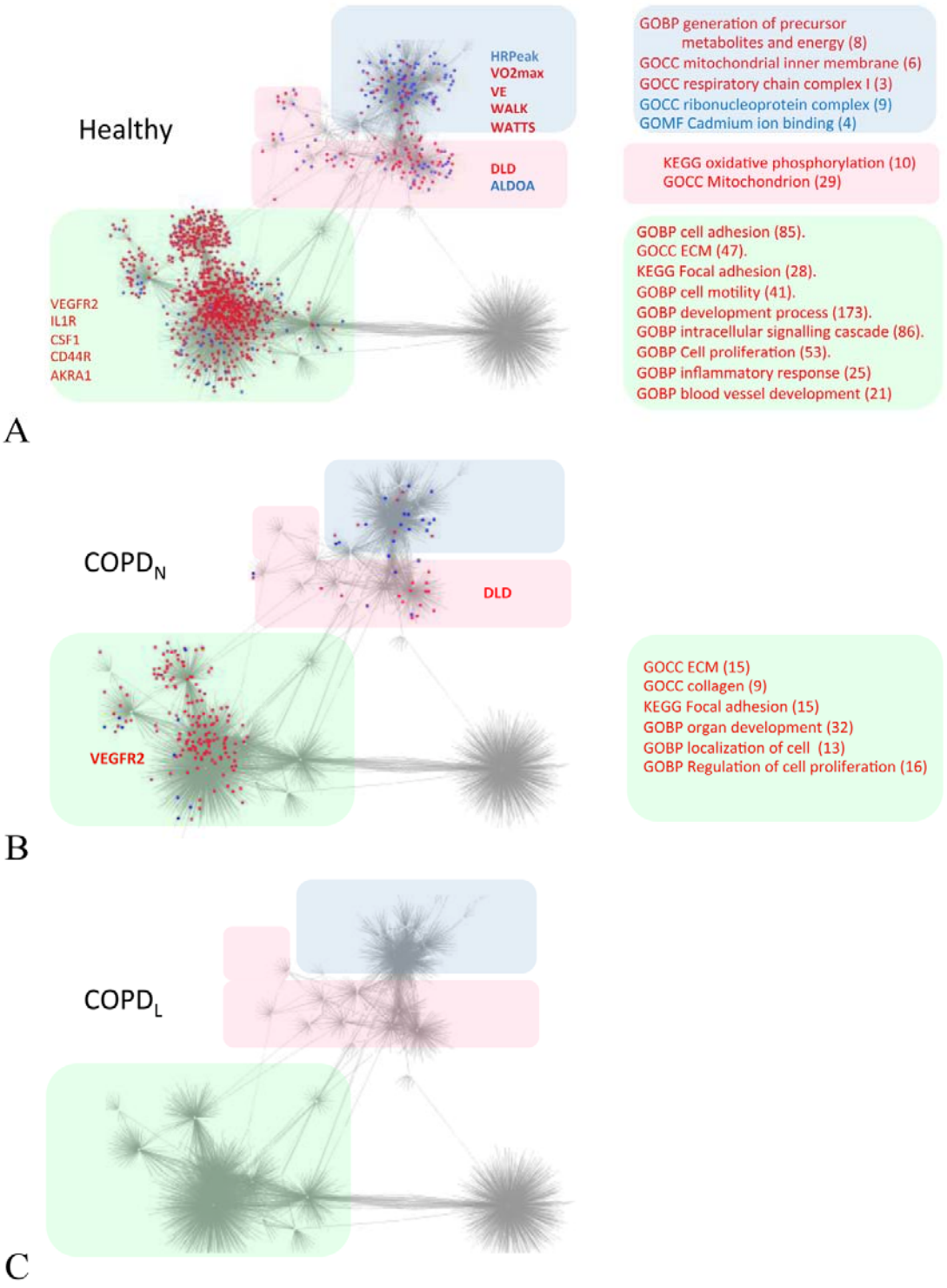


Figure 3-6. Training response of healthy and COPD individuals visualized on the integrated 8 weeks training network.

The networks represent the genes up-regulated (marked in red) and down-regulated (marked in blue) mapped on to the interaction map inferred using the 8 weeks training multi-measurement dataset (shown in detail in Figure 1). Panel A shows genes differentially regulated after training in healthy individuals. Panel B shows genes differentially expressed after training in COPD patients with normal BMI. Network hubs representing genes up-regulated (red) or down-regulated (blue) are listed on the side of the relevant sub-networks. The drastic reduction in the molecular response to training associated with the cytokine and bioenergetics clusters is evident.

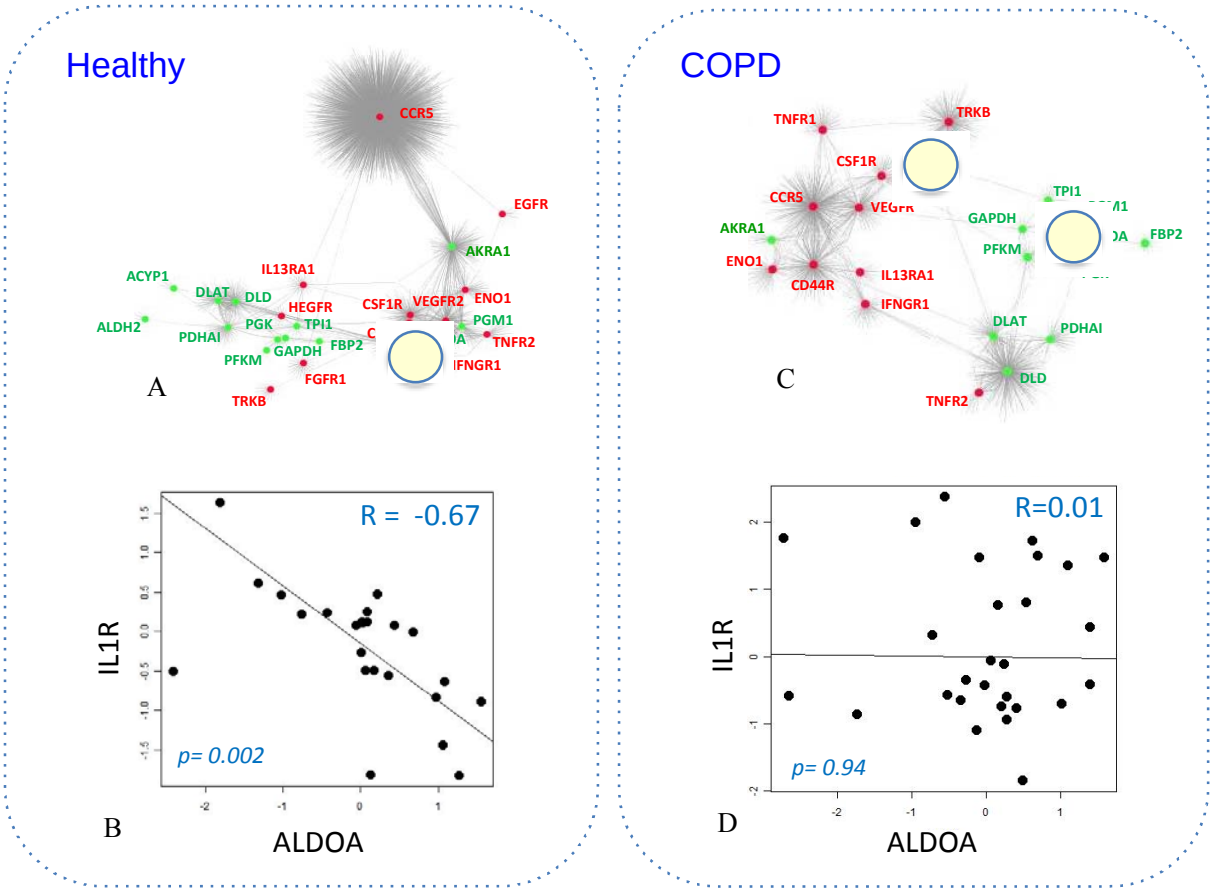


Figure 3-7. Healthy and COPD specific networks.

The figure shows the networks obtained by ARACNE analysis with healthy (panel A) and COPD (panel C) specific datasets (see methods section for details). Panel B and Panel D show the correlation between two landmark genes (IL1R and ALDOA) in the 8 weeks training study for healthy and disease sample biopsies respectively. Note that the network models represent the lack of correlation between tissue remodeling and bioenergetics in COPD patients.

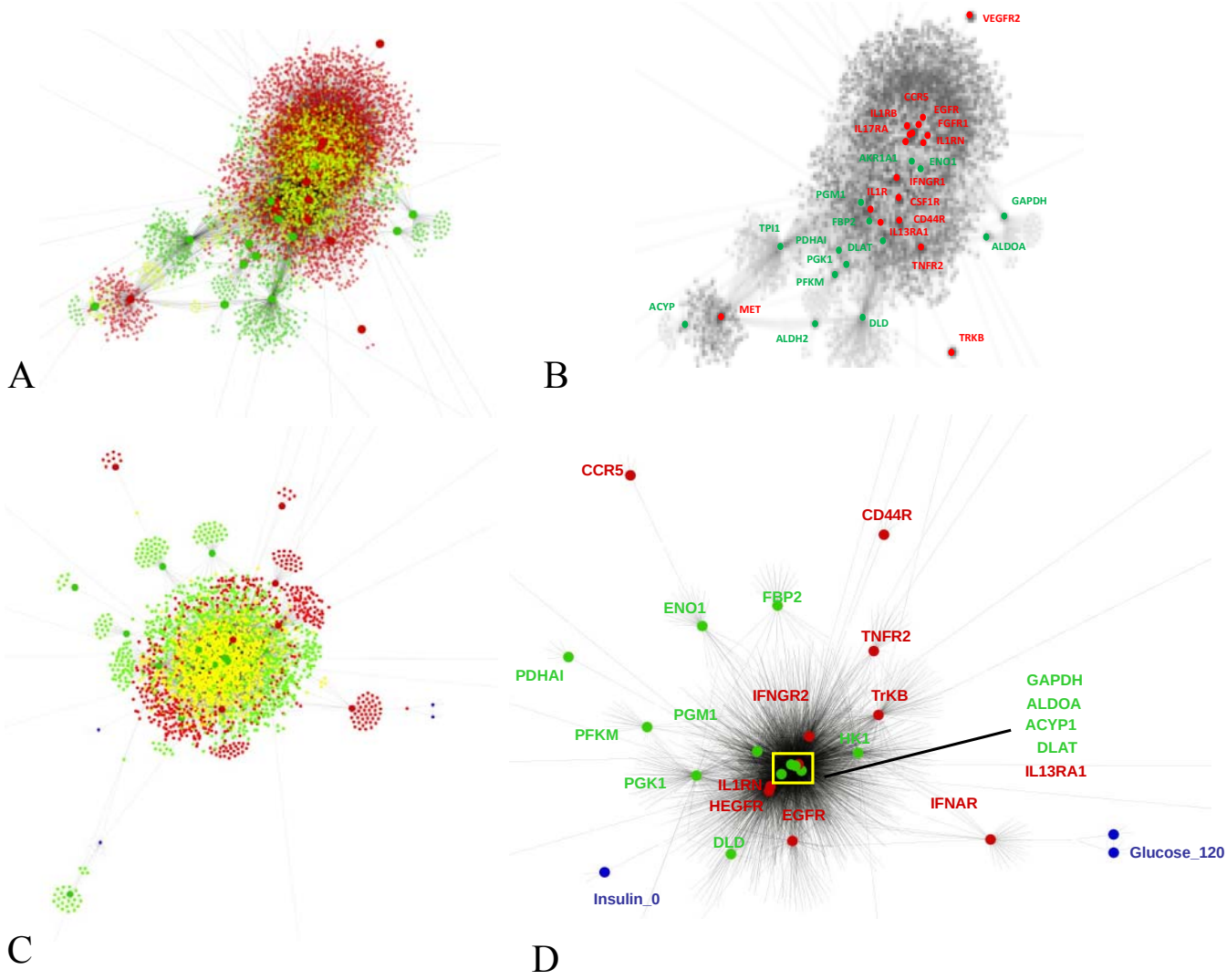


Figure 3-8. Networks representing diabetes and dystrophy muscles.

Panel A and Panel B represent the network inferred from the dystrophy dataset (see methods for details). Panel C and D represent the full interaction networks inferred from the diabetes dataset (see methods section for details). Nodes marked in red represent the neighborhood of receptors of the immune receptors; green nodes represent the neighborhood of central metabolism hubs. From a simple visual inspection it is evident in the two networks that the intermediate metabolism genes are closely linked to the expression of inflammation and growth factor signals.

3.3.3 The molecular response to endurance training highlights a failure of diseased muscles to transcriptionally regulate tissue remodeling and bioenergetics

Healthy individuals responded to training by modulating expression of a relatively large number of genes (3908, **Figure 3.9A**) compared to $COPD_N$ (953, **Figure 3.9B**) and to

$COPD_L$ patients (6 genes). There was a significant overlap (36%, FDR < 10%) at the gene level and an almost complete overlap at the pathway level (**Figure 3.9**) between the response to training in $COPD_N$ and healthy subjects whereas none of the six genes differentially regulated in the $COPD_L$ group were modulated in healthy and $COPD_N$ individuals (**Figure 3.9C**).

In order to identify which part of the network was modulated in response to training, these differentially expressed genes were mapped onto the integrated network described previously (**Figure 3.4**). Genes that were up-regulated after training appear to cluster primarily in proximity to receptors for the cytokines IL-1 (*IL1R*), CSF1 (*CSF1R*), plasminogen (*ENO1*) and the key receptor for the pro-angiogenesis growth factor VEGF (*VEGFR2*) (**Figure 3.6**). A second cluster that is also predominantly populated by genes up-regulated in response to training represents the neighbourhoods of glycolytic enzymes (**Figure 3.6A**) and to the epidermal growth factor receptor (*EGFR*). In contrast, genes that are either up or down-regulated in response to training populated the neighbourhood of physiological variables, representing energy (up regulated) and mRNA processing (down-regulated) functions. The transcriptional response of $COPD_N$ muscles to training followed a very similar pattern although as reported in **Figure 3.7C**, the number of genes involved was much lower (**Figure 3.6B**). As mentioned previously, $COPD_L$ patients do not respond to training with a detectable change in the transcriptional state of the muscle (**Figure 3.6C**). In addition to the transcriptional uncoupling described above, the transition between healthy and diseased individuals is therefore characterised by a reduction in the ability to transcriptionally regulate

expression of genes in both tissue remodeling and bioenergetics pathways.

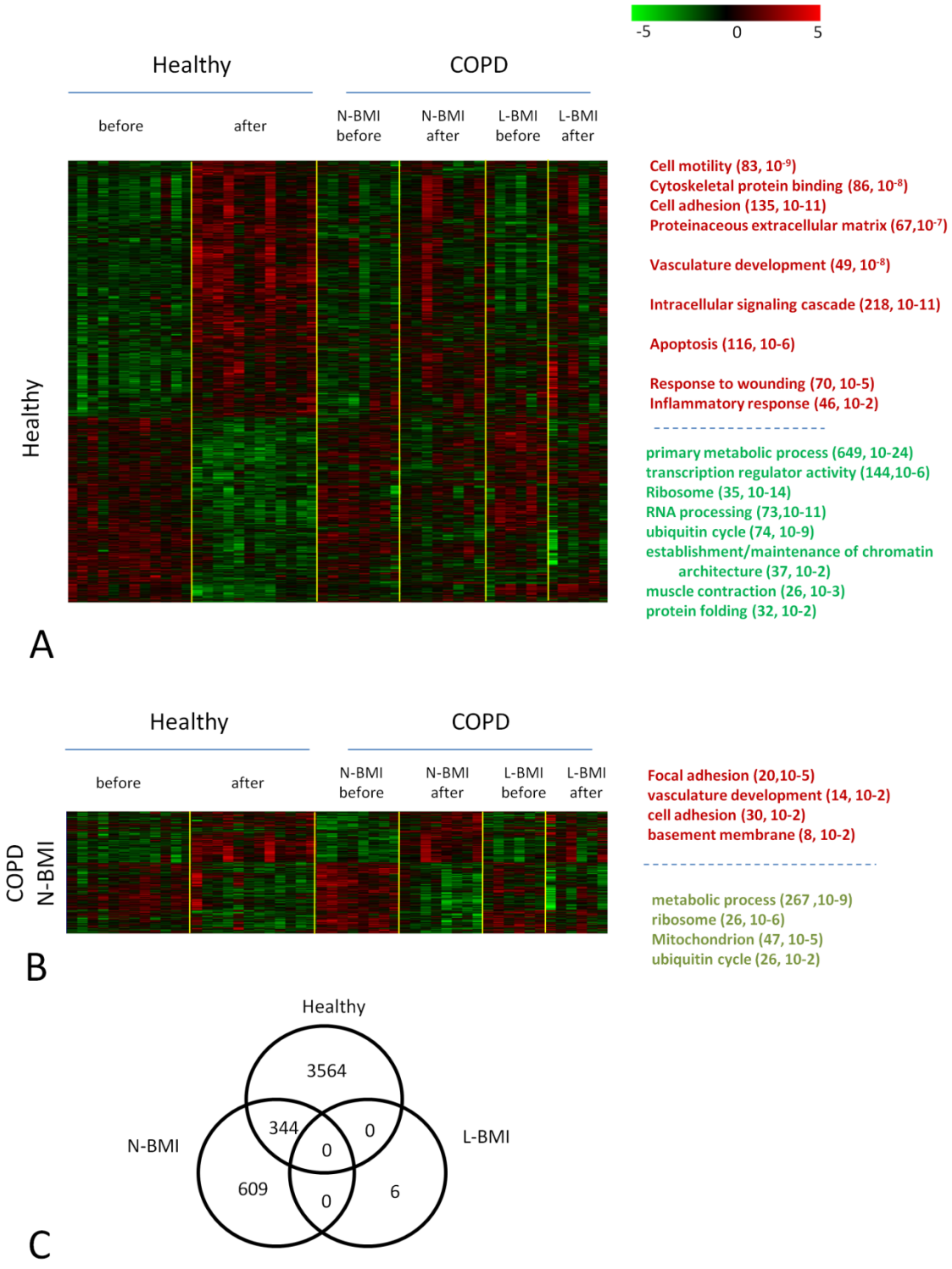


Figure 3-9. Expression profiling analysis of muscle biopsies.

The figure summarizes the results of the microarray analysis of muscle biopsies representing response to endurance training. Panel A shows the heatmap representing genes differentially regulated in response to training in healthy individuals. Panel B shows the heatmap representing genes differentially regulated in response to training in COPD patients with normal Body Mass Index (N-BMI). Functions enriched in the up and down regulated genes are labelled on the right side of the

heatmap and colour coded to discriminate between up-regulated (red) and down-regulated (green) functions. Panel C displays a Venn diagram showing the overlap between differentially regulated genes in response to training in healthy, N-BMI and COPD patients with low Body Mass Index (L-BMI) subjects. It is evident that N-BMI and L-BMI show a progressively less pronounced response to training.

To better characterize molecular networks associated with the growth factor and cytokines components of the network (**see Figure 3.4 and Figure 3.5**) directly connected to receptor hubs and differentially expressed in response to training in healthy individuals (predominantly located in the lower part of the network), these genes were selected and used as input for the Ingenuity Pathway Analysis (IPA) software. The analysis identified 17 networks (see method section for details of the procedure). Among the most significant findings, two interconnected networks linked the transcription of several cytokines (IL1, TNFα, IFNγ, CCL2) and activation of the NF-kB complex (**Figure 3.10A**), leading to the activation of several NF-kB targets related to connective tissue formation (**Figure 3.10B**). **Figure 3.11A** shows the relationship between upregulation of gap junction complexes (JAM, JAM2, JAM3 and TJP1) and activation of important structural components of muscle fibres (e.g. tropomyosin). The network in **Figure 3.11B** shows the important functional link between activation of several Rho GTPases and muscle development, represented by genes encoding for a component of the hexameric ATPase cellular motor protein myosin (MYL5) and (MYH10).

Although not all captured by the network model, 6 myogenic markers were found to be regulated in response to training (**Table 3.7**). Five of these were up-regulated. Consistent with what was previously observed (above), the training-induced expression of all the genes represented in these networks was defective in COPD individuals (**Table 3.7**).

Genes	CS/CT	FDR	NS/NT	FDR	WS/WT

MYO1B	2.424319	0.003434	2.358684	0.043488	NS
MYH10	1.9779	0.010815	NS	NS	NS
MYH9	1.769668	0.01157	NS	NS	NS
MYO15B	1.680785	0.053249	NS	NS	NS
MYLK2	-1.73438	0.060255	NS	NS	NS
MYO6	1.680994	0.060255	NS	NS	NS

Table 3.7. MYOD targets which are significantly different in response to training in healthy patients (FDR 10%, fold> 1.5).

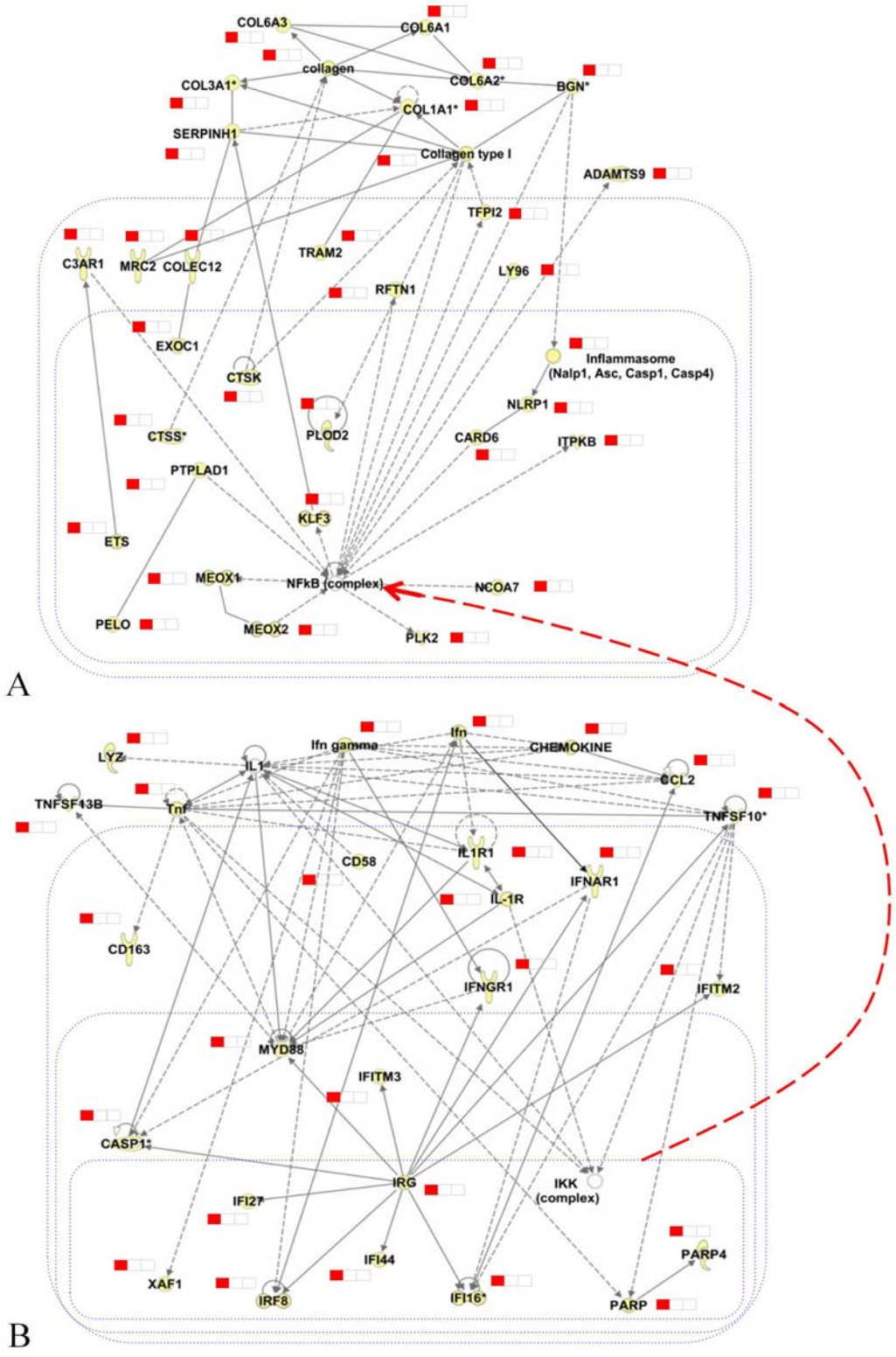


Figure 3-10. Ingenuity Networks Regulated by Endurance Training.

This figure shows the most important findings derived from Ingenuity analysis using genes that are differentially expressed in response to training in healthy individuals and connected to receptor hubs in the previously inferred network. Panel A and Panel B represent two interconnected networks linked the transcription of several cytokines (IL1, TNFα, IFNγ, CCL2) and activation of the NF-kB complex (Panel A), leading to the activation of several NF-kB targets related to connective tissue formation (Panel B).

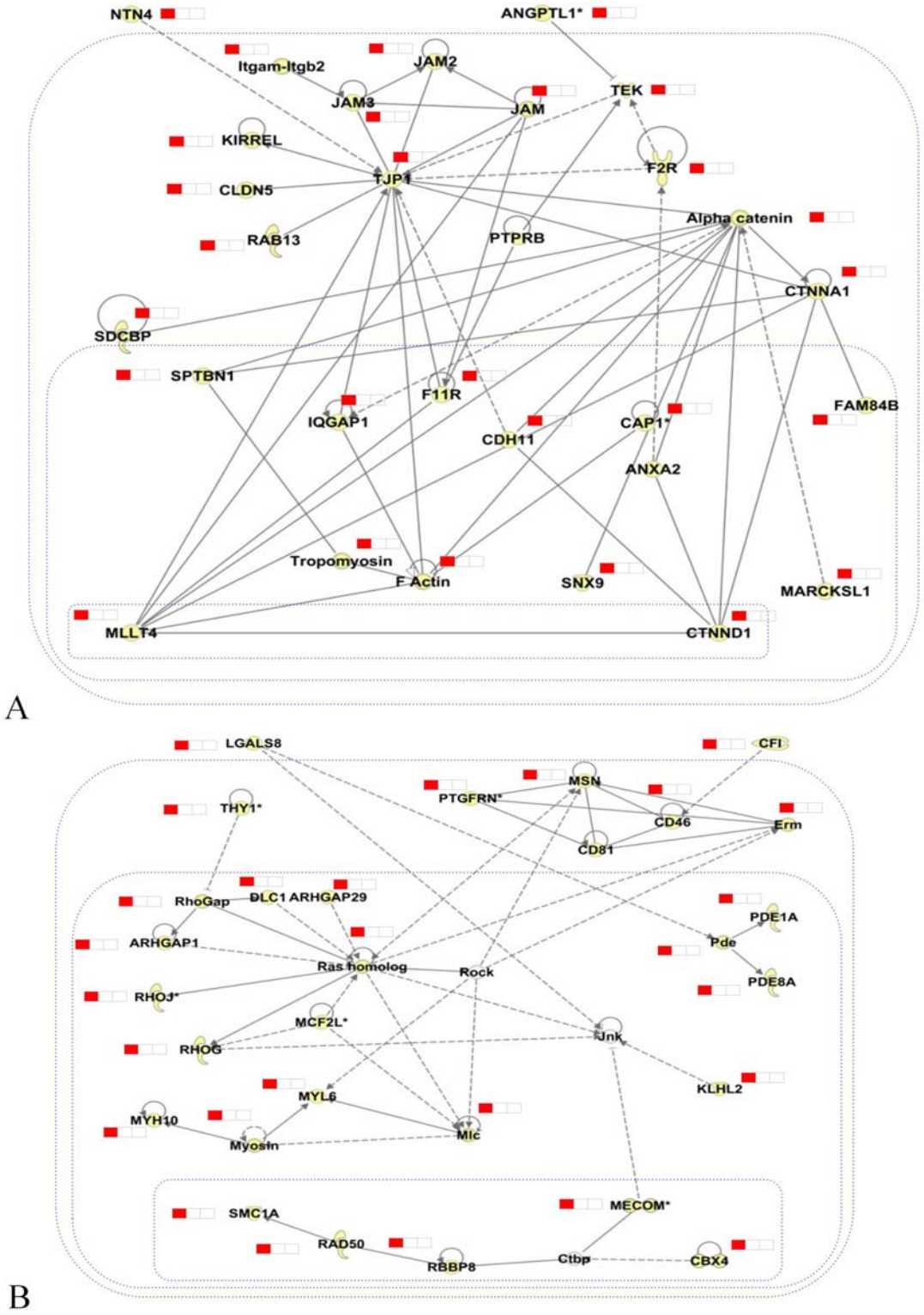


Figure 3-11. Ingenuity Networks Regulated by Endurance Training.

Panel A shows the relationship between upregulation of gap junction complexes (JAM, JAM2, JAM3 and TJP1) and activation of important structural components of muscle fibres (e.g. tropomyosin). Panel B shows the important functional link between activation of several Rho GTPases and muscle development, represented by genes encoding for a component of the hexameric ATPase cellular motor protein myosin (MYL5) and MYH10.

3.3.4 Interleukin-1 promotes activation of tissue remodeling *in vivo* and mirrors the effect of training in healthy human subjects

Because of the strong correlation between expression of IL-1 receptor and genes involved in muscle remodeling and the networks identified by the Ingenuity analysis, IL-1 might be responsible for inducing a significant component of the transcriptional response to training observed in healthy individuals. In order to test this hypothesis, the transcriptional response to intra-venous injections of IL-1β in mice was analyzed in both glycolytic and oxidative muscles. The response of mice to injection of IL1β supported our initial hypothesis. IL-1β treatment induced within 24 hours the up-regulation of 336 genes in oxidative muscle, whereas in glycolytic muscles it induced the up-regulation of 263 genes and the down-regulation of 201 genes (**Table 3.8**). Genes up-regulated in both muscle fibre types were enriched for structural components of the sarcomere (*ACTA1, NEB, CRYAB, ANKRD2*). Apart from this, the response of the two muscle fibre types was very different. Oxidative muscles up-regulated genes encoding for components of the extracellular matrix and organ development including angiogenesis (*SOX18, VEGFA, SERPINF1, ANXA2, PTEN, ENG*), whereas glycolytic muscles modulated genes involved in oxidative phosphorylation (*COX8A, COX5A, NDUFB8, NDUFB5, NDUFS4, NDUFV2*) and ribosomal components (Ribosomal protein chains 10A,s5 and I37A, and mitochondrion ribosomal protein chains 35 and 45).

The high percentage of differentially expressed genes encoding for tissue remodeling functions (extracellular matrix components, organ structure development, cell communication) and bioenergetics (oxidative phorphorylation) which included the oxygen sensitive regulator *HIF-1*, suggested that *IL-1β* has the potential to reproduce a component of the physiological transcriptional response to training that is defective in COPD patients.

In order to assess this hypothesis, genes differentially expressed were mapped in the IL-1β mouse experiment on the human interaction network represented in **Figure 3.4**. 78% and 68

% (265 and 317 genes) of the genes modulated in response to IL-1β treatment in oxidative muscle and glycolytic muscle, respectively, in mice mapped on to the human dataset. Of these, 34 genes were up-regulated after endurance training in healthy individuals (**Figure 3.11**). A significant proportion of these genes were, in oxidative muscle, also directly connected with the receptor for Interleukin 1 in the integrated network model. This observation further validates the model and supports a role for IL-1β in the physiological response to training.

Tissue type	↑↓	#	**Functional annotation**
Oxidative muscle	up	257	**GOCC ECM (64). GOCC sarcomere (7).** GOCC lysosome (8). GOBP anatomical dev structure (41). GOBP angiogenesis (6). GOMF oxidoreductase activity (18) .
Glycolitic muscle	up	170	**GOCC myosin complex (5). GOCC sarcomere (6).** GOBP generation of precursor metabolites and energy (13). KEGG oxidative phos. (8) GOCC mit (20). GOBP gluconeogenesis (5).
Glycolitic musle	down	150	GOCC extracellular space (23). GOBP protein transport (10). GOBP metabolic process (60). GOCC Mit (4)

Table 3.8. Functional annotation of genes changing in response to IL1 injection in oxidative and glycolityc muscle. These functions are not necessarily significantly enriched.

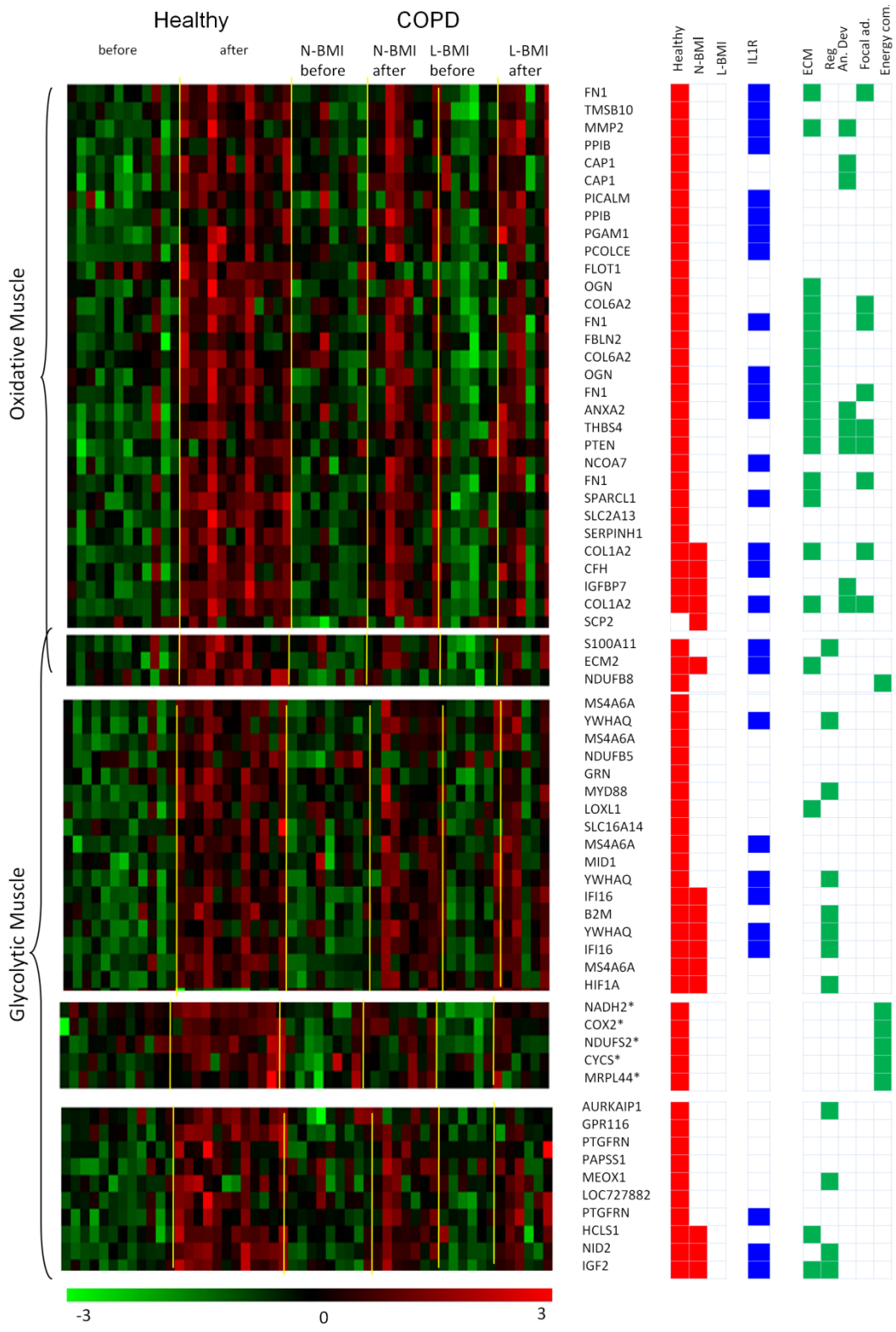


Figure 3-12.Effects of IL-1β in mouse glycolytic and oxidative muscles.

This figure describes the effects of recombinant interleukin 1 on the transcriptional state of glycolytic and oxidative muscles. Panel A represents the heatmap of expression profiles in human muscle biopsies of genes that are differentially regulated in response to training in humans and also induced by IL1β in mouse muscles. Panel B shows the localisation of genes up-regulated in oxidative and glycolytic muscles on the COPD interaction network represented in Figure 3.4.

3.3.5 Decreased training-induced expression of NF-kB target genes in COPD muscles

The lack of activation of tissue remodeling functions in COPD muscles, including the component linked to IL-1β signaling, may be a consequence of the inactivation of myogenic pathways due to over-activation of NF-kB signaling induced by chronic exposure of pro-inflammatory cytokines [16]. If this hypothesis were correct, over expression of a number of NF-kB target genes in diseased muscles should be observed.

A number of NF-kB targets, identified by the IPA software in the 'tissue remodeling' section of the network, were up-regulated during training in healthy individuals, but failed to respond in COPD patients (**Figure 3.10A**). Although interesting, this observation was based on a loose definition of NF-kB targets, as many connections reported in the Ingenuity database are indirect. I therefore further tested the NF-kB over-activation hypothesis by analyzing expression of 94 experimentally validated targets of NF-Kb known in the literature. Against the working hypothesis, there were not any NF-kB target genes differentially regulated between normal and diseased muscles. On the contrary, 8 genes were identified as differentially regulated in response to training in healthy individuals (all up regulated, 10%FDR and fold>1.5) but not in COPD individuals with normal or low BMI (**Table 3.9**). These observations clearly do not support the over-activation hypothesis. On the contrary, they suggest that in COPD muscles training-induced activation of NF-kB is repressed.

genes	FOLD CS-CT	FDR
TFPI2	1.88	0.02
PLAU	1.62	0.02
CD74	1.60	0.03
CCL2	1.69	0.04
ENG	1.41	0.04
CCND1	1.45	0.07
NQO1	1.69	0.08
CD44	1.80	0.09

Table 3.9.NFKB Targets in response to training in healthy muscle.

NFKB Targets which are significantly different in response to training in healthy patients.

3.3.6 Expression of epigenetic histone modifiers discriminates between healthy and diseased muscles and is correlated with peak oxygen consumption

In order to explore an alternative hypothesis that may explain muscle wasting in COPD patients, an unbiased analysis were performed to identify functional pathways differentially modulated between healthy and diseased muscles, by using two-factor ANOVA (taking into consideration both disease and training).

Consistent with the analysis described above, tissue remodeling and energy-associated pathways were identified as significantly differentially expressed for both factors analysed (**Table 3.12**). However, one additional category of 20 chromatin modification enzymes were also differentially expressed (**Table 3.11 and Table 3.10**).

Gene Symbol	Fold Disease vs. Healthy	ANOVA Factor Disease (FDR)
HDAC9	2.30	6.61e-5
HDAC4	-1.48	0.05
SIRT2	1.27	0.0016
SIRT3	1.31	0.02

Table 3.10. Chromatin remodeling complex genes in healthy and disease muscle.

Among these, four histone deacetylase enzymes were identified. These are known to be particularly relevant for controlling expression of muscle differentiation (*HDAC9* and *HDAC4, SIRT2*) and bioenergetics (*SIRT3*) related genes. Consistent with their potential role in COPD muscle wasting, expression of these genes was sufficient to discriminate between diseased and healthy muscles (**Figure 3.13A**). Mapping these on the network model described in **Figure 3.4** show that they localized close to VO_2peak, (**Figure 3.13B**) suggesting that their transcription may be up-regulated by oxygen availability and/or oxidative capacity in the skeletal muscle.

Functional Category	Genes in the Function (10% FDR)
GOCC Chromatin remodeling complex (disease factor)	APPL1, ARID1A, ARID1B, CHD4, CIR1, HDAC4, HDAC7, HDAC9, MBD2, MTA1, PHF21A, RSF1, SIN3A, SIRT2, SIRT3, SMARCA4, SMARCB1, SNARCE1, SUDS3, TBL1XR1
GOCC histone deacetylase complex (exercise factor)	APPL2, CHD3, CHD4, GATAD2A, HDAC1, HDAC11, HDAC3, HDAC4, HDAC5, LOC642954, MBD3, NRIP1, RBBP4, RBBP7, RERE, SAP18, TAL1, TBL1XR1

Table 3.11. Functional annotation of genes identified by two-factor ANOVA.

Functional category	#of genes	Factor
GOBP RNA processing	121	Disease
GOBP ribonucleoprotein complex biogeneis	53	Disease
GOCC mitochondrion	198	Disease
GOCC contractile fiber	33	Disease
KEGG Tight junction	35	Disease
GOCC Chromatin remodeling complex	19	Disease
GOCC histone deacytylase complex	12	Disease
GOMF RNA binding	245	Exercise

GOBP proteolysis involved in cellular protein catabolic process	206	Exercise
GOMF enzyme binding	182	Exercise
GOCC ribonucleoprotein complex	165	Exercise
GOBP protein complex biogenesis	166	Exercise
GOCC extracellular matrix part	50	Exercise
GOBP blood vessel development	89	Exercise
GOBP regulation of cell proliferation	236	Exercise
GOCC contractile fiber	50	Exercise
KEGG ECM-receptor interaction	36	Exercise
GOCC histone deacetylase complex	17	Exercise
GOCC ribonucleoprotein complex	29	Interaction
GOMF GTPase activity	14	Interaction
GOBP RNA processing	21	Interaction
GOCC mitochondrial part	20	Interaction

Table 3.12. Functional annotation of genes identified by two-factor ANOVA.

3.3.7 Genes predicted to be linked to VO2max ARACNE are induced in a mouse model of chronic hypoxia

In order to elucidate whether genes linked to VO2max in the network model are part of the physiological response to hypoxic conditions, I analyzed a public domain microarray study developed by Budak *et al.* (GEO: GSE9400) [185] which represents the response of murine skeletal muscles to 2 weeks hypoxic conditions.

In this study, 45% of genes connected to VO2max in the network model were transcriptionally regulated in the mouse model of hypoxia and that a striking 82% of these were regulated in the direction predicted by the network analysis (**Figure 3.14**). Among these, the genes encoding for the chromatin modifiers *HDAC4* and *SIRT3* were present. *HDAC4* is up-regulated, whereas *SIRT3* is down-regulated in hypoxic mice compare to normal (**Figure 3.14**), which is consistent with the sign of the correlation with VO2max observed in the clinical study (**Figure 3.13A and 3.13C**). Interestingly, many other genes encoding for chromatin modifiers were differentially expressed in response to hypoxia.

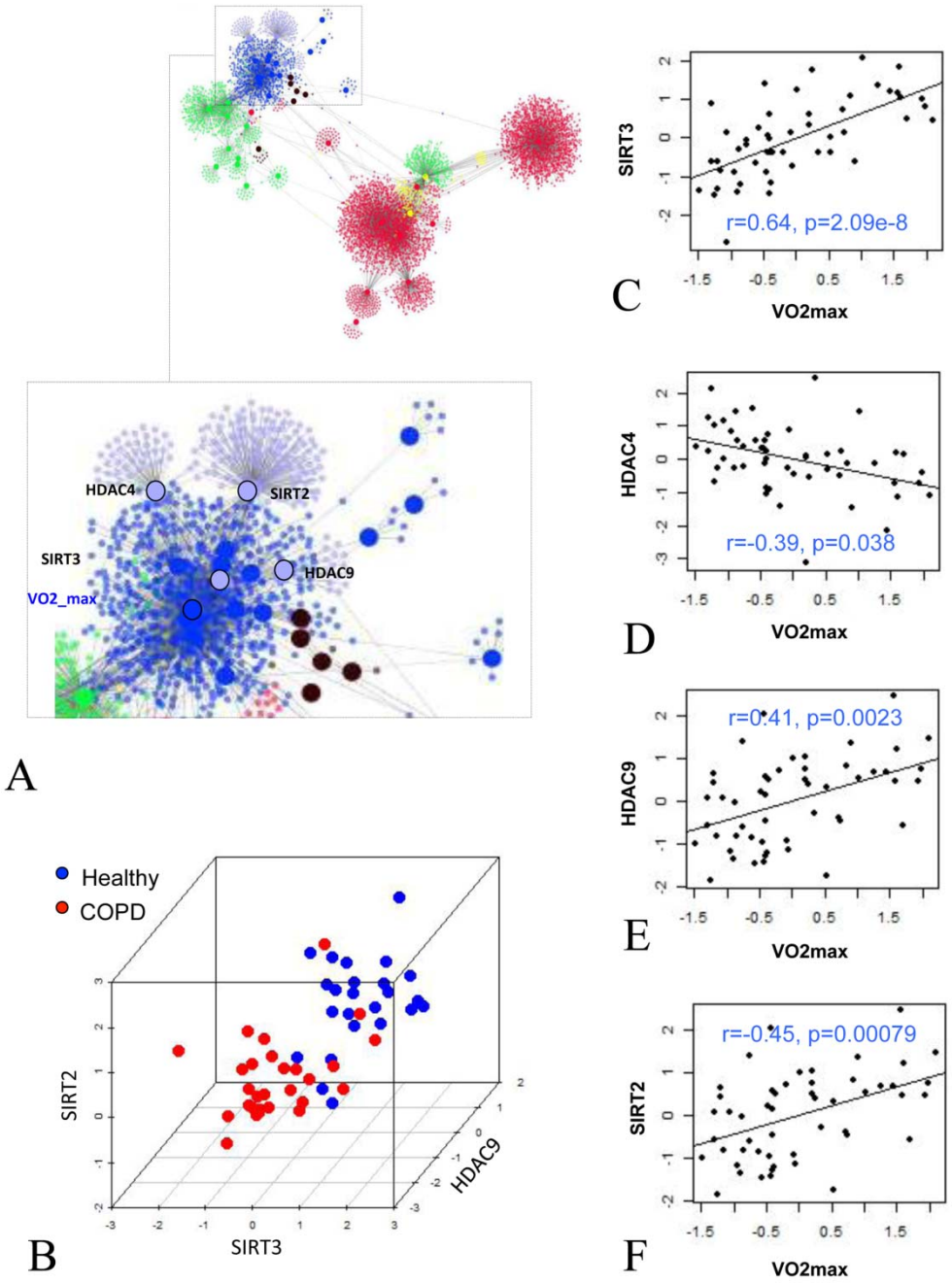


Figure 3-13. Expression of Chromatin Modification Enzymes in COPD and Healthy Muscles.

Panel A represent the close localization of histone deacetylase enzymes to VO_2max in the inferred network. Panel B is a 3D plot representing the expression of histone deacetylase enzymes and the separation between disease and healthy muscles. Panel C, D, E and F represent the scatterplots between VO_2max and the expression of histone deacetylase enzymes.

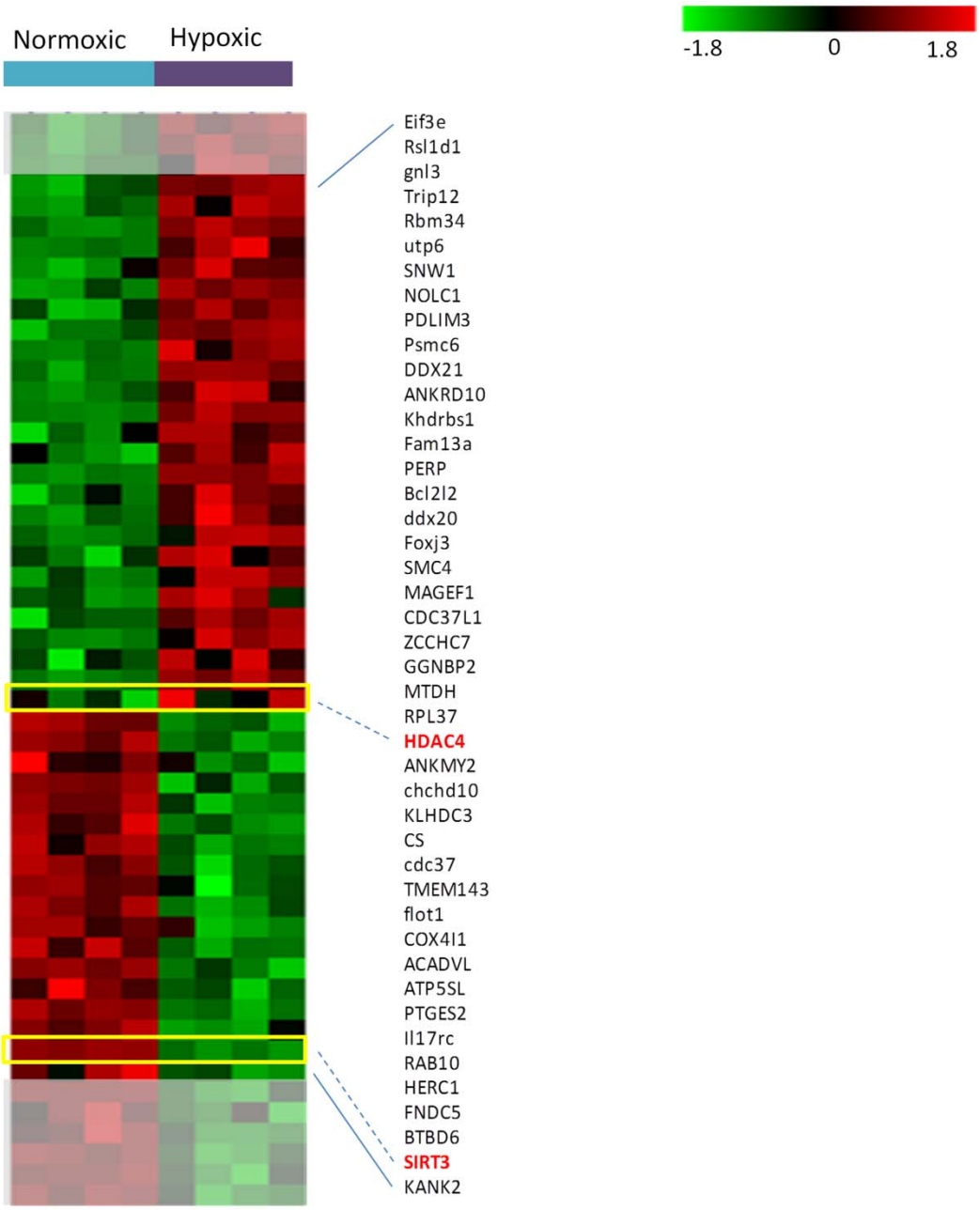


Figure 3-14. Genes represented in the neighborhood of VO2max are transcriptionally regulated in a mouse model of hypoxia.

This heatmap represent the expression of genes in the neighborhood of VO2max that are also differentially expressed in skeletal muscles, in a mouse model of hypoxia. It shows that the majority of these genes (82%) are regulated in the direction predicted by the network (Figure 3.4). The HDAC4 and SIRT3 genes are among these and are highlighted in red. Genes that are not modulated in the direction as predicted are shaded in grey.

3.4 DISCUSSION

The networks developed in this thesis represent the first model linking molecular and physiology measurements in skeletal muscle of COPD patients. It provides convincing evidence that a failure to co-ordinately activate expression of several tissue remodeling and bioenergetic pathways is a specific landmark of diseased muscles. Moreover, the model inferred in this study is consistent with the view that the abnormal expression of a number of histone modifiers potentially regulated by oxygen availability may be responsible for alterations in both tissue remodeling and bioenergetic functions. This hypothesis has important implications as it places cell hypoxia and oxidative capacity as the main drivers for skeletal muscle abnormalities in COPD patients

3.4.1 Response to training in healthy and diseased muscles

Several pathways were identified as transcriptionally regulated in healthy individuals in response to training. These involve the up-regulation of several tissue remodeling genes/pathways (Plasminogen receptor, *VEGF*, pro-inflammatory signals such as *IL1*) as well as modulation of energy and ribosome biogenesis functions. Without exception, the exercise-induced modulation of these pathways is severely impaired in COPD individuals.

The Ingenuity pathway analysis has revealed several detailed mechanisms associated with the inflammation and growth factor receptor component of the network presented in **Figure 3.4**. Two of these networks (**Figure 3.10A and 3.11A**) are part of a larger network linking the effect of pro-inflammatory signals such as IL-1, TNFα and IFNγ to the training induced up-regulation of several components of the extracellular matrix in healthy subjects. Although circulating concentration of IL-1β are largely unaffected by exercise there is evidence of increased local IL-1β levels within skeletal muscle, likely in response to micro-injury of skeletal muscle with increased activity [187]. The analysis of the *in vivo* experiment shows

that indeed a component of the transcriptional response observed in healthy individuals and defective in COPD patients may be mediated by interleukin-1β. Moreover, mouse glycolytic and oxidative muscles respond similarly to IL-1β in respect to the up-regulation of structural components of the muscle but diverge in the regulation of genes involved in energy metabolism and ubiquitination (up-regulated in glycolytic muscle), or extracellular matrix and tissue remodeling including angiogenesis (up-regulated in oxidative muscles).

Patients with mild to moderate COPD have a greater proportion of fatigue-susceptible anaerobic (glycolytic Type II) relative to fatigue-resistant aerobic (oxidative Type I) fibres, suggesting a slow-to-fast transition. It has been proposed that <27% Type I and >29% Type IIx fibres offers a pathological threshold for COPD [188]. This is consistent with a shift towards a more glycolytic enzyme profile, and would contribute to an increase in skeletal muscle fatigability [189]. Although individual fibre phenotypes are well conserved among mammals virtually all human skeletal muscles are of mixed composition, hence mouse muscles with discrete metabolic profiles were used to identify differential responses to IL-1β according to metabolic type.

A mechanistic link between the activity of these tissue remodeling pathways and myogenesis is well supported by the current literature. For instance, the plasminogen receptor ENO1 has been demonstrated to be an important component of skeletal myogenesis by concentrating and enhancing plasmin generation of the cell surface [190]. Interestingly, ENO1 KO mice show severe defects in muscle regeneration following injury [190]. Components of the extracellular matrix, which are induced by signaling from many of the receptors present in the network are also mechanistically linked to myopathies. For example, Col6a1–deficient and Col15a1-deficient mice have a muscle phenotype that strongly resembles human myopathies [191, 192].

Inflammatory and chemo-attractant mediators are also known to be key factors in driving muscle remodeling in the normal physiological response to training [193] and in response to trauma [190]. In the latter, IL1β also promote phagocytosis of trauma-induced cellular debris by macrophages, which themselves can continue to secrete this cytokine up to 5 days post-injury [187]. Ccr2 null mice with cardiotoxin induced injury has been shown to have a delayed angiogenesis and Vegf production compared to wild type mice with muscle fibre size increase observed only after restoration of tissue Vegf [194]. Other studies have also shown the crucial role of VEGF in angiogenesis and in muscle regeneration [195, 196].

There is also strong evidence that bioenergetics and tissue remodeling pathways are linked [197]. It has been shown that genes coding for components of collagen V and collagen VI play an important regulatory role in ECM maturation, where reduced expression promotes apoptosis, mitochondrial dysfunction and muscle degeneration [197]. Other studies have shown that bioenergetics knockouts, such as H6PD null mice is responsible for inducing severe skeletal myopathy by altering sarcoplasmic reticulum redox state [198].

3.4.2 Decreased training induced expression of NF-kB target genes in COPD muscles

A number of *in vitro* and *in vivo* studies have been used in the past to support the hypothesis that a systemic inflammation-driven mechanism leads to inactivation of myogenic pathways in COPD muscles. The analysis presented in this chapter is the first attempt to challenge this hypothesis using genome-wide data, and in a clinically relevant setting. There was no evidence of over-expression of NF-kB target genes in COPD muscles. On the contrary, a subset of direct (**Table 3.9**) and indirect targets (**Figure 3.10A**) of NF-kB were up-regulated in healthy individuals in response to training but not in COPD patients ($p < 0.05$). These

results are consistent with a recent observation demonstrating failure to activate NF-kB in response to acute training in a small subset of COPD patients [81]. Taken together, these results suggest that training-associated inactivation rather than over activation of NF-kB may be a feature of diseased muscles.

Additional observations also argue against a primary role of chronic inflammation in muscle wasting. For example, although an increased levels of cytokines in COPD patients were observed (**Figure 3.1B**) with respect to normal individuals, only a few links between the concentration of these cytokines and the muscle transcriptional state were identified. These results imply that systemic inflammatory signals may have a smaller effect on muscle physiology than previously thought. This latest observation is consistent with previous reports showing that TNFα levels measured in COPD muscles were not significantly higher than in healthy muscles [67].

3.4.3 Is there an epigenetic basis for muscle wasting in COPD?

Since the NF-kB over-activation hypothesis is unlikely to explain the inhibition of muscle remodeling observed in COPD, what could be an alternative mechanism for muscle wasting in COPD?

There are several pieces of evidence in favour of the hypothesis that an imbalance in expression of oxygen-correlated chromatin modifying enzymes, which have shown to be a landmark of COPD muscles. This could explain failure to modulate both tissue remodeling and bioenergetic functions in response to training in these patients. **Figure 3.13** shows that the expression of *SIRT2, SIRT3, HDAC4* and *HDAC9* is sufficient to discriminate healthy and diseased muscles. At the individual gene level, healthy muscles are characterized by a higher expression of *HDAC9, SIRT3* and by a lower expression of HDAC4 (**Table 3.10**) with respect to COPD muscles.

The role of HDAC9 and HDAC4 in muscle development is well documented [199]. For example, HDAC9 is a transcriptional repressor involved in feedback control of muscle differentiation, acting in concert with MEF2 to repress activity-induced genes [200], while HDAC4 is up-regulated in pathological conditions such as muscle denervation [201], and it has been described to be a critical regulator of muscle atrophy by activation of E3 ubiquitin ligases [202]. It is possible, therefore, that a lower expression of HDAC9 and a higher expression of HDAC4 in COPD muscles may be linked to a reduced ability to activate muscle remodeling. Abnormal expression of SIRT2 may also contribute to a failure to activate an appropriate muscle remodeling response. SIRT2 is a NAD^+-dependent histone deacetylase that regulates muscle gene expression and differentiation by forming a complex with MyoD [203]. When over-expressed, this retards muscle differentiation. Conversely, cells with decreased SIRT2 differentiate prematurely. Interestingly, the activity of SIRT2 is dependent on the redox state of the cell [203], which showed evidence of being abnormal in COPD muscles [204, 90, 160]. In the network neighborhood of SIRT2, TXN2 was identified. TXN2 is known to play an important role in protection against oxidative stress [205]. Similarly, GAB1 is shown to play a role in oxidative stress signaling [206] and it is identified in the neighborhood of HDAC9 in the network. Since changes in ROS production are known to influence the expression of HDACs [207], the inferred network may represent this important control mechanism.

SIRT3 is a NAD^+-dependent histone deacetylase that may account for the characteristic loss of transcriptional modulation of bioenergetic genes in response to training in COPD muscles. SIRT3 is localized in the mitochondrial matrix, where it regulates the acetylation levels of metabolic enzymes, including acetyl coenzyme A synthetase 2 [208, 209]. Mice lacking both SIRT3 alleles show hyperacetylation of several mitochondrial proteins, associated with decreased levels of fatty-acid oxidation, and display a selective inhibition of electron

transport chain Complex I activity leading to reduction in basal levels of ATP in several organs [210]. These and other data implicate protein acetylation as an important regulator of mitochondrial function *in vivo*, and it is therefore feasible that an altered expression of SIRT3 in the muscles of COPD individuals may contribute to the observed imbalance in mitochondria functionality. It is interesting that SIRT3 in the model is positively correlated with VO_2peak, suggesting that the lower levels of expression of this enzyme observed in COPD muscles may be the consequence of tissue hypoxia. Consistent with this, histone deacetylase (HDAC) inhibitors reduce HIF-1α protein expression leading to down-regulation of VEGF and other angiogenesis-related genes [211], potentially explaining the reciprocal relationship between extent of muscle capillarity and the degree of COPD [212]. The abnormal expression of a relatively small number of histone modifying enzymes could therefore account for a wide spectrum of abnormal responses observed in the muscles of COPD patients, and may also explain the limited efficacy of training as a therapeutic option. This view is supported by our observation that indeed hypoxia induces modulation of a number of chromatin modifiers in a mouse model of chronic hypoxia and that indeed SIRT3 and HDAC4 are among them (**Figure 3.14**).

3.5 Conclusion

In this chapter, the most accurate system level representation of COPD muscles to date was built. Further work is, however, needed to elucidate the precise mechanism for muscle inactivation. If the mechanism for muscle wasting suggested by our observations on HDACs were to be validated in a clinical setting this would open up a very exciting therapeutic avenue. The use of non-toxic histone deacetylase inhibitors such as valproate has already shown promising in treatment of haematological cancer [213, 214, 215], and may help to restore mitochondrial functionality and the ability to activate muscle remodeling in COPD

patients. In this context it is possible that the appropriate pharmacological regime coupled with physical rehabilitation may lead to recovered muscle functionality, and improved quality of life.

CHAPTER 4: THE DEVELOPMENT OF A NOVEL NETWORK MODULARIZATION PROCEDURE FOR MULTI-LEVEL DATA INTEGRATION

4.1 INTRODUCTION

Despite the fact that the networks inferred and discussed in the previous chapter represent thousands of nodes and edges, their interpretation were relatively straightforward. This was in part a consequence of the fact that the networks were built using a biology driven network inference pipeline (see Figure 3.4). Unfortunately, this strategy is not generally applicable. Large-scale networks are in fact usually challenging to interpret and in this context, modularization approaches provide suitable tools for identifying sub-networks linked to important biological functions or disease outcomes. Several modularization approaches have been developed so far [151, 216, 155, 154, 217]. These approaches are either aiming to identify sub-networks whose genes are differentially regulated between various experimental conditions [151], [216] or are highly co-regulated [155, 154]. Ideker *et al*. [151] proposed a heuristic approach aiming to identify modules in protein-protein interaction networks that are enriched with differentially expressed genes. In their paper, Ideker et al. demonstrated the usefulness of this approach by integrating yeast protein interaction with gene expression data measured from 20 perturbations to the yeast galactose utilization pathway. Scott *et al.* addressed the same challenge using a graph theory approach based on Steiner trees [217]. In

their method, they applied an approximation algorithm developed by Klein *et al.* [218]. Dittrich *et al*. has also tackled this problem using Steiner trees, but they have used an algorithm developed by Lubic *et al.* [219], which can solve Steiner Tree problem to optimality. Therefore, this methodology had excellent performances on the simulated data and over-performed in all instances compared to the heuristic based methodology. Other approaches have been developed to identify functional modules, with highly co-regulated nodes [155, 154]. In these methodologies, edges based scoring functions are introduced in order identify modules with high scores, which consist of edges that are highly co-regulated.

However, a methodology integrating such node and edge properties is still missing. If available, such procedure would allow the identification of network modules integrating three different levels of information. For example, protein-protein interaction modules that are enriched of genes differentially expressed between two or more experimental groups and also co-regulated across a large numbers of experimental perturbations (e.g. different genetic backgrounds or across a patient population). Here, this challenge has been addressed by developing a new methodology derived from the modularization method published in Dittrich et al. [216]. On simulated data, this methodology has a high sensitivity and specificity in recovering network modules of different sizes and with varying degrees of association. Its application to integrate mass spectrometry, expression profiling and knowledge datasets in a real world biological system has revealed a new mechanism controlling cell proliferation in human Glioma.

4.2 MATERIAL AND METHODS

4.2.1 The Modularization Approach

The proposed methodology aims to identify sub-networks within a larger network, whose elements share several functional properties. Various properties could be used to construct and score the initial network depending on the question to be addressed. Hence, it allows integrating three distinct levels of information. This is achieved by a scoring function that integrates nodes (first level) and edges (second level) properties in an underlying network structure (third level). In the application case described in this chapter the underlying network is represented by a large gene interaction network derived from a public domain knowledge database Human Protein Reaction Database (HPRD). The node properties are encoded from p-values derived from a mass spectrometry experiment representing molecular interactors of the RhoE protein. Edge properties are derived from the correlation between pairs of gene expression profiles representing the development of a tumour. In this particular case, the aim is to identify sub-networks with potential interactors of RhoE that are highly coregulated during the tumour development.

The problem of identifying network modules has recently been shown that it could be solved using Prize-Collecting Steiner Tree Problem (PCST) [217, 216]. Briefly, in PCST nodes in the network represent profits and edges represent costs. Networks are scored by computing the sum of the profits subtracted by the sum of the cost of its members. The goal is to identify sub-networks with the highest score. Hence, the node scoring function should transform low p-values with higher node profits. Similarly, since highly correlated genes have higher correlation coefficients the edge scoring function should express these gene-to-gene connections with a lower edge cost. The graph scored using such scoring functions can then

be used as an input to the algorithm developed and implemented in [219] in order to find the network with the highest score.

4.2.1.1 Constructing interaction networks

The first step is the construction of the underlying network. The selection of the data source is a crucial step. A number of gene and protein interaction databases have been developed in the past few years. Some of them are publicly available (for a review [220]) and others are accessible in the context of commercial software packages [221]. In this application, the HPRD database has been used [222, 223]. HPRD is the most comprehensive human protein-protein interaction database being developed to date and contains the large majority of interactions (75%) available in other similar databases such as BIND, DIP and BIOGRID [220]. Using the BioNet R package, an interaction network representing 9392 different proteins linked by 36504 literature-curated interactions have been extracted.

4.2.1.2 Scoring the network

Once the interaction network is constructed, the nodes and the edges in the networks are scored as described below.

Node Scores

Each node is scored as a function of a *p-value* derived from a statistical test. This can be the output of any statistical procedure comparing two or more experimental groups. It may also represent a measure of confidence that assign a particular property to a node such as a target gene. The score for each node is defined as:

$$scoreNode_g = -\log(p), \qquad (1)$$

where g is the score of the gene g and p is the associate p-value of the gene g derived from the differential expression analysis. This basic score for each gene allow us to obtain larger scores for smaller p-values.

Edge Scores

Each network edge is scored as a function of a correlation coefficient. This can be either a standard correlation coefficient or mutual information. For smaller datasets a descriptive measure of dependency such as the Spearman correlation coefficient may be more appropriate whereas MI can be more effective with larger sample size (n>40) [147].

The score of the edge e between genes g_i and g_j, is defined as:

$$scoreEdge_e = -\log(I(g_i, g_j)) * w, \quad (2)$$

where, I is the correlation coefficient (or MI) between genes g_i and g_j, and w is a weight. The final scoring function to maximize is dependent on the edge (2) and node score (1), therefore the contribution of the cost (i.e., edge score) can be increased or decreased using the weight parameter w in (2). The size of the identified sub-network depends on this parameter.

In the study case data, the Spearman correlation between genes in the tumour data has been computed in R.

4.2.1.3 Scoring sub-networks

The aim of the procedure is to identify the highest sub-networks that are scored by computing the sum of the profits subtracted by the sum of the cost of its members. For a given sub-network A with a set of nodes N and a set of edges E, the module score $score_A$ is then defined as follows:

$$score_A = \sum_{g \in N} scoreNode_g - \sum_{e \in E} scoreEdge_e \quad \textbf{(3)}$$

The scores of a node $scoreNode_g$ and the score of an edge $scoreEdge_e$ are computed by applying equations (1) and (2) respectively.

4.2.1.4 The algorithm to identify sub-network with highest score in the interaction network

In order to identify the sub-network with the highest score in the interaction network, the algorithm developed and implemented in [219] was used. This algorithm has been shown by Ljubic *et al.* [219] to solve the Prize Collecting Steiner Tree (PCST) problems to optimality. Ljubic *et al.* have first transformed the mathematical problem of identifying the sub-network with the maximum score into an equivalent problem called Steiner arborescence problem. The authors have then shown that a feasible arborescence with minimum total edge score corresponds to an optimal PCST [219]. Here, in Figure 4.1, an example is shown demonstrating the transformation of an undirected graph into a Steiner arborescence. To identify minimum Steiner arborescence tree, they have used integer linear formulation and a branch and cut framework using the software dhea [219]. For this, several types of constraints (connectivity, asymmetry and flow balance) are introduced and violated constraints are identified iteratively until all constraints are satisfied [219].

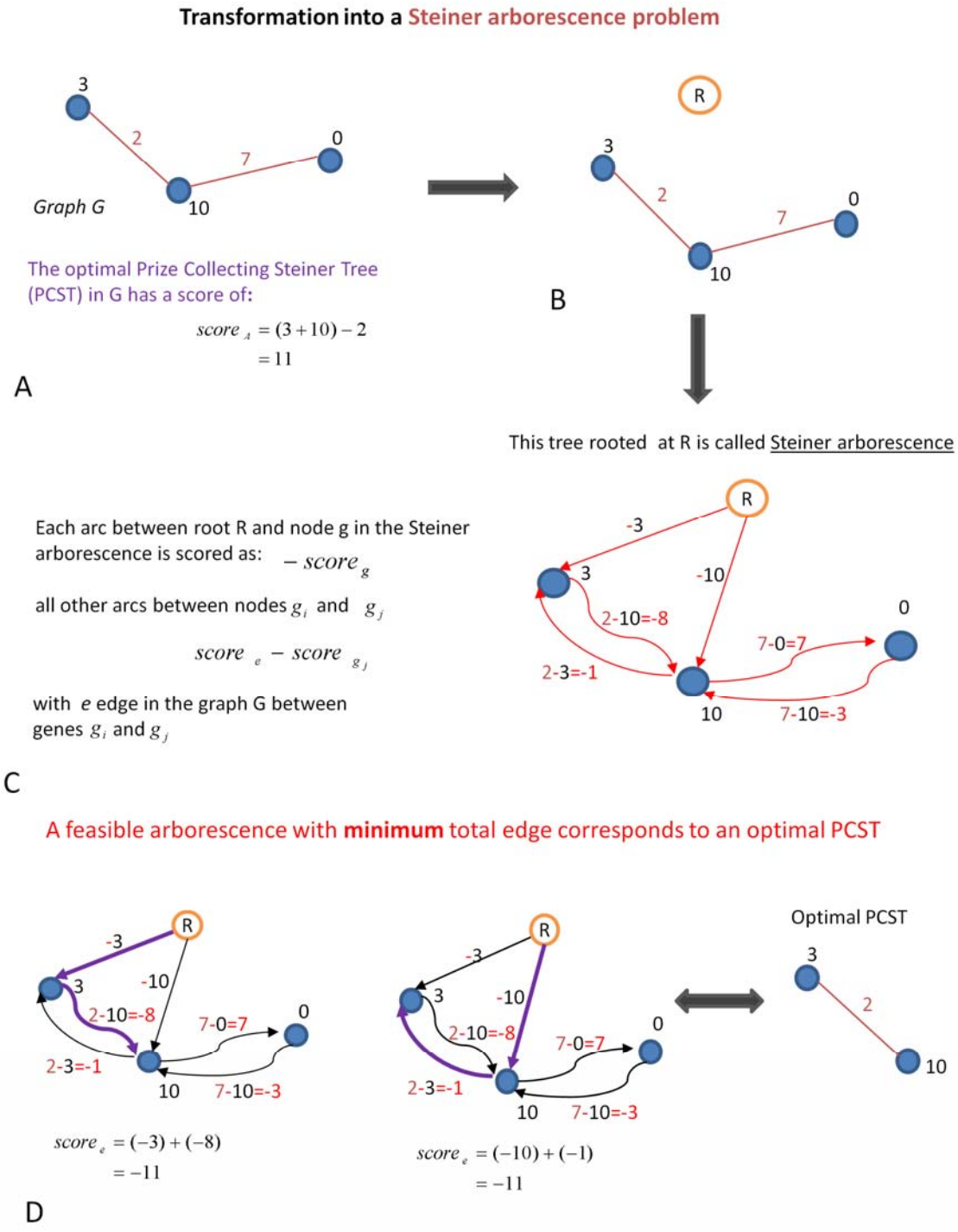


Figure 4-1. Transformation into Steiner Arborescence problem for identifying optimal Prize Collecting Steiner Tree.

Panel A is an example of a graph G with three nodes of weight 3, 10 and 0 and with edges of weight 2 and 7. The optimal sub-network in this graph has a score of 11 and contains the nodes scored by 3 and 10 (calculated using equation 3). In order to identify the optimal sub-network, the graph is first transformed into a Steiner arborescence problem. For this, first an artificial root R is inserted in the graph (Panel B). Then a direct edge from R to all other nodes in G with score higher than zero is introduced and the edge is scored using equations in Panel C. Arcs between all other nodes are introduced and the cost is recomputed using

equations in Panel C. A feasible arborescence with minimum total edge score corresponds to an optimal PCST, i.e. a best solution.

This software is implemented using IBM ILOG CPLEX Optimizer, which is a high-performance mathematical programming solver for integer linear programming. For academic users, CPLEX is available with no-charge via IBM's Academic Initiative program.

4.2.1.5 Correcting the network score for random effects

It is of crucial importance to test the likelihood of the module of being occurred by chance. On this account, the z-score of the sub-network separately for edges and nodes were computed as follows:

For a given sub-network A of size K = k, an edge z-score is then defined as:

$$z\_score\_edges_A = \frac{\sum_{e \in A} scoreEdge_e - \mu_k}{\sigma_k} \qquad (4)$$

where μ_k and σ_k are respectively the mean and standard deviation of network edge scores (as defined in [2]) computed for 1000 random sub-networks of size K.

Similarly, for a given sub-network A of size K = k, the node z-score is defined as:

$$z\_score\_nodes_A = \frac{\sum_{g \in A} scoreNode_g - \mu_k}{\sigma_k} \qquad (5)$$

where μ_k and σ_k are respectively the mean and standard deviation of network node scores (as defined in [2]) computed for 1000 random sub-networks of size *k*. In this application, these were estimated from the module score (calculated using equation (4) and (5)) distribution derived from resampling 1000 random sub-networks of size K.

4.2.1.6 Searching for independent multiple sub-networks

The edge scoring function allows controlling the contribution of the cost by using the parameter *w* in equation (2). This results to the identification of sub-networks with various sizes. Each identified subnetwork is associated with a z-score computed from resampling random sub-networks of the same size by using equation (4) and equation (5). It is possible to compare these networks by their z-scores and choose the network with the highest z-score divided by the square root of the edge or node size respectively.

The sub-network with highest score is removed from the interaction network and then the search procedure is restarted to identify other independent sub-networks iteratively. These iterations are repeated until there are no networks with acceptable p-values calculated using (3) and (4). The original algorithm can provide also the suboptimal solutions using hamming distance [219]. However, the iterative procedure allows identifying completely independent networks that are relatively small and easy to interpret.

4.2.1.7 Method validation on a simulated search space derived from a real data

In order to assess the performance of the method, two simulated datasets have been generated to represent search spaces with different properties. The objective was to first generate networks with a real network topology but with non-statistically significant nodes and edge properties. Within these networks, modules of a defined size were then selected and scored with varying degrees of statistically significant node and edge properties. The aim was to test the ability of the procedure to recover these modules.

Construction of the simulated search space

The real underlying network structure was derived by protein-protein interaction data extracted from the HPRD database and covering the whole gene complement represented in Affymetrix platform (Human U133 Plus2 Gene Chips). This has resulted in a network

consisting of 4540 nodes (genes) and 14903 edges (interactions). In order to score the network edges, MI values were computed from a randomized gene expression dataset. This dataset represent 44000 probes and 54 samples and it is described in Chapter 3 section 3.2.2.

Significant network modules were constructed as follows. Three modules of size 22 in one simulated dataset and three of size 55 in the other dataset were first selected randomly from the global network. Within these networks, modules edges were scored by sampling MI values from normal distributions representing high, medium or low statistically significant MI intervals (**Figure 4.2**). The intervals are defined by partitioning the area between the 90% quantile of the MI values from real dataset and the random dataset as described in [154]. Genes belonging to the selected modules are set signal p-values uniformly distributed between 0 and 0.10 and background noise p-values uniformly distributed between 0 and 1 as proposed in [216].

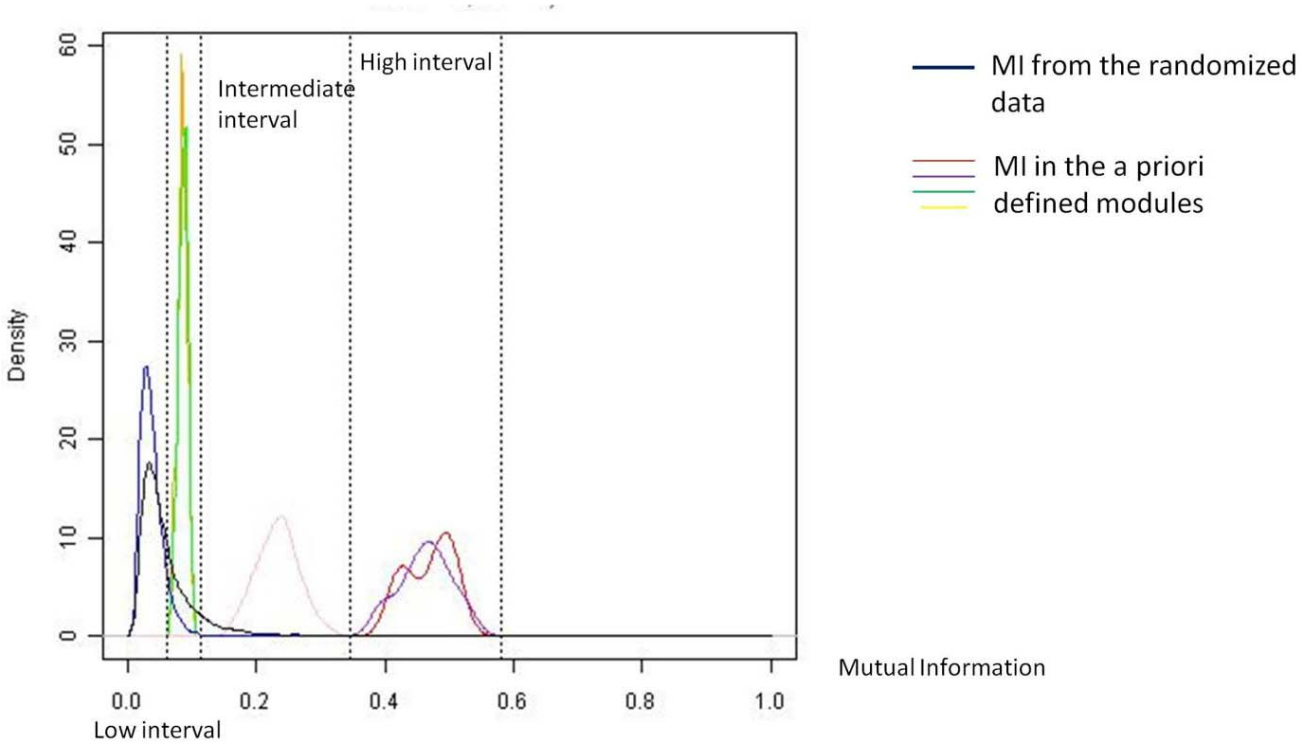


Figure 4-2. Generation of a priori defined modules in the simulated search space.

Method performance

In order to test the accuracy of the procedure, the precision and recall were computed for recovering different size and degree of gene-to-gene correlation (see method section for details). The modules with the highest degree of internode mutual information were identified by the procedure with a high recall and precision, regardless of the network size. More precisely, the module of size 22 has a recall of 0.8 and a precision of 1 (**Figure 4.3A**) and the module of size 55 has a recall of 0.9 and a precision of 1 (**Figure 4.3B**). Network modules of intermediate node correlation (**Figure 4.2**) were also identified with high recall and precision. The smaller module (22 nodes) has a recall of 0.79 and a precision of 0.90 (**Figure 4.3C**), whereas the largest module (55 nodes) was identified with a precision of 1 and recall 0.74 (**Figure 4.3D**). Network modules of low correlation between nodes has resulted to precision 1 and recall of 0.62 for the large module (55 nodes) and precision of 0.84 and recall of 0.52 for the small module (22 nodes).

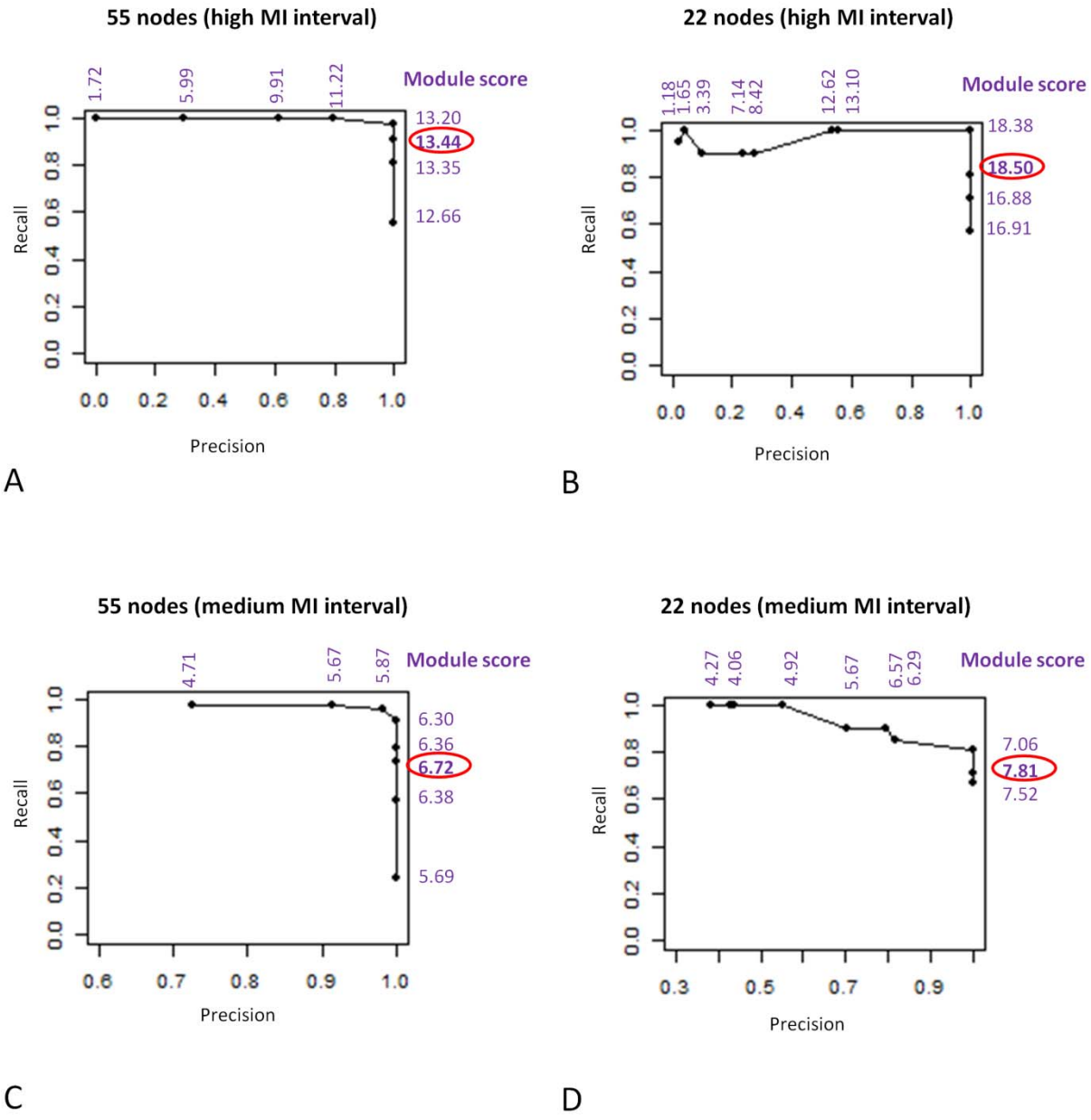


Figure 4-3 Accuracy of the identified modules.

Panel A and B represent the modules pre-defined in the high MI interval respectively with 55 nodes or 22 nodes. Panel C and D represent the modules pre-defined in the intermediate MI interval respectively with 55 nodes or 22 nodes.

Comparison of using p-values vs. FDR as an input to node scoring function

As using corrected p-values as an input in the node scoring function could influence the performance, the p-values have been corrected by performing a Benjamini-Hochberg

multiple correction. The node scores in the simulated search space was consequently recomputed, which was then used as an input in the methodology. The performance was then reassessed for large module in the high MI interval. From the identified modules 4 out of 6 were also identified using p-values as node scores with an overlap higher than 80% (Table 4.1). This analysis has shown that the highest scored module with the largest score allows identifying a priory defined network with 55 nodes with a recall of 0.98 and precision of 1.

Run	Nodes scored from p-values			Nodes scored from FDR			Overlap
	Recall	Precision	Size	Recall	Precision	Size	
1	0.55	1	30	0.66	1	36	83%
2	0.90	1	49	0.90	1	49	100%
3	0.98	1	53	0.98	1	53	100%
4	1	0.29	88	1	0.55	96	91%

Table 4.1. Comparisons of modules identified from different node inputs.

This table represents the modules identified using p-values for scoring the nodes (column 1) and Benjamini-Hochberg corrected p-values (column 2). These modules share a large overlap (larger than 80%) between the two inputs.

Performance compared with other algorithms

Two other methods have been used for comparing the performance with the methodology presented here. For this, the same simulated search space presented in section 4.2 with a priori defined networks of 55 nodes has been used as an input to FUMO [154] and jActiveModules [151]. FUMO aims to identify functional modules with high mutual information and jActiveModules aims to identify functional modules with genes differentially expressed between experimental conditions. Table 4.2 shows that FUMO identify modules with high recall but low precision. The modules identified by jActiveModules have high precision but low recall. Overall, the method presented in this chapter has a better

performance than these two methods. The algorithm used in this chapter has been also shown to outperform jActiveModules by Dittrich *et al.* [216].

Module #	FUMO		jActiveModules		Integrative approach	
	Precision	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
m1	0.63	0.18	0.15	0.80	1	0.90
m2	0.94	0.27	0.17	0.91	1	0.74
m3	1.00	0.15	0.17	0.89	1	0.62

Table 4.2. Performance comparison with other methods.

This table shows the accuracy of the modules in high, intermediate and low MI intervals identified by FUMO (first column), by jActiveModules (second column) and by the integrative approach presented in this chapter (third column).

4.2.2 Application to real data

The chicken chorioallantoic membrane (CAM) model and expression profiling

The transcriptional dynamics of tumor cells in a model of human glioblastoma cancer development (U87 cell line) was studied. This data has been generated in the group of Bikfalvi. U87 human glioma cells (American Type Culture Collection) were maintained in DMEM with 10% FBS, antibiotics, and l-glutamine. Fertilized chicken eggs (*Gallus gallus*) (EARL Morizeau, Dangers, France) were handled as described [224].

On embryonic day 10, a plastic ring was placed on the CAM, and 3 million to 5 million U87 cells in 20 µl of medium were deposited after gentle laceration of the surface. Digital photos were taken under a stereomicroscope (Nikon SMZ800). When required, tumor size-matching was done based on tumor volume calculation: $V = 4/3\pi r^3$, with $r = 1/2\sqrt{(d1 \times d2)}$.

CAM and glioma tissue was extracted together from CAMs at 10 time points in 12 hour intervals over 5 days and snap frozen in LN2. Total RNA was extracted by using RNeasy

mini kit (Qiagen, Courtaboeuf, France). RNA quality and quantity were assessed by agarose gel electrophoresis and optical density measurement. RNA was reverse transcribed (cDNA Synthesis System Kit, Roche Diagnostics), purified and labelled using Cy3 (Megascript D7 Kit, Ambion, Austin, TX). Labelled cRNA containing both chicken and human cRNA were hybridised to Agilent Chicken V1 and Human GE V2 microarrays following manufactures standard protocols. Arrays were scanned using an Agilent HD Array scanner. The microarrays were generated by Sarah Durant, Philipp Antczak and Kim Clarke (University of Birmingham).

Detection of potentially cross-hybridising probes on the Chicken and Human arrays

Two colour microarrays were used to determine human and chicken probes that could potentially cross hybridise to chicken and human microarrays respectively and thus contribute to a false positive signal. RNA from pre-implantation CAM and U87 cells were labelled with Cy3 and Cy5 dyes respectively. Pooled labelled cRNA from CAM cells and U87 cells were hybridised to both Agilent Chicken V1 and Human GE V2 arrays. The microarray data was quantile normalised and the background signal for each dye was subtracted. A contribution score was then calculated for each human probe on the chicken array and vice versa. The score is defined as the signal intensity divided by the total signal for both dyes. Probes with a signal higher than 50 in a linear scale and a contribution greater than 15% were defined as potentially cross-hybridising. 9128 human probes and 5272 chicken probes were removed from further analysis. This dataset was processed by Kim Clarke (University of Birmingham).

Normalization and Annotation of the Human and Chicken Arrays

Array data was quantile normalised and the background signal was subtracted using the statistical software package R. Where multiple probes were available for a gene, the highest

average expression amongst those probes was used for further analysis. This dataset has been processed by Kim Clarke (University of Birmingham).

Survival analysis

In order to test whether the module derived from the approach developed in this thesis is also predictive of clinical outcome, a publically available dataset [225] has been analyzed. This dataset consists of 24 grade III and 50 grade IV glioblastoma patients. For each patient the survival time has been recorded either to current time if the patient is alive or to the day of death. An analysis has been performed by the authors in the original study published in [226] that allowed grouping patients into poor and good survival groups using a partial least squares procedure. In this thesis, the differentially expressed genes have been identified between the poor and good survival groups using a t-test. The t-statistics has been used to rank the genes in order to test by using a GSEA analysis whether there was a significant enrichment in the module identified by the modularization approach developed in this chapter.

4.3 RESULTS

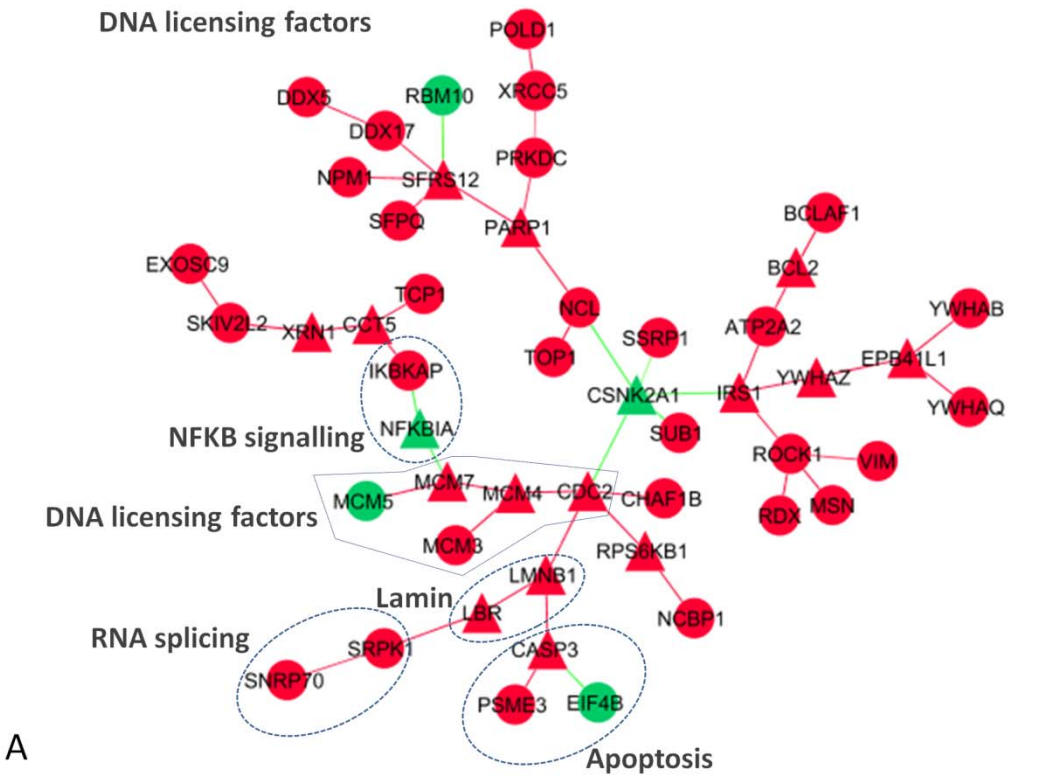
4.3.1 A novel modularization approach identifies new role for RhoE during tumour development in CAM

It has been shown that RhoE has a role in the regulation of cell proliferation and survival in U87 cells in vitro [227]. In this study, the role of RhoE targets in an *in vivo* CAM model was addressed. I have aimed to identify modules describing the regulation of effector functions controlled by RhoE during the development of a fully vascularised tumour. For this, the transcriptional data derived from an *in vivo* CAM model and putative RhoE targets identified by immuno-precipitation followed by mass spectrometry were integrated in the modularization framework described in this chapter. First, the interaction network was built of 3399 genes and 9640 interactions resulted from the integration between the HPRD database and the Human tumour cell time course dataset available from the CAM model. The nodes and edges were then scored using the equation (2) and (3). The identified modules in this case are aiming to identify highly co-regulated genes in the neighborhood of the putative RhoE interactors. In total, 4 modules have been identified by this procedure (**Table 4.3, Figure 4.4A and Appendix Figure A4.1**). These modules are functionally enriched in genes related to cell cycle, DNA repair, cell death, proliferation and extracellular matrix components (**Table 4.1**). The module with the best scored (m1) shows that RhoE is strongly linked with nuclear proteins involved in DNA replication (**Figure 4.4A**). The genes in the identified module can separate between low and high survival in glioma patients classified with grade 3 and grade 4 tumours. Consistent upregulation of genes in patients with bad prognosis with median survival of 237 days compare to good prognosis patients with median survival of 4.8 years identified by GSEA analysis (**Figure 4.4B**).

M#	size	Z-score	p-value	Functional Annotation
m1	47	12.96	< 1e-16	GOCC nuclear lumen [26]. GOBP DNA metabolic process [13]. KEGG cell cycle [8]. SMART MCM [4]. GOBP DNA replication [8]. GOBP DNA repair [7]. GOBP Apoptosis [11].
m2	71	11.03	< 1e-16	GOCC nuclear lumen [32]. GOMF ATP binding [16]. GOBP DNA repair [7]. GOBP Cell death [12].
m3	41	7.44	< 1e-11	KEGG Cell cycle [6]. GOBP Regulation of cell proliferation [9]. KEGG ECM-rec interaction [4]. GOBP Positive reg of cell proliferation [9].
m4	51	6.34	< 1e-6	GOBP Protein transport [11]. GOCC nuclear lumen [16].
m5	34	1.80	0.07	

Table 4.3. Modules identified from the search procedure.

The second column indicates the size of the module, the third column is the z-score derived from (4), the forth column represents the associated p-value, and the fifth column represent the functional annotation of the module derived from DAVID.



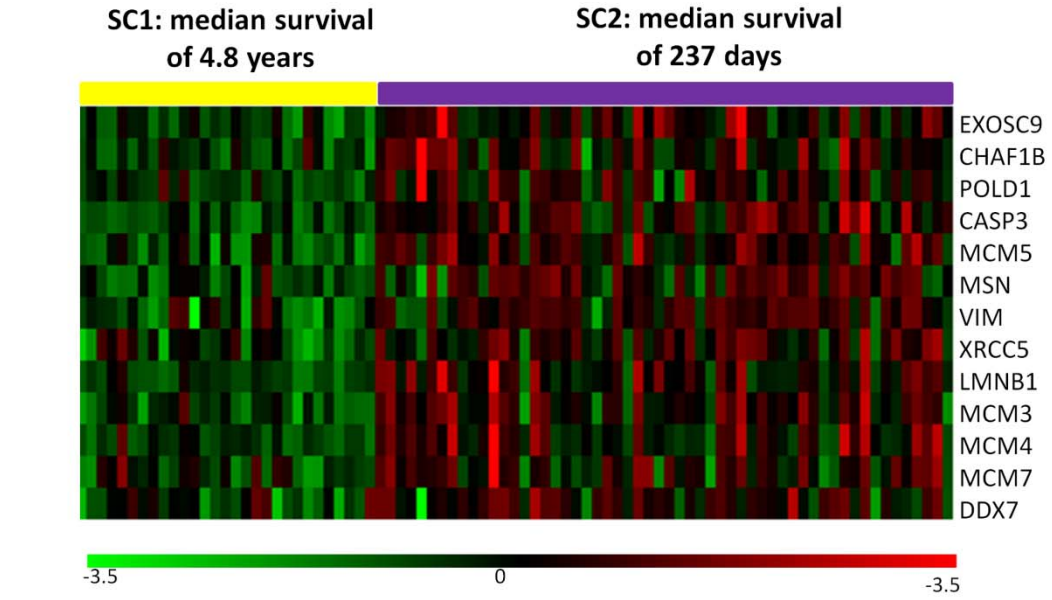


Figure 4-4. The module identified by integration of the CAM model, RhoE targets and interaction networks and the survival analysis.

Panel A shows the module identified by the approach in this chapter. Panel B shows the heatmap of the genes significantly different between SC1 and SC2 patients that are enriched in the module as identified by GSEA analysis. The heatmap shows that these genes are upregulated in patients with low survival (SC2) and downregulated in patients with high survival (SC1).

4.3.2 Experimental validation shows that RhoE is localized in the nucleus and its inactivation alters the localization of MCM3

The network has shown that RhoE is strongly linked with nuclear proteins involved in DNA replication (e.g. MCM3). To test whether RhoE is involved in nuclear localization, western blot and immuno-fluorescence staining analysis were carried out by two collaborators Sarah Durant and Dr. Neil Hotchin. The results have shown that RhoE is localized in the nucleus in U87 cells (**Figure 4.5**) and MCM3 accumulates in the nucleus of RhoE-depleted cells (**Figure 4.6).**

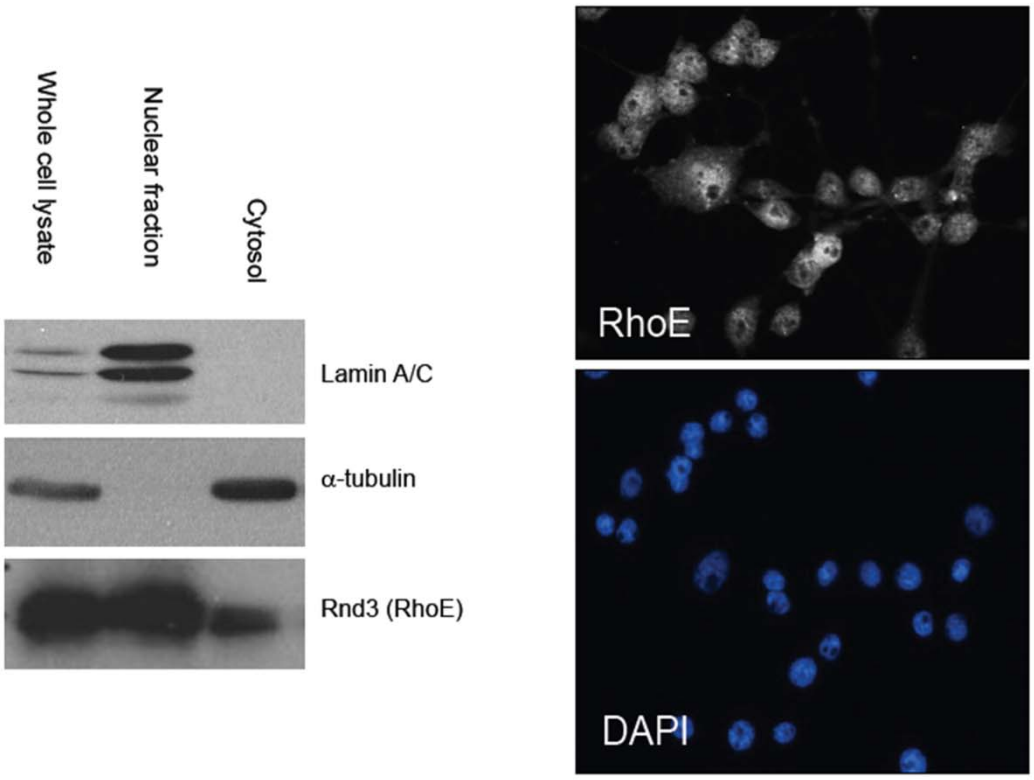


Figure 4-5. RhoE is localized in the nucleus.

Western blot and immune fluorescence staining shows that RhoE is localized in the nucleus in U87 cells (Panel A)

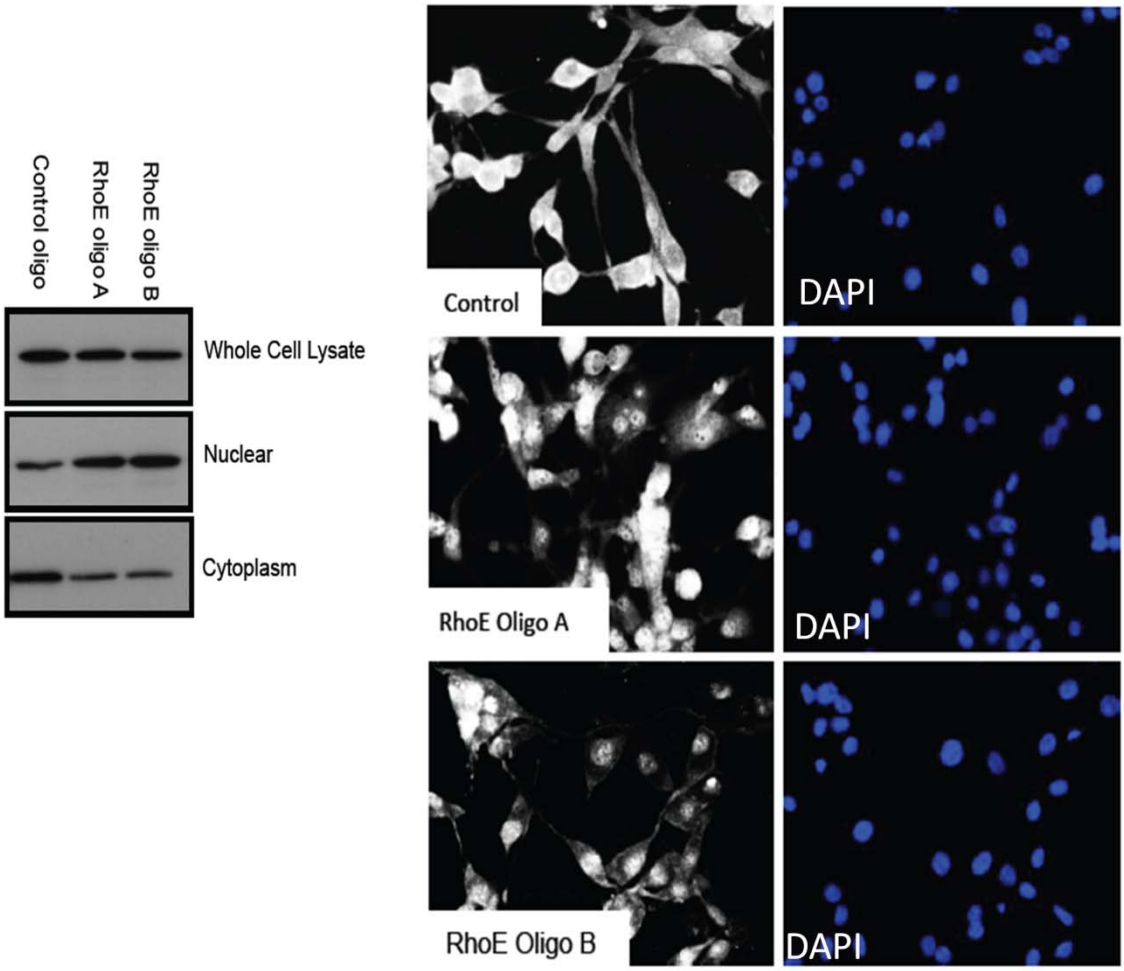


Figure 4-6. RhoE inactivation alters the localization of MCM3.

Western blot and immune fluorescence staining shows that MCM3 accumulates in the nucleus of RhoE depleted cells.

4.4 Discussions and Conclusions

In this chapter, a novel modularization approach has been developed and tested. This procedure allows the integration of several datasets into one framework. The performance on the simulated data has shown accurate identification of a priori defined networks. The case study has also shown the usefulness of the method in generating new hypothesis that are experimentally testable.

Identification of overlapping modules

As genes can be associated with several functions, the modules in reality could be overlapping and the modules might have fuzzy borders. The methodology described in this chapter allows only identifying non-overlapping modules. One way to overcome this issue is to identify several sub-optimal solutions at each iteration using the hamming distance as proposed in the original algorithm (dhea) developed by Lubic *et al.* [219]. This would allow identifying networks that share common genes between the functional modules.

Other applications

The approach developed in this chapter can be used to integrate various types of data as long as the nodes are associated with p-values and edges with a metric similar to correlation. The initial interaction network can be constructed as presented in this chapter from the literature or by using pairwise correlation from the data. Other types of application of this methodology could be based in integrating genome-wide association study (GWAS) focussing on SNPs association with a disease and co-regulation from gene expression dataset. In a recent study, GWAS and PPI information was integrated into a modularization approach in breast cancer and pancreatic cancer [228]. For this, the authors have used the most significant SNP for each gene which is located within the 20kb immediately upstream or downstream of the gene [228]. In the methodology described in this chapter, a similar approach can be used to score the nodes and co-regulation from gene expression could be used to score the edges allowing the integration of multi-level datasets.

Another application of this methodology could be to identify network modules whose genes are transcriptionally correlated through a sample series and at the same time differentially regulated between two or more experimental conditions in two different systems. In this case, the interaction network is represented by a graph where a differential expression *p-value* is a node property and gene-to-gene correlation is an edge property. Indeed, in the next chapter,

this approach will be performed to integrate a clinical dataset, differentiation data of hMSCs and interaction networks into one framework for studying bone and cartilage degradation in various conditions.

CHAPTER 5: DYNAMICAL MODELS FOR UNDERSTANDING BONE AND CARTILAGE DIFFERENTIATION FROM MESENCHYMAL STEM CELLS IN NORMAL AND PATHOLOGICAL SCENARIOS

5.1 INTRODUCTION

hMSCs have the ability to differentiate towards bone, fat and cartilage cells and as such they are key for the maintenance of the musculoskeletal system. Aberrant differentiation of hMSCs has been observed in ageing and in various diseases including multiple myeloma and arthritis [104, 106, 107, 108]. In ageing, the number of adipocytes in bone marrow increases, resulting in the appearance of fatty marrow [104]. The interplay with fatty marrow reduced osteoblastic bone formation and increased osteoclastic bone resorption causes sustained bone loss [104]. Reduced osteoblastic differentiation and increased activity of osteoclasts has also been shown to contribute in rheumatoid arthritis [229, 230, 231, 232] and in multiple myeloma [106, 107, 108].

Activation and inhibition of several signaling pathways, growth factors and hormones have been found to play a crucial role in the process of hMSC differentiation towards the three lineages in both healthy tissue and patients with bone and cartilage degradation. For instance, PPARG activation is required for normal adipogenesis and the activity of PPARG has been observed to be increased with ageing, which could contribute to increased activity of adipogenesis in ageing. Moreover, PPARG inhibition is required for normal osteogenesis and the high activity of PPARG can at least partially explain the reduced osteogenesis activity during ageing. The transcription factor RUNX2 is known to be a regulator of osteogenesis activity, which has been shown to be inhibited in hMSCs from MM patients potentially by soluble factors (IL7), Wnt antagonist (DKK1), secreted frizzled (sFRP3 and sFRP2), and VCAM [106, 107].

Although these studies have contributed to the identification of several key pathways, the exact mechanisms of hMSC differentiation are still unclear. Several microarray studies aiming to understand the dynamics of differentiation of hMSCs towards bone [38, 233] and cartilage [234] have been previously reported. However, construction of time dependant gene regulatory networks reflecting these processes have not been achieved yet. This represents a crucial task as it could enhance our understanding of the differentiation processes. In this chapter, the development of the first genome-wide dynamical models of hMSCs differentiation towards the three lineages is described. On a general level, these models successfully identify several signaling pathways in the most upstream events leading to the activation of effector pathways such as ossification and ECM in the osteogenesis and the chondrogenesis model. These models also allowed the identification of potential candidates that could enhance the efficiency of differentiation in ageing and in disease condition. More specifically, this study predicts that knock-down of DUSP6 and over-expression of MYC in hMSCs derived from MM patients could enhance the osteogenic differentiation. It also

predicts that inhibition of TBX3 in hMSCs derived from elderly individuals could enhance osteogenic differentiation. The analysis pipeline developed is general and could potentially be used for identification of targets in other diseases associated to bone or cartilage degradation.

5.2 MATERIAL AND METHODS

5.2.1 Datasets

5.2.1.1 hMSCs differentiation and microarray datasets

Osteogenic differentiation from hMSCs

This data has been published by Piek et al [38], but for clarity the details of the differentiation protocol is reported here. hMSCs were purchased from Lonza (Walkersville, MD) cultured *in vitro* up to passage 8 and then differentiated towards bone, cartilage and fat cells. For osteogenic differentiation, 2.0×10^4 cells per cm2 were seeded in DMEM supplemented with 10% fetal bovine serum (a selected lot from Lonza Walkersville, Inc.), 100 U/ml penicillin. The next day, cultures were sub-confluent and cells were switched to osteogenic differentiation medium (DM) composed of high-glucose containing Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (Lonza Walkersville, Inc.), 100 U/ml penicillin, 100 µg/ml streptomycin, 10−7 M dexamethasone (DEX), 0.2 mM ascorbate and 10 mM β-glycerophosphate, and 10−8 M 1α, 25-dihydroxyvitamin D3 (1,25(OH)2D3; Calbiochem. Cells treated with DM in the absence of dexamethasone (further referred to as −DEX) were used as negative controls and are also referred to as "untreated." Medium was refreshed every 3 days. Histochemical staining and flow cytometric analysis of ALP-positive cells indicated that 1,25(OH)2D3 strongly potentiated osteoblast differentiation induced by DEX, resulting in cultures that contained more than 70% ALP-positive cells after 9 days of treatment.

hMSCs differentiation towards adipocytes

For adipogenic differentiation, 4.0 x 10^4 cells per cm^2 were seeded DMEM supplemented with 10% fetal bovine serum (a selected lot from Lonza Walkersville, Inc.), 100 U/ml penicillin and 100 µg/ml streptomycin. The next day, the cells were switched to adipogenic differentiation medium (AD), which consisted of the above proliferation medium now

supplemented with 10^{-6} M dexamethasone, 10 μg/ml insulin (R&D Systems), 10^{-7} M rosiglitazone (Campro Scientific, The Netherlands), and 500 μM IBMX (3-isobutyl-1-methylxanthine, Sigma-Aldrich, St. Louis, MO). Medium was refreshed every 3 days. Cells treated with proliferation medium (PR) were used as controls. (Generated in the group of Everardus J.J. van Zoelen)

hMSCs differentiation towards chondrocytes

For chondrogenic differentiation, hMSCs were trypsinized and 2.5x 10^5 cells pelleted in a 10 ml round bottom tube (Greiner Bio-One, Monroe, NC) for 10 min at 250xg. Cell pellets were subsequently cultured for 21 days in chondrogenic differentiation medium, consisting of proliferation medium supplemented with 6.25 μg/ml insulin, 6.25 μg/ml transferrin, 6.25 ng/ml sodium selenite, 5.35 μg/ml linoleic acid, 400 μg/ml proline, 1 mg/ml sodium pyruvate, 10^{-7} M dexamethasone, 50 μg/ml sodium L-ascorbate (all obtained from Sigma-Aldrich, St. Louis, MO) , in the absence (INC) or presence (TF) of 10 ng/ml recombinant TGFβ1 , and supplemented with 50 ng/ml GDF5 (TG). Medium was refreshed every 3 days. Growth factors were obtained from R&D Systems. (Generated in the group of Everardus J.J. van Zoelen)

Microarray datasets

In total, 396 RNA samples were obtained from triplicate experiments of all three lineages, four biological conditions and each measured at 11 time points (0, 1, 3, 6, 12, 24, 48, 72, 120, 192 and 288 hours after onset of treatment). RNA was extracted using TRIzol® according to the protocol provided by the manufacturer (Invitrogen). For each sample, 5 μg of RNA were reverse transcribed into double-stranded cDNA, and used as a template for the preparation of biotin-labeled cRNA, as previously described (Vaes, et al., 2006). A total of 10 μg of biotin-labeled cRNA was hybridized to the Human Genome U133A Array (Affymetrix), after which hybridization signals were amplified using a streptavidin–biotin amplification procedure. Arrays were hybridized and scanned with a GeneChip G3000 scanner (Affymetrix). Data

were quantified using GCOS 1.2 software (Affymetrix). (Generated in the group of Everardus J.J. van Zoelen)

5.2.1.2 Multiple Myeloma

hMSCs were extracted from the bone marrow of 10 healthy donors, 10 premalignant patients, and 10 Multiple Myeloma patients (5 with bone lesion and 5 without bone lesion) was performed. Patients with MM were recruited from specialist clinics at the Royal Wolverhampton Hospitals NHS Trust and Heart of England NHS Trust (United Kingdom). The study received approval from the local ethics committees at South Birmingham, Birmingham East, North, and Solihull, and informed written consent was obtained in accordance with the Declaration of Helsinki in all cases. Bone marrow aspirates (~1ml) were obtained from patients at various stages of disease and mononuclear cells were purified by density centrifugation using Lymphoprep (Nycomed, Oslo, Norway). Plasma was collected after cell separation and snap frozen at -80°C. Microarray gene expression analysis employing Affymetrix ® GeneChip technology was performed using Human Exon 1.0 ST Array Gene Chips according to the manufacturer's suggested protocols. This dataset has been generated by Dr. Sarrah Essex (University of Birmingham, UK).

5.2.1.3 Ageing

Linkage with ageing cells has been explored using a microarray dataset developed by Wagner et al [235], dataset available in GEO, accession number GSE12274 (Affymetrix Human Genome U133 Plus 2.0 Array). This represents the transcriptional state of bone marrow mesenchymal stem cells from individuals ranging from 21 to 95 years old (4 young age: 21, 24, 24, 25; 4 middle age: 44, 50, 53, 55 and 4 elderly age: 79, 85, 85, 95 patients).

5.2.1.4 RA-OA

hMSCs were extracted from the bone marrow of 12 rheumatoid arthritis patients and 6 osteoarthritis patients diagnosed on the basis of characteristic x-ray findings and the absence of features suggestive of inflammatory arthritis. Ethical approval for the use of this material was given by the local research ethics committee and all patients gave written informed consent (LREC reference 5735). Fibroblasts were maintained in medium consisting of 81.3% RPMI 1640, 10% FCS, 0.81X MEM non-essential amino acids, 0.81mM sodium orthopyruvate, 1.62mM glutamine, 810U/ml penicillin and 81µg/ml streptomycin, at 37°C in a humidified 5% CO2 atmosphere. mRNA was extracted from all fibroblasts at passage 5, under identical conditions. Microarray gene expression analysis employing U133 plus 2.0 Affymetrix® Genechips was performed. This dataset has been generated by Filer A in the group of Buckley CD (University of Birmingham, UK).

5.2.2 Data processing and differential expression analysis

5.2.2.1 Microarray data processing

The microarray generated from the differentiation of hMSCs (Human Genome U133A Array) and the ageing data (Human Genome U133 Plus 2.0 Array) were normalized using the R library rma, which converts CEL files using a robust multi-array average (RMA) expression measure with the help of probe sequences. Data were quantile normalised.

The myeloma data (Human Exon 1.0 ST Array Gene Chips) has been normalized following the RMA workflow for Affymetrix Gene ST available in the R library oligo [236]]. Gene intensities were calculated using core probeset, which are supported by RefSeq annotation and expected to lead to more reliable signal [237]. Data were quantile normalised.

5.2.2.2 *Identification of statistically significantly expressed genes during differentiation of hMSCs towards cartilage, bone and fat cells*

To identify genes that are differentially expressed during differentiation of hMSCs into the three lineages, ANOVA analysis has been performed comparing the 10 time points for each treatment. The p-values derived from ANOVA analysis have been corrected by performing Benjamini-Hochberg multiple correction [137]. Genes differentially expressed were selected using a threshold of false discovery rate (FDR) smaller than 5% unless otherwise stated.

5.2.2.3 *Identification of statistically significant differences in diseases and in ageing*

In order to establish a link between the differentiation of hMSCs and abnormal microenvironment of Myeloma patients, RA-OA patients and ageing individuals, a differential expression analysis has been performed. For this, a t-test was performed to address the transcriptional differences: between myeloma patients with and without bone lesion, between healthy and myeloma patients. To identify genes that are positively or negatively co-regulated with ageing, spearman ranking coefficient and calculated the associated p-values by using a bootstrap approach. To identify genes that are significantly different between RA-OA patients, a SAM analysis have been performed [186].

5.2.3 Overview of the modelling strategy

The overarching objective of this study is to construct regulatory networks underlying normal hMSC differentiation and identify abnormal response in ageing and in relevant human pathologies. The overview of the analysis strategy is summarized in (**Figure 5.1**).

The task of inferring a global gene regulatory network from functional genomics datasets is very challenging, mainly because of the relatively small number of samples available compared to the number of variables to model. Here, this challenge has been addressed using two complementary approaches. In the first approach, initially the gene expression dataset

have been simplified using a procedure that identifies a relative small number of clusters of highly correlated genes. The average expression profile is then used as a measure of gene module transcriptional activity. Different modules are then linked into a directed network using time delay ARACNE, a well validated network inference procedure. These high-level dynamical models therefore represent the relationship between the transcriptional response of a relatively small number of gene modules. From these models, sub-networks have been identified, which are linked to key differentiation steps (e.g. the regulation of important effector functions) or perturbed in aged or diseased hMSCs (**Figure 5.1A**). In the second complementary approach, the network modularization procedure described in Chapter 4 has been applied to directly integrate pathway information (either physical interaction or pathway diagrams such as KEGG) with co-transcription in normal cell differentiation and differential expression in ageing and diseased hMSCs (**Figure 5.1B**). Using the results of a RNAi screening [45] a separate analysis integrating pathway information, co-transcription and the knowledge of genes that when inactivated induce a loss of differentiation have also been performed (**Figure 5.1C**).

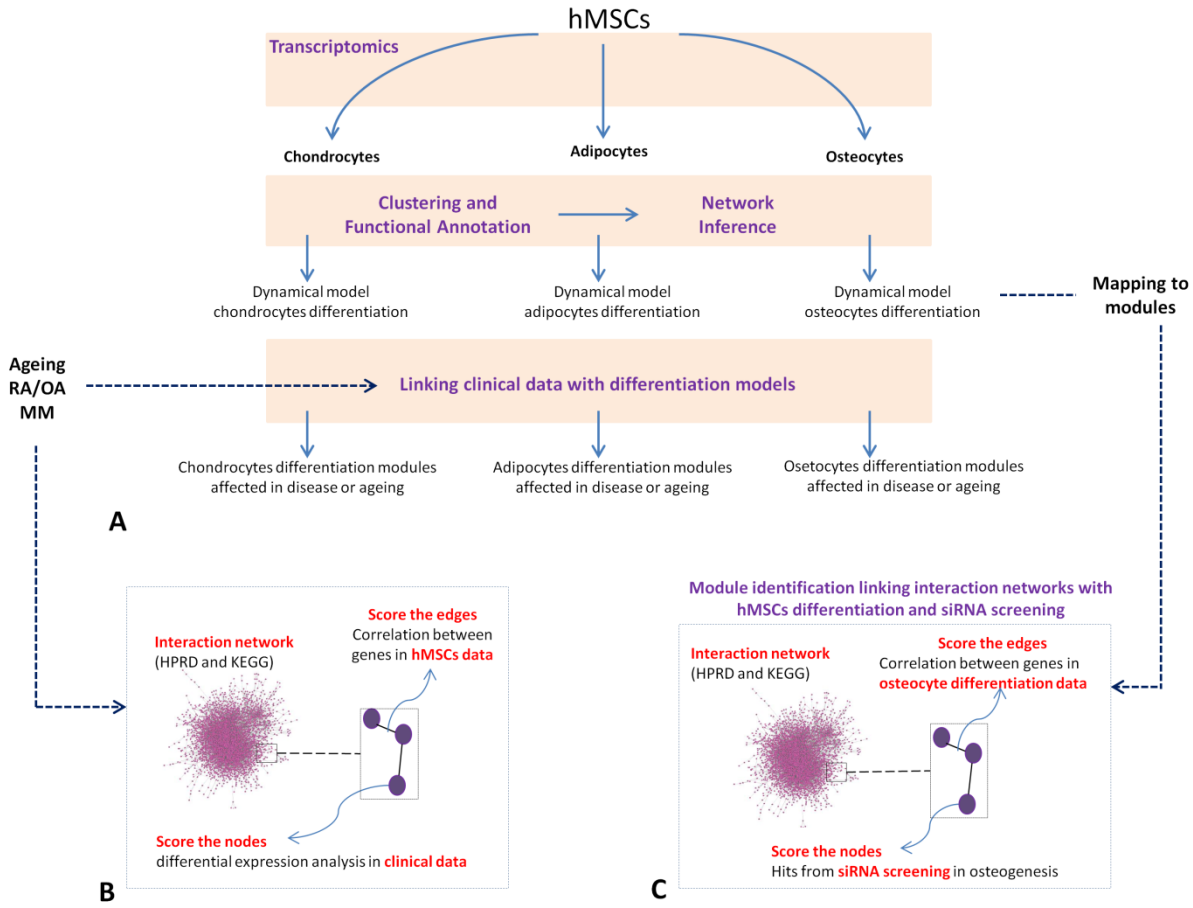


Figure 5-1. Overview of the analysis strategy.

This figure shows the overview of the analysis strategy. Panel A shows the inference of high-level dynamical models representing chondrogenic, adipogenic, and osteogenic differentiation from hMSCs. Panel B represents the network modularization procedure to directly identify gene regulatory networks transcriptionally regulated in normal differentiation and enriched in genes that are differentially expressed in aged and diseased hMSCs**.** The same approach was also used to integrate osteogenesis differentiation data, with siRNA screening and interactions networks (Panel C). The dynamical models were mapped to the modules for further characterization (Panel B and Panel C).

5.2.4 Development of high-level dynamical models of hMSC differentiation

This modelling strategy is based on an initial step aiming at reducing the complexity of the transcriptional response to a relatively small number of gene modules, characterized by highly correlated gene expression profiles. These modules were identified by performing a SOTA clustering [238] using a 90% variability threshold in each cluster. Once the procedure has identified the optimal number of clusters, the average gene expression profiles (cluster

centroid) are computed in each module and this is used as the input of the network inference procedure Time Delay ARACNE [239]. Time Delay ARACNE computes mutual information between pairs of genes at various time shifts and determines the significant connections by using a stationary bootstrap [239]. This led to graphical models representing the dynamics of cell differentiation towards bone, fat and cartilage cells. The significant connections were selected by using a threshold that is expected to lead to 2.5% false positive. For this, the parameter standard deviation multiplier is set to 2 (ksd=2) with a time delay set to 1 (d=1). Heatmaps representing the average expression profiles of each module in the models have been generated. In the time course profile of the module, the first point with a minimum of 50% of the change in the expression occurs has been assigned as the earliest change. Each module was then annotated for functionally enriched KEGG and GO terms, using the web-based tool DAVID [138, 139]. A threshold of 10% FDR was chosen to define statistically enriched pathways.

5.2.5 Identification of modules linked to ageing and diseases in the differentiation models

In order to identify network modules whose activity is affected in ageing and disease, each module is tested for enrichment in genes expressed in hMSCs and correlated with age (dataset with GEO accession GSE12274) or in genes differentially expressed between hMSCs derived from healthy and diseased individuals (datasets derived from MM patients and RA-OA patients). This was achieved by applying the GSEA methodology. In this tool, first genes are ranked by a metric derived from phenotypic correlation or differential expression analysis between experimental conditions (such as t-statistics). Then a score is computed for identifying at the top or bottom of the ranked list significant enrichment in sets of genes that could be defined by the user [240]. In this analysis, gene modules in the differentiation models are used as functional sets and the genes are ranked using t-stats in the

disease datasets and correlation coefficient in the ageing dataset. A weighted enrichment statistic was chosen to calculate the enrichment score that reflects the degree to which a module is overrepresented at the top or bottom of the ranked list. A cut-off of 10% FDR was selected to define significantly overrepresented modules in the disease and ageing data.

5.2.6 Model validation using knowledge databases

I have tested whether the interactions between the modules in the differentiation models are also supported in the literature using a workflow implemented in the commercial knowledge managements system Metacore [241] (**Figure 5.2**). The tool allows identifying enrichment of interactions between sets of genes by using all types of interactions (binding, transformation, phosphorylation, dephosphorylation, cleavage, covalent modification, transport, catalysis, transcriptional regulation) that are included in the database. The modules that are in the differentiation models and subsets of modules that are over-represented in ageing or disease datasets are used as sets in the tool. The significantly over-connected sets are identified using the z-score and p-values (**Figure 5.2**). The z-scores are computed by comparing the number of links between interacting modules (r) with the expected mean value obtained from the hypergeometric distribution (obtained by multiplying number of incoming and outgoing links between modules and divide by the total number of links in the database) [241]. The p-values are also computed using the hypergemoteric distribution [241].

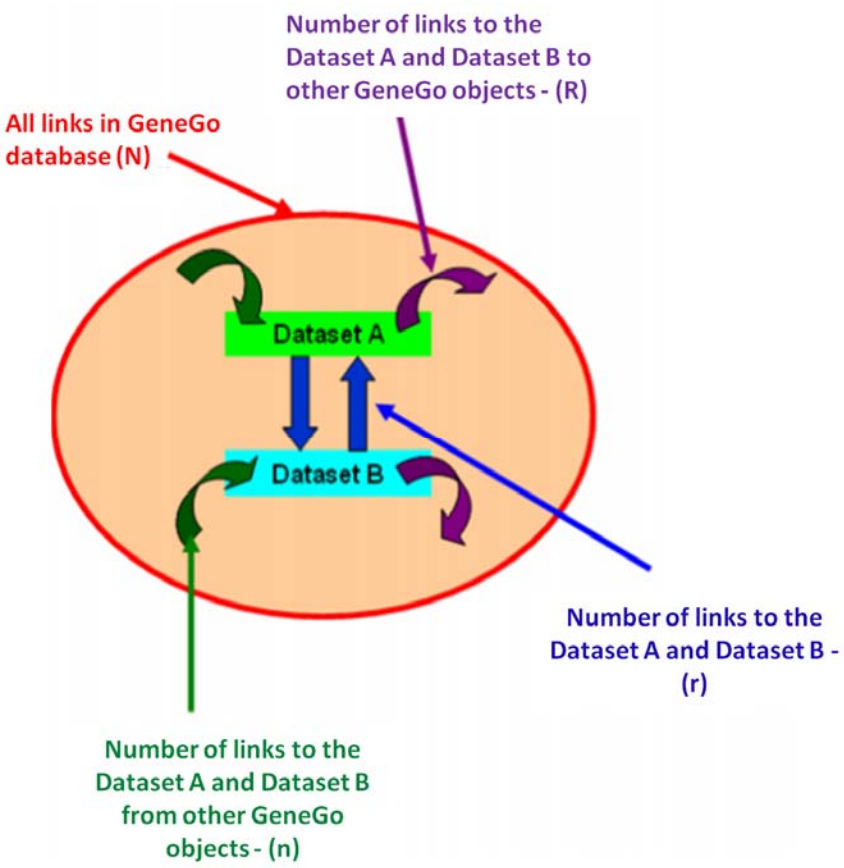


Figure 5-2. Metacore analysis to validate the interactions in the differentiation models, figure as published in MetaCore from GeneGo Inc.

This analysis allows identifying the links between components in the differentiation model that are also supported in the literature. The significant interactions between datasets are identified using the z-score and associated p-values. The z-scores are computed by comparing the number of links between interacting datasets (r) with the expected mean value obtained from the hypergeometric distribution (obtained by multiplying number of incoming and outgoing links between datasets and divide by the total number of links in the database).

5.2.7 Direct inference of gene sub-networks representing hMSC differentiation in the normal physiological process and in ageing and disease

In addition to mapping the genes that are affected in disease in the differentiation models, the novel modularization approach (described in Chapter 4) is applied for the direct identification of gene networks linked to hMSCs differentiation in normal individuals, ageing or disease. This methodology allows identification of modules, which share mul arge

biological networks derived from KEGG and p-p interaction networks extracted from the HPRD database (**Figure 5.1B**). The procedure identifies the borders of the modules by maximising a scoring function that integrates node properties (e.g. genes differentially expressed in ageing or in myeloma) with edge properties (e.g. gene expression profile correlation during the differentiation time course).

The same method has also been used to identify potential key regulators of osteogenic differentiation by integrating the results of an RNAi screening for osteoblast differentiation genes [45] (**Figure 5.1C**). In this case, the nodes of the network were scored by the RNAi screening results and the edges were scored by the correlation computed from the osteogenic differentiation data.

5.3 RESULTS

5.3.1 Transcriptional changes during differentiation of hMSCs towards bone, cartilage and fat cells

The first step of the analysis strategy (**Figure 5.1**) is to identify genes differentially expressed in hMSCs differentiation. This analysis has lead to 2506, 1673, 1659 significantly differentially expressed genes during adipogenic, osteogenic and chondrogenic differentiation respectively. These genes were clustered by SOTA analysis in 48 modules in adipogenesis, 40 modules in osteogenic differentiation and 49 modules in chondrogenic differentiation **(Appendix Table A.5.1, Table A.5.2, Table A.5.3**).

Functional analysis of individual modules has shown significant functional enrichment in several processes, many known to be relevant in hMSC differentiation. In **adipocytes** differentiation, clusters were functionally enriched in PPARG signaling, glutathione metabolism, response to oxidative stress, mitochondrion, DNA replication, valine, leucine degradation, fatty acid metabolism, cell cycle, regulation of cell proliferation (**Table 5.1**). The modules representing the differential expression during **osteogenic differentiation** were enriched in ECM, Ossification, Wnt signaling, focal adhesion, growth factor activity, apoptosis, and skeletal system development (**Table 5.2**). The modules of the **chondrogenic differentiation** were enriched in regulation of cell proliferation, cell cycle, ECM, growth factor activity, inflammatory response, mesenchymal cell differentiation, oxidation reduction, Wnt signaling, mitochondrion (**Table 5.3**).

Cluster #	# of genes	Functional Annotation
C11	125	**1. GOCC mitochondrion [37], (4.78). 3. Response to organic nitrogen [5], (2.2). 4. Lipid biosynthetic proc [10], (2.79).**
C12	57	1. **GOCC Mitochondrion [10],** oxidation reduction [9], KEGG thryptophan [3](2.12).
C13	82	1. **KEGG PPAR signaling [5], (2.42). 2. GOBP lipid homeostasis [4], (2.09). 3.**
C16	34	**1. KEGG Lysosome [6], (4.21). 3. GOBP Embryonic organ dev. [5], (2.26).**
C19	124	**1. GOBP Resp to oxidative stress [11], (2.66). 3. KEGG Glutathione met [6], (2.29). 4. GOBP Protein complex biogenesis [13] , (2.27). 7. GOMF regeneration [6], (1.92). 11. Regulation of cell proliferation**
C22	28	1. INTERPRO Fibronectin type 3 [4], (1.35). 2. GOMF TF activity [6], (0.5).
C27	49	1. KEGG TGFB signaling [4], (1.77).
C30	104	**1. GOCC ECM [13], collagen [5], (3.96). 3. GOMF Integrin binding [5]. 4. Protein complex binding [8] (1.82). 6. GOMF Oxidation reduction [8] (0.96).**
C37	137	**1. GOBP Cell cycle [26], (13.56)**
C38	138	**1. GOBP Cell cycle [31], (15.27)**
C39	20	1. GOBP RNA splicing [3], (1.47).
C40	15	1. Regulation of transcription [4], (1.49)
C41	87	2. GOMF Calcium ion binding [14], (2.44). 3. GOMF Actin binding [9](2.2).
C44	138	**1. GOCC Chromosomal part [14], (3.9). 4. KEGG Cell cycle [11], (2.33).** 9. GOBP Regulation of cell morphogenesis involved in differentiation [35 (1.28).
C5	93	**6. GOBP Fatty acid metabolic process [7], (1.6)**
C51	91	**2. GOBP DNA replication [11], (3.14). GOMF ATP binding [19]**
C6	108	**1. GOCC mitochondrial part [35], (8.77). 5. GOBO energy derivation by oxidation of organic compunds [12], (4.94). 6. Valine, Leucine degradation [10], Fatty acid met [5], (4.46)**

Table 5.1. Functional annotation of the genes in modules identified in the adipocytes differentiation

The functions in bold are statistically significant at 10%FDR. Only modules with enrichment are included in this table.

Clusters	# of genes	Functional annotation
C10	56	2. Ossification [3] , (1).
C13	15	1. GOCC ECM [4], (1.06).
C17	19	**1. GOBP Response to hormone stimulus [8], (3).**
C20	41	**1. GOBP Response to organic substance [10], (2.89).** 3. GOBP MAPK [3], (1.36).

C21	43	2. KEGG B cell rec signaling pathway [3], (1.81).
C33	72	**1. GOBP cell cycle [25], (14.74). 7. DNA repair [13], (4.03).**
C39	37	1. Ossification [3], (1.47).
C40	53	**1. GOBP Amino acid activation [9], (5.56). 2. Resp to hypoxia [7], (3.48).**
C44	23	**1. GOBP Apoptosis [6], (2.54). 2. Regulation of cell proliferation [8], GOMF Cytokine activity [4] (2.03).**
C45	74	6. GOCC Adherens junction [6], (1.32). 7. GOMF growth factor activity [6], (1.3). GOBP Cell proliferation [6], (1.25).
C47	56	**1. KEGG Focal adhesion [6], Wnt signaling [4] (1.56). 2.** GOMF Calcium ion binding [10], (1.5).
C50	12	1. GOCC ECM [5], (1.42).
C51	17	1. GOCC ECM [5], (1.28). 2. GOBP Reg of cell proliferation [4], (1.13).
C52	36	**1. GOCC ECM [11], (3.29).** 2. Ossification [4], (2.43). 3. GOBP cell adhesion [7], (1.66).
C7	78	1. GOBP regulation of signal transduction [6], (1.66). 2. GOMF Oxidation reduction [8], (1.57).
C8	73	**1. GOCC ECM [18], (6.62). 2. GOBP Inflammatory resp. [9], (2.62). 5. GOCC apoptosis [11] (1.7).** 7 .GOBP Glutathione met proc [3], (1.56).
C9	90	**1. GOCC ECM [14], (3.22).** 2. GOBP Skeletal sys dev [8] (1.55).

Table 5.2. Functional annotation of genes in modules identified from the osteogenic differentiation data.

The functions in bold are statistically significant at 10%FDR. Only modules with enrichement are included in this table.

Clusters	# of genes	Functional annotation
clust #1	33	**1. GOCC ECM [9], (2.54).** 4. Regulation of cell proliferation [7], (1.31).
clust #12	57	1. GOMF Enzyme inhibitor activity [5], (1.05).
clust #13	25	1. GOBP Response to organic substance [6], KEGG ErbB sig pathway [3], (1.75).
clust #19	47	**1. GOBP Negative reg of cell proliferation [7], (2.41).** 2. GOBP Regulation of hydrolase activity [5], (1.27).
clust #2	40	**1. GOCC ECM [6], (2).** 3. GOBP Regulation of inflammatory resp [3], (1.54). 4. GOBP Mesenchymal cell differentiation [3], (1.2).
clust #21	14	**1. GOBP Regulation of apoptosis [6], (2).**
clust #24	26	**1. GOCC ECM [8], (2.14).**
clust #25	36	3. GOBP Reg of cell proliferation [6], (1.09).
clust #3	32	**1. ECM [13], (4.4). 2. GOBP Ossification [4], (2.73).** 3. GOMF Growth Factor binding [4], (2.2).

clust #31	99	**2. GOCC Focal adhesion [7], (2.55). 3. GOMF Ras GTPase binding [5], (2.27).**
clust #32	39	1. GOBP Oxidation reduction [6], (1.97).
clust #33	43	1. GOBP Apoptosis [6], (1.3). 3. KEGG Wnt signaling [3], (1.15). 4. GOCC mitochondrion [7], (1).
clust #35	40	**1. GOBP cell cycle phase [6], (1.73).** 2. GOMF ATP binding [10], (1.72). 4. GOBP phosphate metabolic process [8], (1.3).
clust #36	16	1. GOBP cell cycle [4], (1.9).
clust #37	13	**1. GOBP cell cycle [8], (4.41).**
clust #39	30	**1. GOBP Regulation of phosphorylation [6], (2.31). 2. GOMF growth factor activity [5], (2.02).**
clust #40	39	1. GOBP Cell cycle [8], (1.54). 4. GOBP negative regulation of cell differentiation [4],(1.21).
clust #56	45	**1. GOCC ECM [11], (2.92).** 5. GOMF Calcium ion binding [7], (1.17).
clust #58	40	1. GOBP Neuron differentiation [4], (1.44).
clust #7	13	1. GOBP Inflammatory resp [3], (1.9). **2. GOCC ECM [5], (1.61).**
clust #9	13	**1. GOCC ECM [5], (1.99).**

Table 5.3. Functional annotation of genes in modules identified from the chondrocytes differentiation data.

The functions in bold are significantly enriched (10%FDR). Only modules with enrichement are included in this table.

5.3.2 High level dynamical models of hMSC differentiation

Osteogenic differentiation model

The analysis above describes the complexity of the transcriptional response of hMSCs during differentiation with a relatively small number of co-ordinately regulated gene clusters (modules), each one with a specific functional enrichment profile. In order to develop an interpretative model representing the dynamics of cell differentiation, a dynamical modelling approach has been applied to a dataset representing the average gene expression profile of each module. The inferred model represents a temporal hierarchy of transcriptional profiles where the most upstream module (C21) include genes represented in the KEGG receptor signaling pathway (NFKBIA, FOS, PIK3CA) that are upregulated as early as 1h post

stimulation (**Figure 5.3**). The other most upstream module C44 is downregulated 6h after stimulation and this module is enriched in genes represented in the gene ontology terms regulation of cell proliferation (IL6, CCL2, IL8, KLF11, LIF, ADAMTS1, PLAU, NGF) and cytokine activity (IL6, CCL2, IL8, LIF). Modules in the intermediate response are enriched in genes belonging to the functional terms growth factor (C45), oxidation reduction (C7), cell cycle (C25) and response to hypoxia (C40). Consistent with the timeframe of phenotypic differentiation, more downstream modules (C51, C52 and C13) are linked to the gene ontology terms ECM and ossification (FHL2, MMP2, CTGF, IGFBP3).

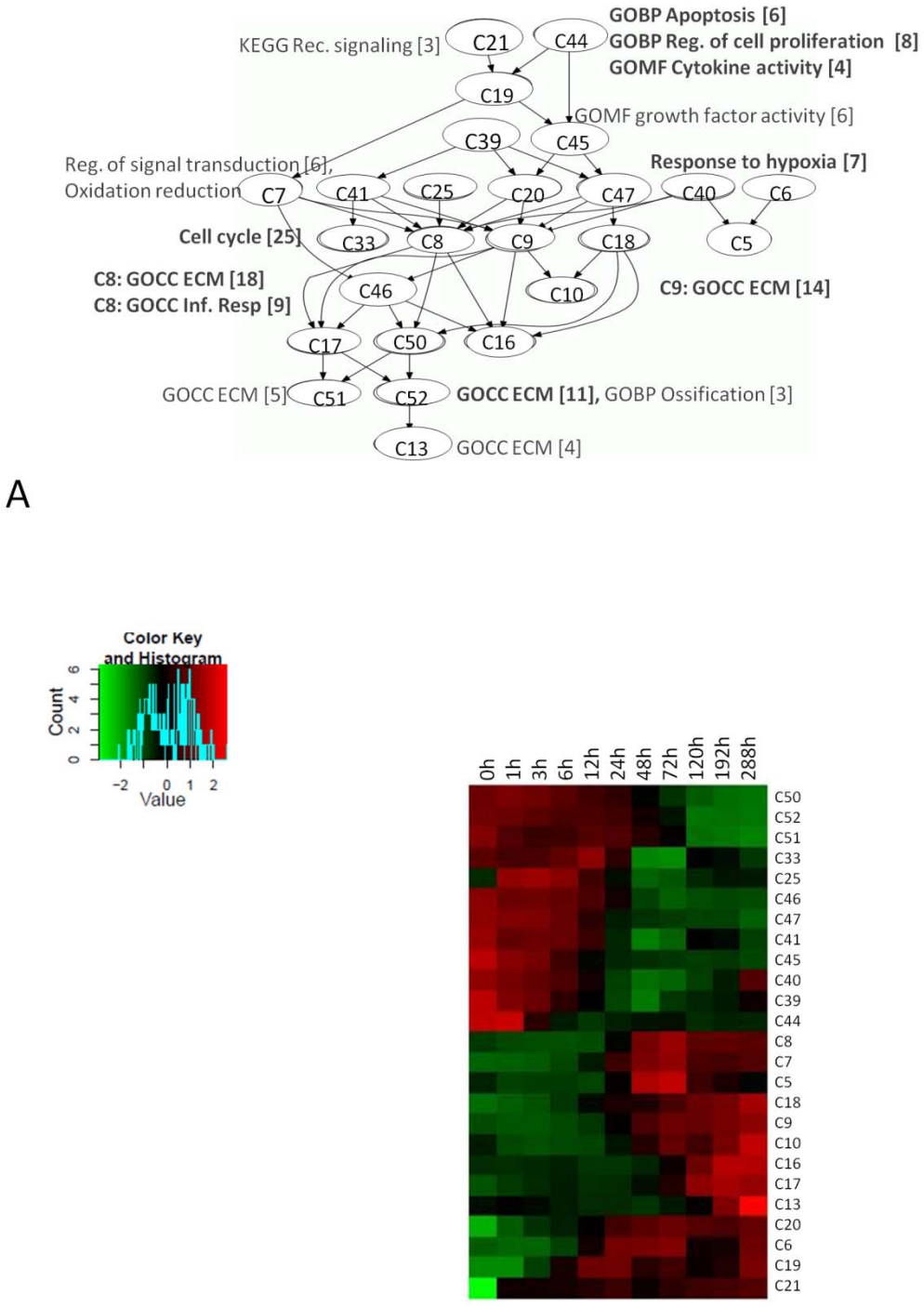


Figure 5-3. The osteogenic differentiation from hMSCs.

This figure shows the hierarchical models representing the differentiation process towards bone cells (Panel A). Each module (node) in the model represents a number of genes that share similar expression pattern during the differentiation. The averaged expression pattern of genes in each module is represented in Panel B. Data are standardized to represent above average values in red and below average values in green. The modules were functionally annotated individually using DAVID analysis. The interactions between modules are inferred using the time delay ARACNE.

Adipocytes differentiation model

The model representing hMSC differentiation towards adipocytes shows a similar high level structure with more upstream modules representing signaling and regulatory pathways and more downstream nodes representing effector functions (**Figure 5.4, Panel A**). For instance, C22 that is upregulated after 12h (**Figure 5.4, Panel B**) contains the transcription factor MYC, the receptors IFNGR1 and IL10RB. The other early modules are functionally enriched in DNA replication (C51), ATP binding (C51), oxidative stress response (C19), glutathione metabolism (C19), regulation of cell morphogenesis involved in differentiation (C44), and cell cycle (C44 and C38). The late components show downregulation of module C30 enriched in ECM (LOX, COL1A2, TIMP2, LAMB2, COL16A1, SPARC, COL5A2, COL5A1, ERBB2IP) and collagen (COL1A2, COL16A1, COL5A2, COL5A1) after 72h and upregulation of module C5 enriched in fatty acid metabolism after 120h (ACOX1, ELOVL6, MECR, SC5DL, ABCD2, ACSL4, HPGD).

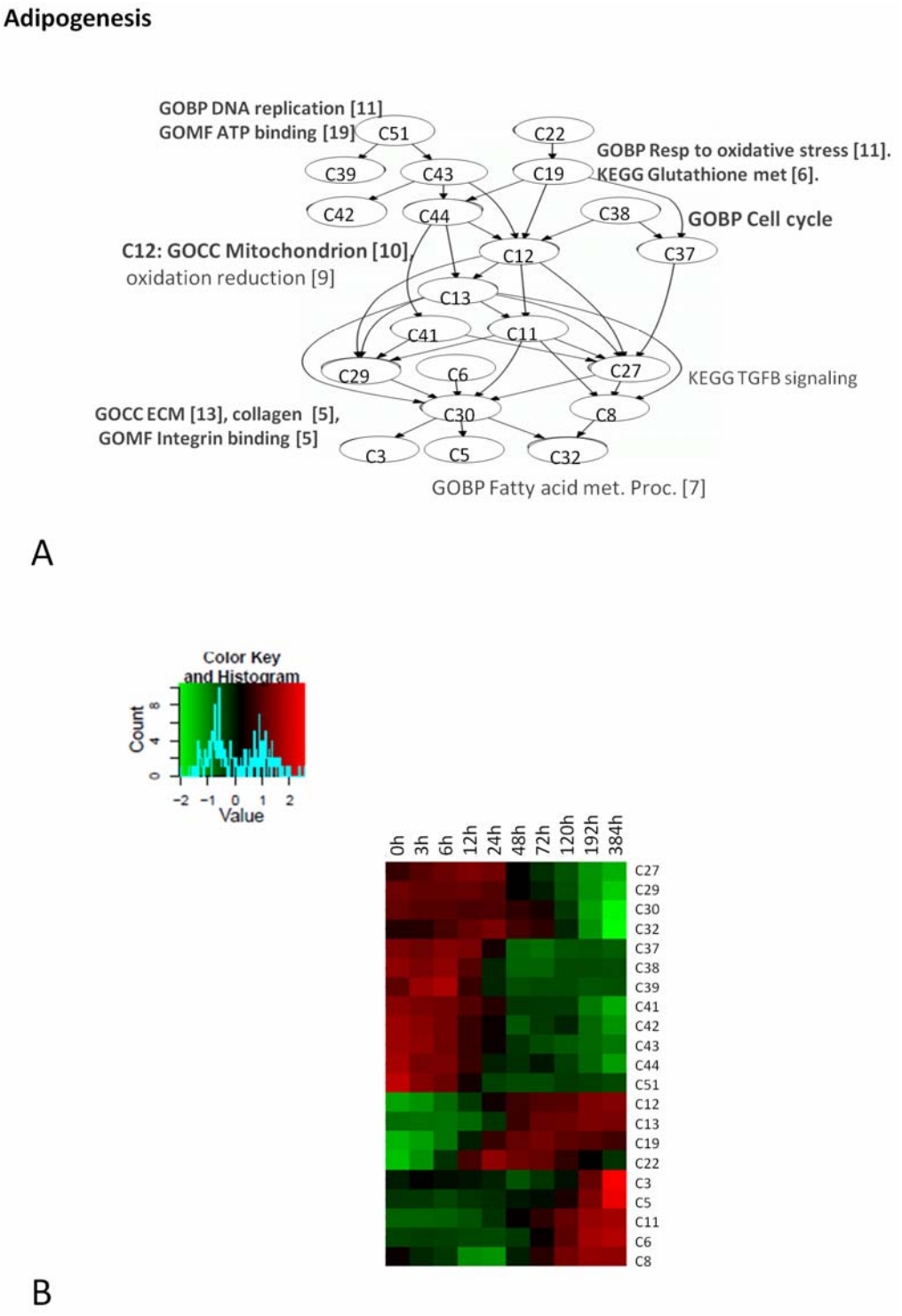


Figure 5-4. The hMSCs differentiation towards adipocytes

This figure shows the hierarchical models representing the differentiation process towards adipocytes (Panel A). Each module (node) in the model represents a number of genes that share similar expression pattern during the differentiation. The averaged expression pattern of genes in each module is represented in Panel B. Data are standardized to represent above average values in red and below average values in green. The modules were functionally annotated individually using DAVID analysis. The interactions between modules are inferred using the time delay ARACNE.

Chondrocytes differentiation model

Similarly to the two previous differentiation models, several signaling components appear at the upstream of the model (**Figure 5.5, Panel A**). The most upstream module represented by C39 is enriched in growth factor activity (FGF5, FGF2, DKK1, NRG1, NGF). The average expression pattern of this module decreases after 12h (**Figure 5.5, Panel B**). Other upstream nodes such as C40, C19 and C11 also contain cytokines/growth factor receptors (IL1R1, IFGR) and signaling components (SMAD3 and STAT1). The intermediate modules are enriched in cell cycle, regulation of cell differentiation, enzyme inhibitor activity and ECM. The more downstream nodes are enriched in oxidation reduction, ECM and calcium ion binding. As expected, the known markers of chondrocytes appear in downstream modules such as in C56 (ACAN) and C3 (COMP). Interestingly, C3 also contains hypertrophic cartilage biomarkers such as MMP13 and COL10A1 and this module have increased expression pattern overtime.

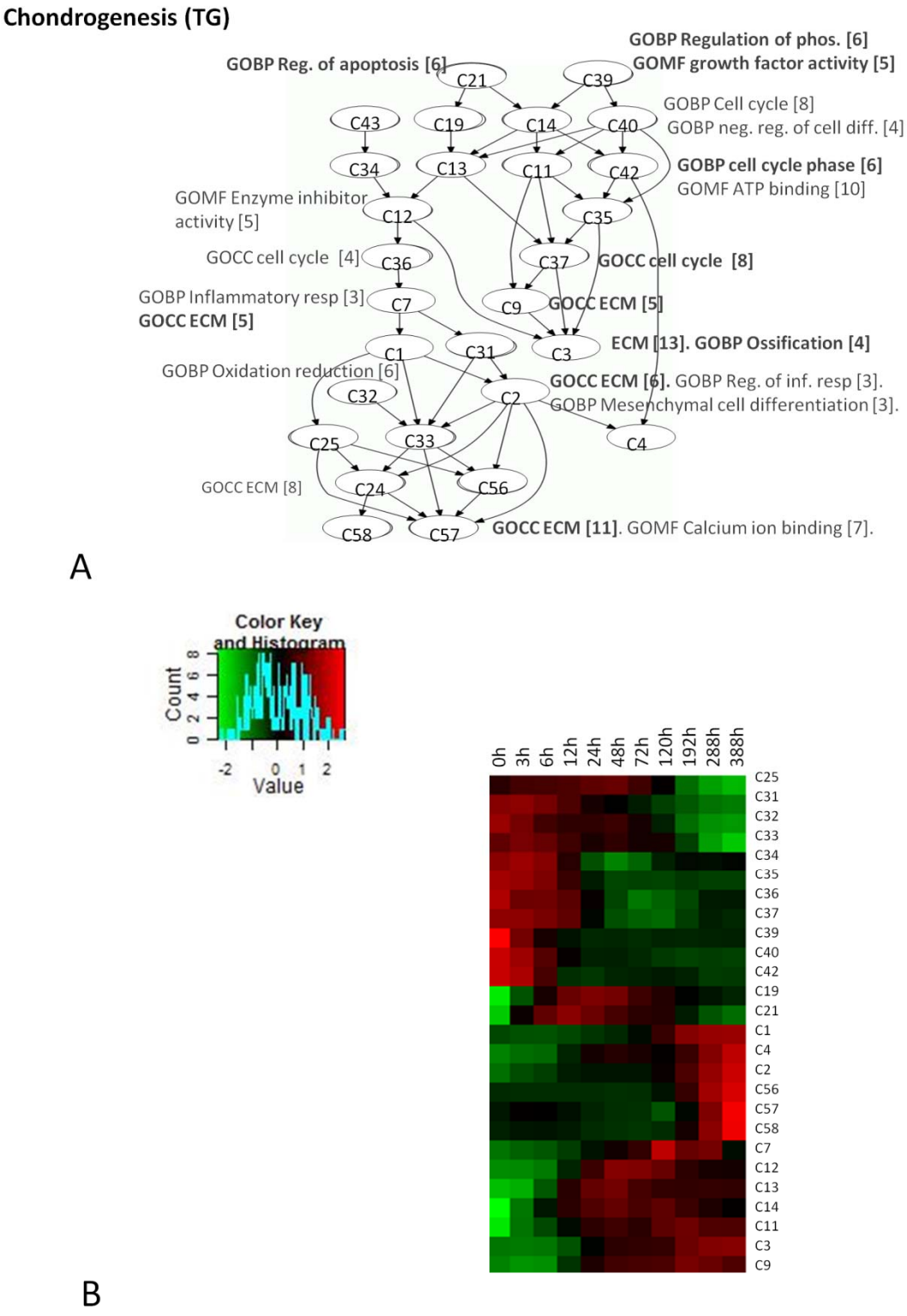


Figure 5-5. The hMSCs differentiation towards chondrocytes.

This figure shows the hierarchical models representing the differentiation process towards chondrocytes (Panel A). Each module (node) in the model represents a number of genes that share similar expression pattern during the differentiation. The averaged expression pattern of genes in each module is represented in Panel B. Data are standardized to represent above average values in red and below average values in green. The modules were functionally

annotated individually using DAVID analysis. The interactions between modules are inferred using the time delay ARACNE.

5.3.3 The transcriptional state of undifferentiated hMSC identifies model components linked to bone fragility in ageing and disease

The models described above represent a credible sequence of early signaling events followed by the activation of effector pathways known to be involved in the differentiation towards bone, cartilage or fat. Although these models represent physiological cell differentiation they can be used to formulate hypothesis on the mechanisms underlying the degeneration of the muscle-skeletal system in ageing and diseases. In order to achieve this objective, each gene module is tested for enrichment in genes correlated to age of the donors or in genes differentially expressed in relevant disease conditions (e.g. myeloma and arthritis). Using this approach, several modules have been detected within the osteogenic, adipogenic and chondrogenic differentiation models (**Figure 5.6**).

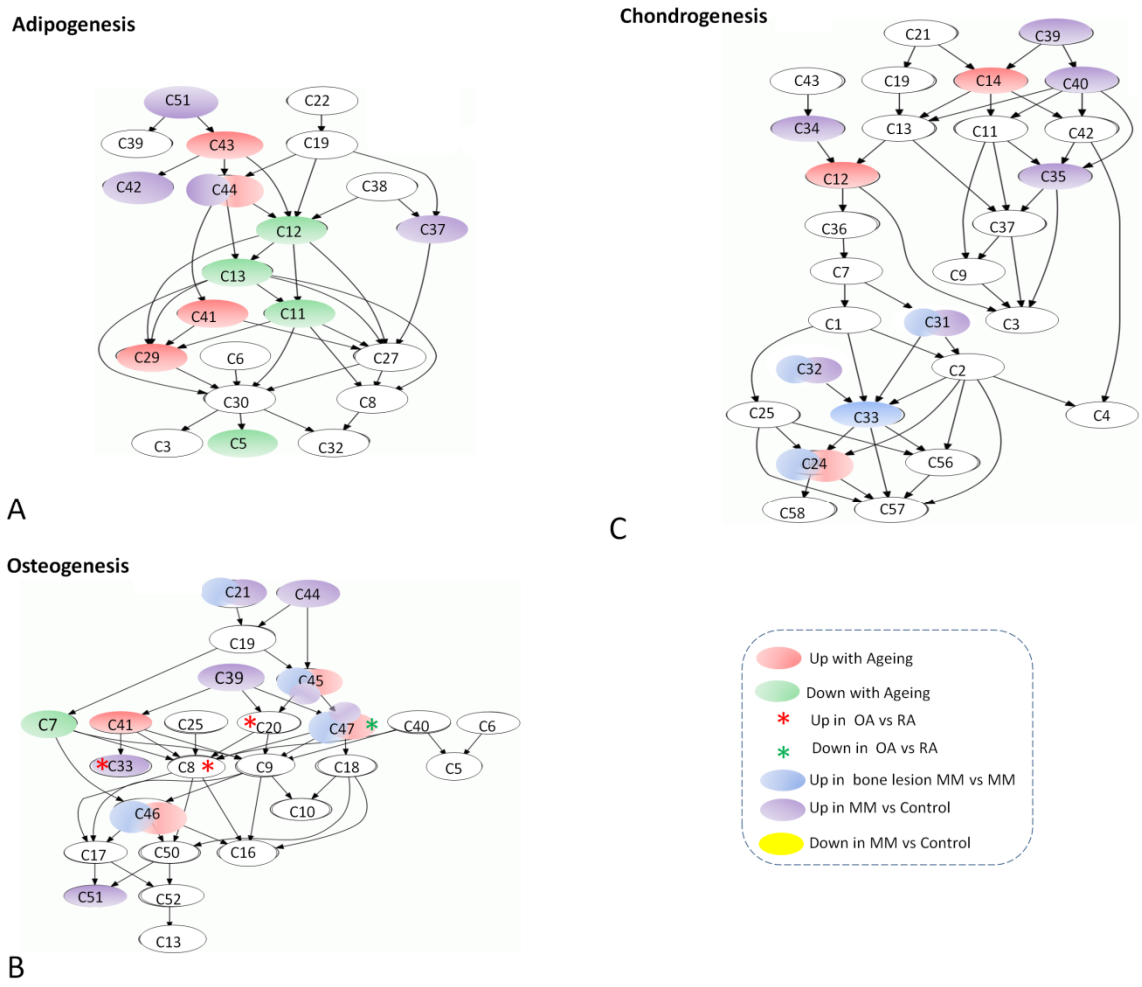


Figure 5-6. Linking the modules in the differentiation models with diseases and ageing

This figure shows the components of the adipogenic, osteogenic and chondrogenic differentiation (Panel A, Panel B, Panel C respectively) that are enriched with genes differentially expressed in disease or in ageing identified by GSEA analysis. Red and green components represent positive and negative correlation with ageing respectively. Purple and yellow components represent up and down regulation in MM compared with control. Blue components represent upregulation in MM patients with bone lesions vs. MM patients without bone lesions. Components marked with a red and green star represent upregulation and downregulation in OA patients compared with RA patients respectively.

Age dependent hMSC differentiation modules

The analysis performed using the GSEA has identified 8 modules that are enriched of genes correlated to the age of hMSCs donors.

The adipogenesis modules over-represented with the downregulated genes in ageing are functionally enriched in genes represented in the GO terms oxidation reduction and mitochondrion, whereas the genes that are upregulated during ageing are enriched in genes represented in the GO terms calcium ion binding, Wnt signaling, blood vessel development and GTPase regulator activity (**Table 5.4**). The adipogenesis and osteogenesis modules that are overrepresented with upregulated genes during ageing have a high degree of overlap (26%). The overlap includes NOTCH2, signaling components such as WNT5A, WNT5B, MAP4K5, RAC2, CXCL6, IL4R, FGF2. The overlap is also preserved at functional levels as these osteogenesis modules are functionally enriched in Wnt signaling pathway, ECM, blood vessel development, GTPase activity (**Table 5.4**). On the other hand, the modules of adipogenesis and osteogenesis models that are overrepresented with upregulated genes during ageing have a small overlap with those in the chondrogenesis model. The upregulated genes during ageing that are over-represented in chondrogenesis modules are functionally represented in oxidation reduction activity and apoptosis.

	Enrichment in the osteogenesis	Enrichment in the adipogenesis	Enrichment in chondrogenesis
Positive coregulation with ageing		C1. GOCC Endoplasmic reticulum [14], GOMF calcium ion binding [7]	
	C12. GOBP blood vessel dev. [4]	C3. GOBP blood vessel dev. [6]	
	C16. GOMF GTPase activity [4]	C7. GOMF small GTPase regulator activity [6]	
	C1. KEGG Wnt signaling [5]	C8. GOBP Wnt rec sig pathway [4]	
	C4. GOCC Extracellular region part [8]		
			C2. GOBP oxidation reduction [4]
			C3. GOBP reg. of apoptosis [5]
Negative co-regulation with ageing		C1. GOBP oxidation reduction [16]	
		C2. KEGG PPAR sig. Pathway [8], GOBP regulation of lipid storage [4]	
		C5. GOCC mitochondrion [17]	

Table 5.4. Functional annotation of genes that are changing during ageing contributing to over-representation of modules in differentiation models

Functions in red are identified from the list of genes that are positively co-regulated with ageing, whereas functions in green are identified from the list of genes that are negatively coregulated with ageing. First, second and third column represent the functional enrichement of ageing-genes over-represented in osteogenesis modules, adipogenesis modules and chondrogenesis modules respectively.

MM dependent hMSC differentiation modules

Patients affected by MM develop bone lesions that can seriously affect quality of life and survival (Chapter 1, section 1.3.2 for an overview). Here, the approach is based on the hypothesis that bone lesions are at least in part the result of a defective differentiation towards bone and cartilage. Hence, each module is tested in the differentiation models for enrichment in genes differentially regulated in MM. Two different comparisons are considered. These are: 1) hMSCs derived from MM patients with bone lesions compared to hMSCs derived from MM patients with no bone lesions; 2) hMSCs derived from healthy individuals compared to MM patients regardless of their bone lesion status.

This analysis identified modules in the three differentiation models that were significantly enriched in disease genes. The genes enriched in the osteogenesis modules are functionally enriched in growth factors, fibroblast growth factor receptor (FGF5, FGF1, FGF2), regulation of apoptosis, GTPase activity, blood vessel development (SEMA5A, ARHGAP22, TGM2, RBPJ, FGF1, FGF2, RASA1, PLAU), positive regulation of cell differentiation (LIF, INHBA, BDNF, IL6, KLF10, NRG1, FGF2) and cell adhesion (**Table 5.5**). The genes enriched in the chondrogenesis modules are functionally enriched in actin cytoskeleton, protein complex biogenesis, regulation of apoptosis (TUBB2C, SMAD3, STAT10), regulation of cell adhesion (SMAD3, NRG1, THBS1, TPM1, CITED2), angiogenesis (VEGFC, THBS1, CXCL12, ANXA2) and regulation of cell migration (**Table 5.5**). The genes in the adipogenesis modules are functionally enriched in glycolysis (LDHA,

ENO2, PFKP, HK1, PGK1), positive regulation of signal transduction, GTPase binding, blood vessel development (SHB, VEGFC, RASA1, CYR61, ANXA2) and cell adhesion (**Table 5.5**).

Osteogenesis-MM	Adipogenesis-MM	Chondrogenesis-MM
C1. GOMF growth factor activity [10], fibroblast growth factor receptor [3]		
	C2. GOCC actin cytoskeleton [10]	**C1. GOCC actin cytoskeleton [13]**
	KEGG Glycolysis /Gluconeogenesis [5]	
C4. Regulation of apoptosis [15]		C6. Regulation of apoptosis [11]
C5. GOMF GTPase activity [6]	**C7. GOMF small GTPase regulator activity [10]**	C15. GTPase activity [6]
C6. GOBP blood vessel development [8]	C23. GOBP Blood vessel dev [5]	C8. GOBP Blood vessel dev [6]
C7. GOBP regulation of cell proliferation [17], **GOBP Positive regulation of cell differentiation [7]**		
C8. GOBP enzyme linked receptor protein signaling pathway [12]		**C10. GOBP enzyme linked receptor protein signaling pathway [9]**
C13 GOBP positive regulation of signal transduction [6]	C4. GOBP positive regulation of signal transduction [8]	C27. GOBP positive regulation of signal transduction [6]
C16. GOBP Cell adhesion [9]	C35. GOBP Cell adhesion [7]	**C7. GOBP Cell adhesion [6]**

Table 5.5. Functional annotation of genes that are changing in MM contributing to over-representation of modules in differentiation models

RA-OA dependent hMSC differentiation modules

This analysis identified modules only in the osteogenesis models that were significantly enriched in RA-OA genes (**Figure 5.6**). The genes enriched in the osteogenesis modules (C33, C8, C20) are functionally with inflammatory response (IL1R, IGF2, THBS1, CD40) and transcription factor activity (SOX4, FOXM1, TRIM22).

Common osteogenic differentiation modules over-represented in MM, in RA-OA and in ageing

The two upstream modules of the osteogenesis model are over-represented in MM patients and in ageing (C45, C47). The subsets of genes contributing to the enrichment of these modules are upregulated in MM patient compared with control and these are positively coregulated with ageing. Moreover, C47 is also over-represented with genes that are upregulated in RA patients compared with OA. As C45 and C47 are downregulated after 12h and 24h, the downregulation is required during the normal osteogenic differentiation. However, the high expression level of these genes in the bone marrow of ageing individuals, MM patients and RA patients points out that these components might potentially contribute to the abnormal osteogenic differentiation.

5.3.4 Knowledge driven validations and construction of gene networks controlling hMSC differentiation

The model described above represent a high level view of the transcriptional response during differentiation. By identifying gene modules whose transcription is affected in ageing and disease, areas of the differentiation networks that are likely to be important for physiological differentiation as well as being crucial for pathogenesis have been identified. Although this allows formulation of general hypothesis, it does not provide a tool to identify regulatory networks at gene-level resolution. In order to address this challenge knowledge from the literature has been integrated into the models. In addition to increase in resolution, this approach also allows some degree of model validation.

To first validate the interaction in the differentiation models using the information available in the literature, an analysis has been performed in Metacore that allows identification of significantly over-connected interactions between the modules. This analysis has shown that

an important proportion of these interactions are significantly supported in the literature. More precisely 36%, 30% and 21% of the interactions present respectively in the adipocytes, osteogenic and chondrocytes differentiation models are also significantly supported in the literature (**Table 5.6, Table 5.7, and Table 5.8).** Several links between the subset of the modules that are over-represented in disease or ageing and the rest of modules are also supported in the literature (**Table 5.9**).

Set 1	Set 2	Actual	p-value	z-score
C11	C13	137	1.807E-12	7.833
C11	C29	148	5.669E-06	4.712
C12	C38	193	6.046E-12	7.442
C12	C13	77	3.713E-08	6.126
C12	C29	90	4.394E-04	3.584
C13	C29	169	9.086E-09	6.117
C13	C27	62	2.988E-04	3.776
C27	C37	102	3.697E-03	2.843
C29	C30	141	4.761E-06	4.765
C29	C41	178	4.766E-03	2.706
C37	C38	460	3.123E-36	13.93
C39	C51	47	3.992E-03	2.908
C43	C51	40	1.224E-03	3.387
C8	C13	55	2.491E-05	4.571

Table 5.6. Knowledge based validation of interactions in adipocytes differentiation model

The first two columns in the table Set1 and Set2 represent clusters in the differentiation model. The third column is the number of connections between genes in Set1 and Set2 that are supported in the literature.

Set 1	Set 2	Actual	p-value	z-score
C45	C20	56	3.793E-03	2.903
C45	C20	56	3.793E-03	2.903
C41	C33	353	2.269E-30	12.77
C19	C33	77	1.614E-11	7.826

C39	C41	173	3.046E-07	5.365
C45	C44	60	1.233E-04	4.064
C50	C46	7	4.135E-02	2.172
C40	C5	19	2.308E-02	2.274
C41	C8	284	5.240E-09	6.087
C20	C8	75	1.504E-02	2.31
C25	C8	42	2.945E-02	2.05
C41	C9	237	1.552E-04	3.774
C20	C9	67	2.931E-02	2.015

Table 5.7. Knowledge based validation of interactions in osteogenic differentiation model

The first two columns in the table Set1 and Set2 represent clusters in the differentiation model. The third column is the number of connections between genes in Set1 and Set2 that are supported in the literature.

Set 1	Set 2	Actual	p-value	z-score
C1	C2	31	6.658E-03	2.768
C11	C37	14	1.676E-05	5.631
C19	C21	18	3.341E-03	3.204
C2	C4	31	8.553E-06	5.163
C2	C31	71	1.955E-04	3.881
C2	C56	27	2.160E-04	4.141
C2	C24	25	1.240E-03	3.51
C31	C33	67	3.036E-03	2.963
C4	C42	10	5.734E-05	5.522
C7	C31	43	6.672E-05	4.355
C9	C37	15	1.521E-07	7.512

Table 5.8. Knowledge based validation of interactions in chondrocytes differentiation model

The first two columns in the table Set1 and Set2 represent clusters in the differentiation model. The third column is the number of connections between genes in Set1 and Set2 that are supported in the literature.

Model	Set 1	Set 2	Actual	p-value	z-score
adipogenesis	clust #13	Ageing-C11	72	1.577E-11	7.891
adipogenesis	Ageing-C13	Ageing-C11	34	2.604E-07	6.162

adipogenesis	clust #29	Ageing-C11	80	4.010E-05	4.326
adipogenesis	clust #38	Ageing-C12	160	8.533E-12	7.468
adipogenesis	clust #13	Ageing-C12	57	5.124E-08	6.219
adipogenesis	clust #29	Ageing-C12	77	7.359E-06	4.806
adipogenesis	Ageing-C13	Ageing-C12	25	3.147E-05	4.864
adipogenesis	clust #11	Ageing-C13	88	1.090E-09	6.842
adipogenesis	clust #12	Ageing-C13	40	1.294E-05	4.895
adipogenesis	clust #29	Ageing-C13	105	8.548E-05	4.053
adipogenesis	clust #51	Ageing-C43	32	2.793E-03	3.121
adipogenesis	clust #38	MM-C C37	169	3.769E-08	5.823
adipogenesis	clust #13	MM-C C44	67	9.268E-04	3.386
chondrogenesis	clust #31	C39 MM-C	45	2.912E-04	3.861
chondrogenesis	clust #37	C39 MM-C	13	6.222E-04	4.112
chondrogenesis	clust #31	C40 MM-C	52	1.814E-03	3.191
chondrogenesis	clust #37	C40 MM-C	17	3.173E-03	3.246
osteogenesis	clust #44	MMBL-MMNBL C45	18	3.134E-03	3.232
osteogenesis	clust #44	MM-C C45	29	9.068E-05	4.416
osteogenesis	clust #20	MM-C C45	25	7.134E-03	2.78
osteogenesis	clust #41	MM-C C39	158	4.836E-09	6.267
osteogenesis	clust #41	MM-C C33	166	1.231E-11	7.385
osteogenesis	clust #44	Ageing C45	20	3.019E-03	3.214
osteogenesis	clust #33	Ageing C41	87	5.805E-20	11.37

Table 5.9. Knowledge based validation of interactions in differentiation models (adipogenesis, chondrogenesis and osteogenesis) and diseases components

In order to obtain further information on the genes that may be involved in the underlying mechanism of differentiation in normal or in disease, the genes in the over-connected key upstream and downstream modules were used to construct gene regulatory networks in Metacore. These networks revealed several targets that could potentially influence the differentiation processes which are also aberrant in ageing or in diseases.

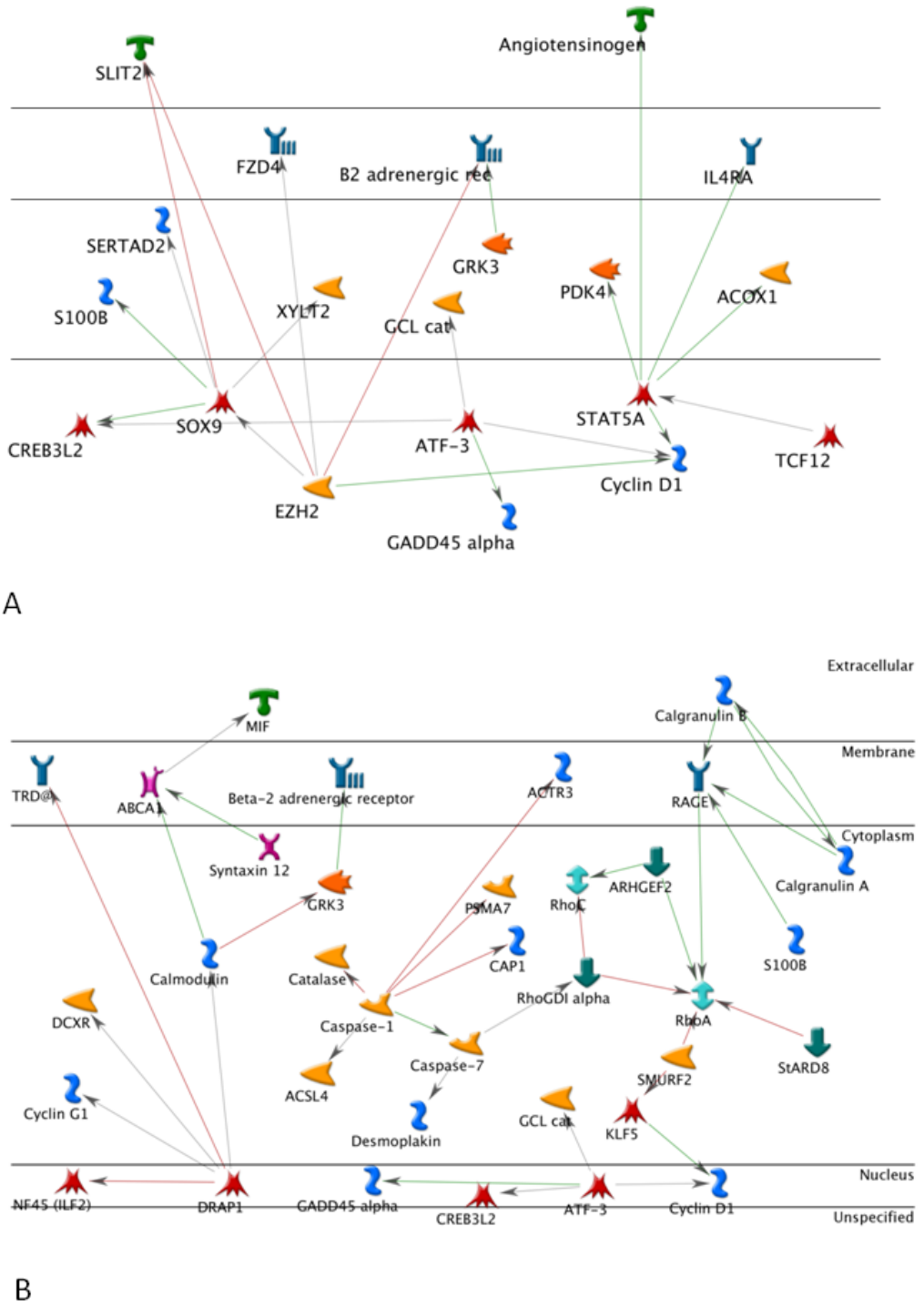


Figure 5-7. Gene regulatory networks derived from adipogenesis modules

These networks are derived from the Metacore analysis using as input the genes in the modules of ad. Panel A shows the gene regulatory network constructed using early component (C12) that are abnormal in ageing (C43, C44) and late component in the adipogenesis module (C5). This network shows that STAT5A is directly regulating several metabolic components and it is downregulated during ageing. Panel B shows the gene regulatory network constructed using early components that are abnormal in MM patients (C42, C44) and late components in the adipogenesis module (C5, C12, C32). This network reveals that small GTPases upstream components that are also abnormal in hMSCs of MM patients are interacting with several effector downstream components.

Gene regulatory networks identified by integrating knowledge in the adipocyte differentiation model

The gene regulatory networks identified by Metacore analysis from the adipocytes differentiation model suggests that the transcription factor STAT5A (gene module C12) may be directly regulating several genes encoding for metabolic enzymes represented in the downstream gene module C5 (PDK4, ACOX1 and AGT) (**Figure 5.8, Panel A**). STAT5A transcription is upregulated 24h after the differentiation stimulus (**Figure 5.4, Panel B**) and it is negatively correlated with ageing (**Figure 5.9**).

Several small GTPases (RhoA, RhoC, RhoGDI) are in the upstream modules of the adipocyte differentiation model (C42, C44) and are interacting with apoptosis genes (CASP1, CASP7), calcium binding proteins (S1008A, S1009A, S100B) and catalase (CAT, GCLC), which are on the downstream modules (C5, C32) (**Figure 5.8, Panel B**). The small GTPases are also significantly up-regulated in MM patients compared with control patients (**Figure 5.10**).

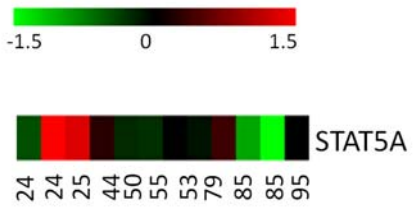


Figure 5-8. Expression pattern of STAT5A in ageing.

This figure shows the heatmap of STAT5A during ageing. STAT5A is upregulated in young individuals (red) and downregulated with ageing.

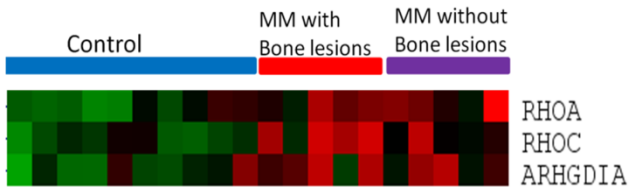


Figure 5-9. Expression pattern of small GTPases in MM.

This figure shows the heatmap of small GTPases in MM patients with and without bone lesions and control individuals. These genes are upregulated in MM patients (red) and downregulated in control patients (green)

Gene regulatory networks derived from the chondrocytes differentiation modules

Consistent with several findings, known regulators of chondrogenic differentiation including transcription factors SOX9 and TRPS1 are directly connected to several genes in the three downstream modules (C2, C3, C57) containing articular and hyperthrophic chondrocytes biomarkers (**Figure 5.11, Panel B**). These modules are also potentially regulated via SMAD3 activation (in C40), (**Figure 5.11, Panel A**), which is upregulated in Myeloma patients compared with control (**Figure 5.12)**.

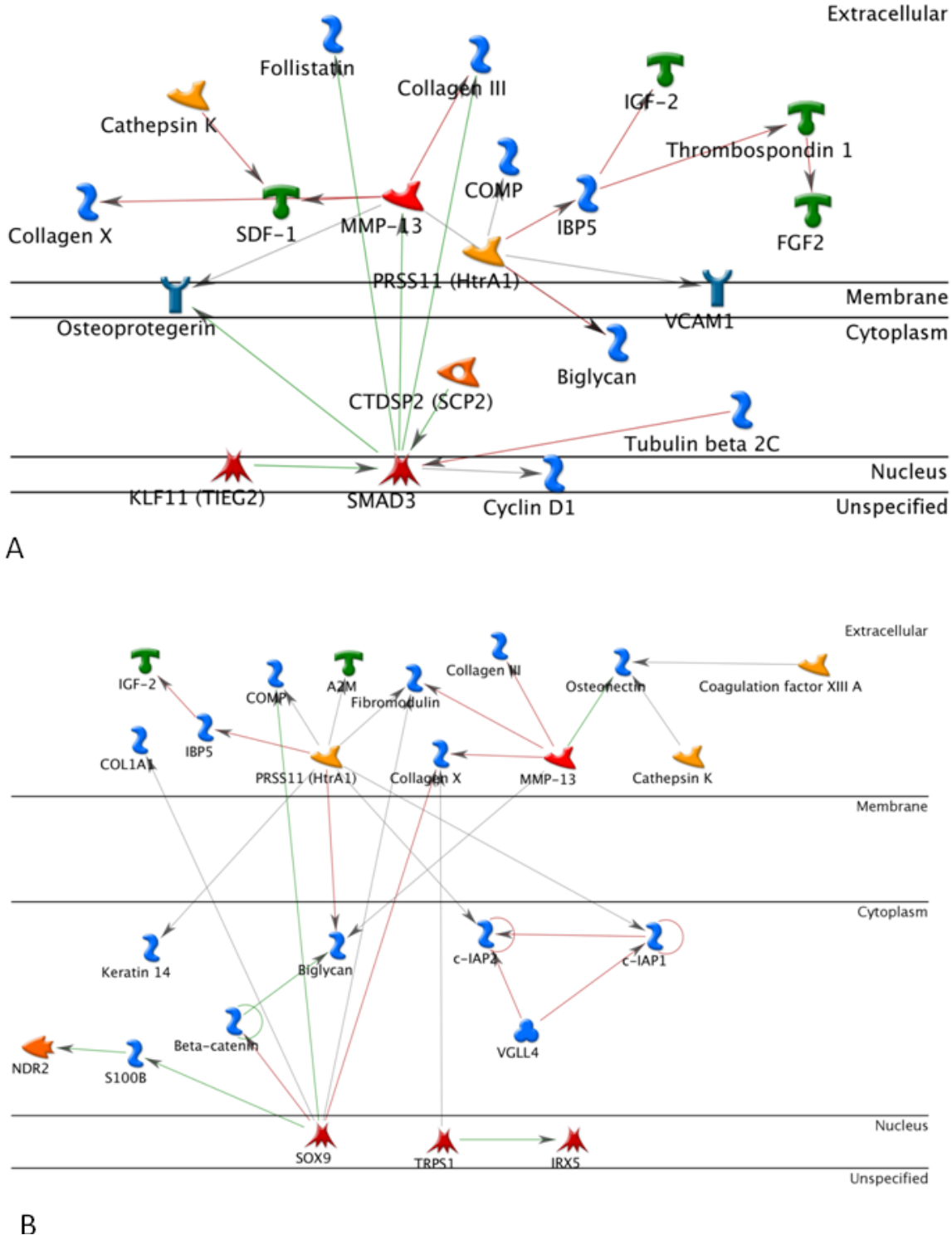


Figure 5-10. Gene regulatory networks derived from chondrogenesis modules.

These networks are derived from the Metacore analysis using as input the genes in the modules of chondrogenesis. Panel A shows the gene regulatory network constructed using early components that are abnormal in MM and late components of the chondrogenesis module (C3). Panel B shows the gene regulatory network constructed using early components and late components in the chondrogenic modules (C2, C3, C57).

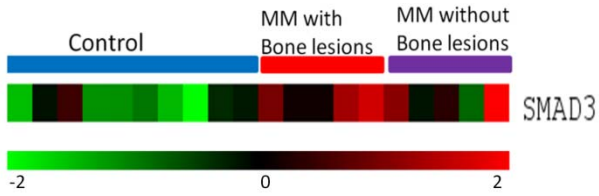


Figure 5-11. Expression profile of SMAD3 in MM.

This figure shows the heatmap of SMAD3 in MM patients with and without bone lesions and control individuals. It shows that SMAD3 is upregulated in MM patients (red) and downregulated in control individuals (green).

Gene regulatory networks derived from the osteogenic differentiation modules

Two potential key players have been identified in the upstream modules that might have an influence in the most downstream modules in the osteogenic differentiation model. These are RBPJ component of notch signaling (in C39) (**Figure 5.13, Panel A**), and ETV5 (in C41) (**Figure 5.13, Panel B**), which are downregulated after 12 hour and 24h respectively. Interestingly, these two genes have higher expression in myeloma patients with bone lesions than patients without bone lesions and control individuals. Therefore, this observation leads to the hypothesis that knocking down these genes in MM patients might enhance the differentiation towards bone cells.

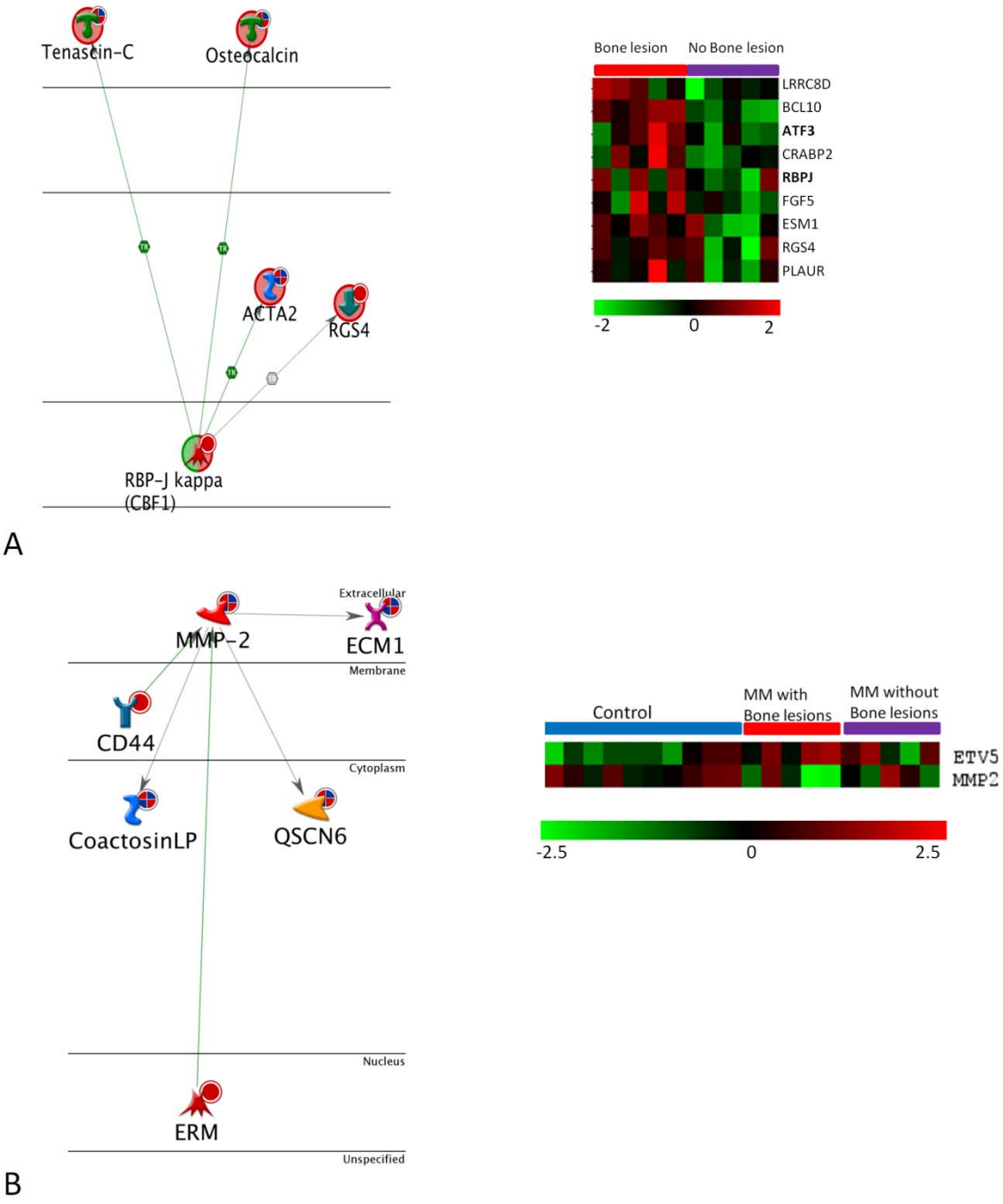


Figure 5-12. Gene regulatory networks derived from osteogenesis modules

These networks are derived from the Metacore analysis using as input the genes in the modules of db. Panel A shows the gene regulatory network constructed using early components that are abnormal in MM and late components of the db modules (C51, C52). This network shows that RBP-J is a potential regulator of late components including osteocalcin. Panel B shows the gene regulatory network constructed using early components and late components in the db modules (C2, C3, C57).This network shows that ERM (ETV-5) is a potential regulator of late components including MMP2.

5.3.5 Identification of gene regulatory networks integrating biological knowledge, gene co-expression and disease linkage

The procedure identified several networks that consist of genes highly co-regulated in the normal hMSCs differentiation process and at the same time these are differentially expressed in disease or ageing. These modules revealed a number of novel potential targets that may play an important role in healthy and pathological scenarios.

One of the most interesting modules derived from the integration of the osteogenic differentiation and myeloma data is significantly enriched in response to hormone stimulus, signaling pathways, blood vessel development, regulation of apoptosis, ECM, cell migration, ossification and Jak-Stat signaling pathway (**Table 5.10**). The module revealed that MYC plays a central role and it is interacting with several genes that are aberrantly expressed in MM patients (**Figure 5.14**). Overexpression of Myc in hMSCs from healthy patients treated with BMP2 has been shown to enhance osteogenic differentiation [38]. Myc is downregulated in the MM patients compared with control individuals. Hence, an experiment overexpressing Myc in hMSCs derived from MM patients represents an excellent setting to test whether it could enhance the osteogenic differentiation in these patients.

The module with the highest score derived from the integration of ageing and adipocytes differentiation datasets is functionally enriched with response to hormone stimulus, regulation of apoptosis, TCA cycle, small GTPases mediated signal transduction, adherens junction, lipid homeostasis, mitochondrion (**Table 5.11**). This module contains several known factors required for adipogenesis differentiation including PPARG, RXRA and CEBPA. Most of the genes identified connecting several of these genes and RAC1 also belongs to upstream module C44 in ad (**Figure 5.15**). This is consistent with the previously identified gene regulatory networks using the hierarchical models and Metacore. In this

analysis, several small GTPases have been identified to be upstream of the adipocytes differentiation model and connected with effector functions.

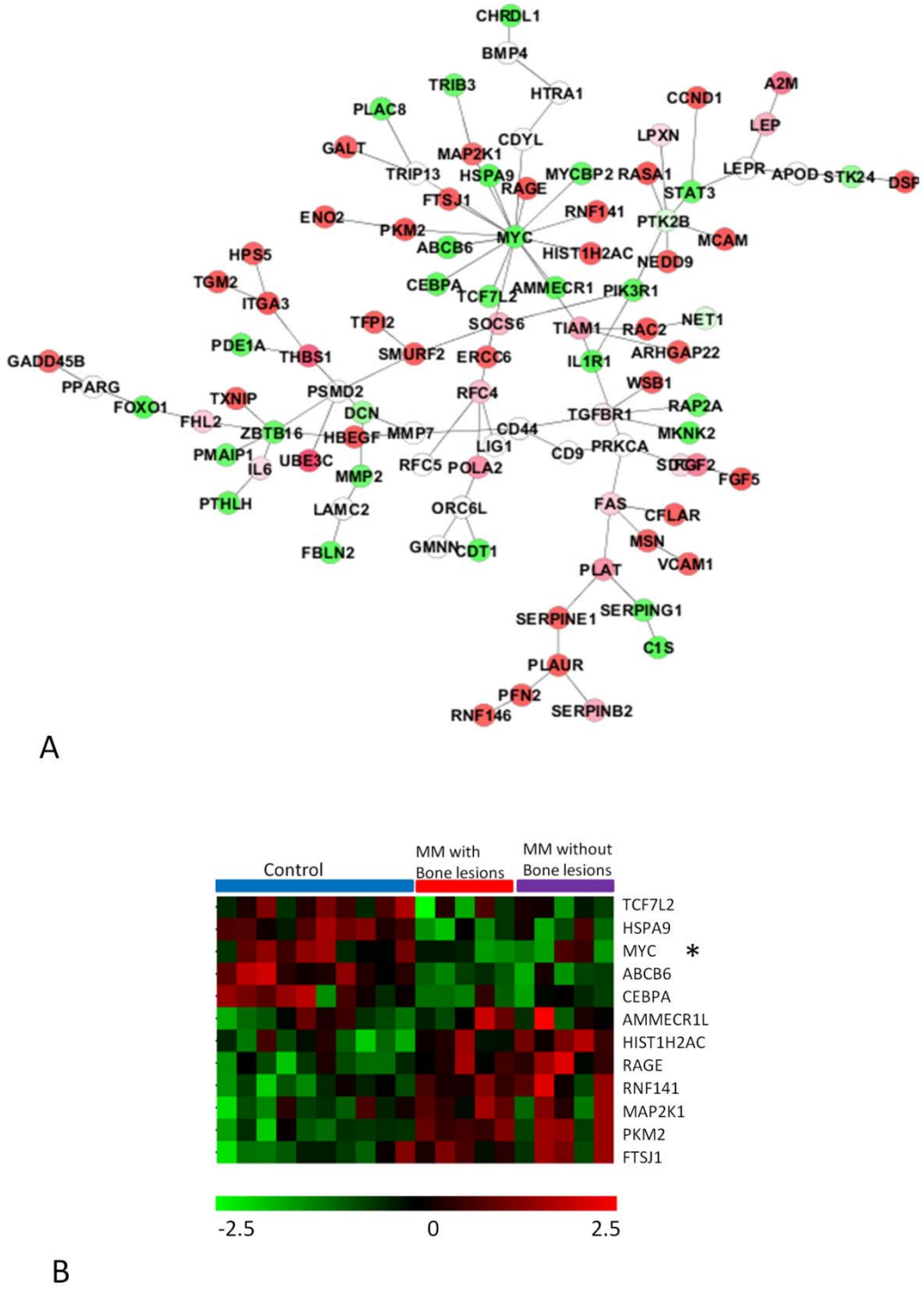


Figure 5-13. Module identified integrating interaction networks, osteogenic dataset, and MM patients

This module has been identified using the novel modularization approach that allows integration of interaction networks, osteogenic dataset, and MM dataset (Panel A). It shows that Myc is a hub gene

and its expression is down-regulated in hMSCs of MM patients. The genes interacting with Myc are also significantly different between MM and control hMSCs (Panel B).

Module enriched from osteogenesis-MM
GOBP Response to hormone stimulus [19]
GOBP Blood vessel development [14]
GOBP Enzyme linked receptor protein signaling pathway [14]
GOBP Regulation of apoptosis [20]
GOCC ECM [23]
GOBP Positive regulation of cell motion [8], GOBP cell migration [11]
GOBP Response to oxidative stress [6]
GOBP ossification [5], GOBP osteoblast differentiation [3]
KEGG Jak-STAT signaling pathway [7]

Table 5.10. Functional annotation of modules identified by integrating interaction networks, osteogenic differentiation data and MM

Functional annotation of the module from ad-ageing
GOBP Response to hormone stimulus [11]
GOMF transcription factor binding [17]
GOBP Regulation of apoptosis [16]
KEGG TCA cycle [5]
GOBP Small GTPases mediated signal transduction [9]
GOCC adherens junction [7]
GOBP lipid homeostasis [4]

GOCC mitochondrion [14]

Table 5.11. Functional annotation of modules identified integrating interaction networks, adipocytes differentiation data and ageing.

Functions in bold are significantly enriched (10%FDR).

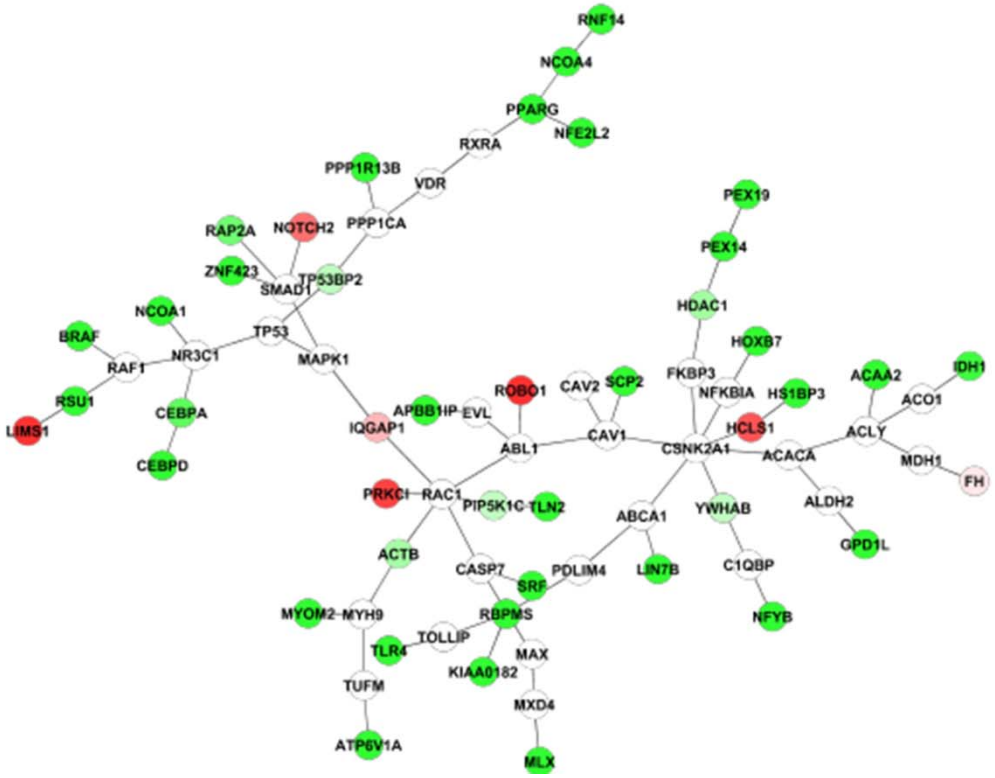


Figure 5-14. Gene regulatory networks derived from interaction networks, adipocytes dataset, and ageing.

This module has been identified using the novel modularization approach that allows integration of interaction networks, adipocytes dataset, and ageing data (Panel A). This panel shows that several markers of adipocytes are down-regulated in ageing. This module also shows that RAC1 plays a role of a hub connecting several components, which place RAC1 as a potential regulator.

5.3.6 Direct integration of RNAi osteoblast differentiation phenotypic screenings into network discovery procedure

Recently, a novel study has reported 54 potential targets that can enhance the osteogenic differentiation measured with ALP activity [45]. In order to identify potential regulators of these targets, the modularization method described in the previous chapter has been applied.

In this application, the node scores are derived from the RNAi screening and the edge scores are derived from the correlation between genes in the osteogenic differentiation data. One module has been identified from this analysis (**Figure 5.16**), which contains several potential regulators such as chromatin modifiers SIRT2 and HDAC6, the GTPase HRAS and the growth factor receptor EGFR. HRAS is a hub gene in the module and interestingly a previous study has reported that MSCs from mouse stimulated with strontium differentiates towards bone cells though HRAS signaling pathway and knock-down of HRAS causes reduced ALP activity and mineralization. Indeed, EGFR signaling [242] and SIRT2 [243] have been reported to reduce the differentiation potential of MSCs from mouse towards bone. The hub of the module HRAS has been previously shown to regulate negatively the differentiation of mouse derived MSCs induced with strontium [244]. In this study, the authors have reported that the knock down of HRAS in these cells has lead to reduced mineralization and ALP activity 14 and 21 days after the initiation of differentiation towards bone cells [244]. Hence, HRAS might also be an important player in hMSCs and a regulator of the other validated targets in the module.

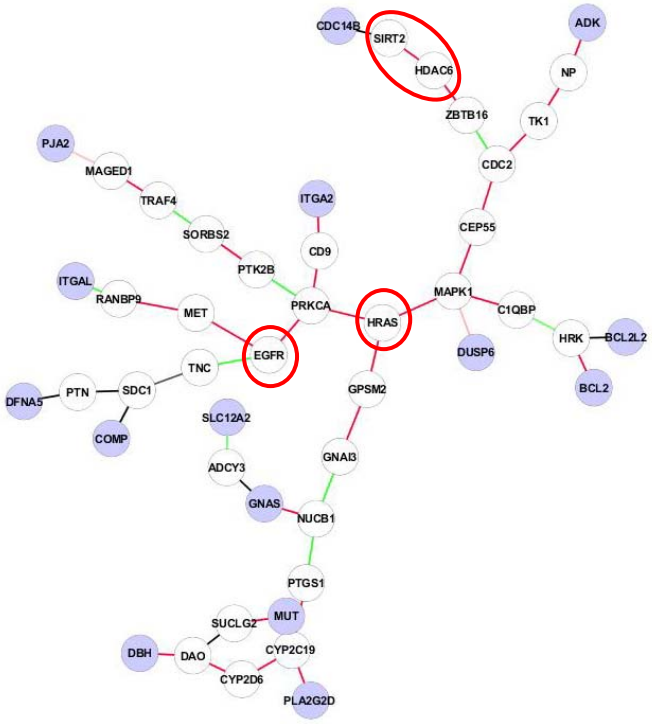


Figure 5-15. Module derived from direct integration of RNAi phenotypic screenings and normal differentiation of hMSCs.

Knock-down of genes represented in purple has leaded to enhanced ALP activity [45]. Edges represented in red are positively coregulated during the time-course of normal hMSCs differentiation, whereas edges represented in green are negatively coregulated in the same data.

5.3.7 Linking the siRNA screening and the differentiation model with the disease condition leading to bone deformation

In the siRNA screening study [45], 7 out of the 54 potential targets are also significantly different in the osteogenic differentiation timecourse data. The authors have done further studies for 12 targets out of 54 by measuring mineralization in which 6 have proven to enhance the differentiation towards bone cells. 3 out of these 6 targets are also in our model. One of these DUSP6, belonging to the one of the most upstream module C44 (Figure 5.3, Panel B), is also significantly upregulated in the MM patients compared with healthy individuals (**Figure 5.16**). Another hit is TBX3 (C13 in the model) and it is significantly

upregulated in ageing. Therefore, inhibiting DUSP6 in hMSCs from MM patients and TBX3 in elderly patients can enhance the differentiation towards octeogenic cells.

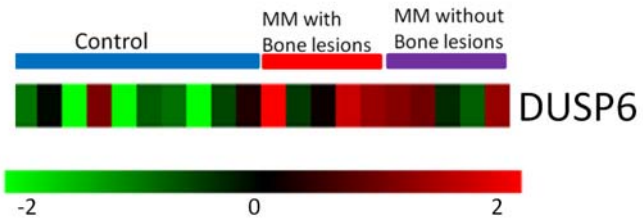


Figure 5-16. Expression pattern of DUSP6 in MM patients

This figure shows the expression pattern of DUSP6 in MM patients with and without bone lesions and control individuals. DUSP6 is downregulated in control individuals (green) and upregulated in MM patients.

5.4 DISCUSSIONS

5.4.1 Inhibition of Notch leads to enhanced osteogenesis in MM patients with bone lesions

The inferred osteogenesis model lead to the hypothesis that downregulation of RBPJ is required during osteogenesis and higher levels of RBPJ is observed in the MM patients with bone lesions compared with control individuals. RBPJ is a transcriptional regulator that plays a central role in Notch signaling, a signaling pathway involved in cell-cell communication that regulates a broad spectrum of cell-fate determinations. The potential role of RBPJ in bone formation is supported by the observation that constitutively active Notch has been shown to suppress the endochondral bone formation and osteoblastic differentiation [245, 246]. Moreover, in a recent study the knock down of Notch has been shown to enhance osteogenic differentiation from hMSCs, which were collected from Myeloma patients with bone lesions [107].

5.4.2 The role of Myc in the MM patients

Myc is a hub in the network module obtained by integrating interaction networks, osteogenesis and MM datasets. This is consistent with the observation that overexpression of Myc in hMSCs from healthy patients treated with the differentiation factor BMP2 has been shown to enhance osteogenesis [38]. Interestingly, Myc is downregulated in the MM patients compared with control individuals. Hence, an experiment overexpressing Myc in stem cells derived from MM patients represents an excellent setting to test whether this could enhance the osteogenesis process in these patients.

The same network locate the gene MAP2K1 in direct connection with MYC. Continuous inhibition of MAPK signaling pathway with the specific inhibitor MAP2K1 has been shown to promote early osteoblastic differentiation and mineralization [247]. The fact that MM

patients have higher level of MAP2K1 compared with control patients suggest that MAP2K1 might be at least in part responsible of the abnormal osteogenesis capacity in these patients.

5.5 Conclusions

This study allowed constructing the detailed gene regulatory networks describing the adipogenesis, osteogenesis and chondrogenesis model. Systems biology approaches have been very useful to achieve this task. The inferred models could be used by the community to identify potential targets to better understand the normal differentiation process and in disease conditions. Indeed, a number of potential testable targets have been identified and these might enhance osteogenesis process in multiple myeloma patients with bone lesions and in ageing. Further work is required for experimental validations of these targets. In-vitro tests followed by animal models represent an excellent potential for the validation of these targets and understanding their role in bone formation.

CHAPTER 6: GENERAL DISCUSSION

6.1 A new perspective in understanding muscle wasting in COPD patients

In this thesis, the effectiveness of systems biology approaches were shown for the integration of various sources of information that allowed us to tackle several questions including the role of systemic and local inflammation in skeletal muscle wasting in COPD patients. Application of such approaches lead to the first model in the community underlying the molecular abnormalities in the skeletal muscle of these patients. The work carried out in Chapter 2 and Chapter 3 has shown that muscles from COPD patients have a lack of response in tissue remodelling and bioenergetic pathways. One possible mechanism for the abnormalities observed in COPD muscle has been previously hypothesized to be based on systemic inflammation [72]. The work carried out in Chapter 3 has not identified a link between the systemic inflammation, and the muscle transcriptional state. Instead, a potential epigenetic mechanism that might drive the skeletal muscle wasting in these patients has been proposed. Therefore, this analysis has lead to the hypothesis that treatments based on histone deacetylase inhibitors might be a potential direction for developing effective drug treatments for muscle wasting in COPD. Indeed, a recent study using myotubes from dystrophic mice treated with VPA for 24h lead to hypertrophy and decreased apoptosis [248]. Several future studies are required to establish the usefulness of this treatment and the safety. Further experimental models could be studied to elucidate the effectiveness of such treatments before setting up clinical trials. For instance, a smoking mouse model treated with and without VPA

and exercise training or a hypoxic mouse model could provide additional experimental validation for disease treatment. Various doses and time points should also be studied for the optimization of the treatment. Multi-level measurements would be necessary to understand aberrant molecular mechanisms and link these with the phenotypic changes and the response to treatment. This could include physiological measurements and omics datasets such as microarrays, metabolomics and proteomics. The analysis of such data could be carried out as inspired by the approaches presented in this thesis using reverse engineering and modularization techniques for integration of the various data types. In the case of the animal models being successful further clinical studies could be carried out. It would certainly be appropriate given the results of this thesis to study epigenetic markers in COPD patients before and after drug treatments combined with exercise training, using technologies like ChIP-seq. Sequencing technologies would be an excellent tool to identify specific genetic traits that lead to muscle wasting in COPD patients as only 40% suffer from it. This strategy might even allow the identification of sub-patient groups and predicting the likelihood of suffering from muscle wasting. Such clinical trials would require the collaboration between clinicians, wet lab scientists and systems biologists. Collaboration efforts from different backgrounds are absolutely crucial for the developments of effective treatments.

6.2 Future work and challenges in using stem cells for degenerative diseases

6.2.1 From normal osteogenic and chondrogenic differentiation models to bone and cartilage deformation

In this thesis, models highlighting precise molecular changes during the differentiation process were developed from normal hMSCs using integrative systems biology approaches. Constructing such models from disease patients associated with bone lesions and cartilage differentiation aids allows understanding the disease mechanisms and identifies potential key

players. As such data is currently not available, generating time-course microarrays from hMSCs of these patients represent an excellent future experimental setting. For the constructions of the models, the same approach presented in Chapter 5 could be performed. Potential key players that play a role during normal osteogenesis *in vitro* were identified and these seem to be abnormal in ageing and MM bone marrow. Validation of these targets *in vitro* could lead to identify targets that could enhance differentiation in disease condition. For this, overexpression of Myc, knock-down of DUSP6 in cells collected from MM patients and knock down of TBX3 in cells collected from individuals representing young and old individuals represent a possible experimental setting. If these experiments can indeed lead to enhanced osteogenesis process in these patients, it would be a great opportunity to carry out time-course microarrays. This will allow understanding the response to these treatments in the process of differentiation in these patients.

6.2.2 From in-vitro studies towards effective treatments in diseases with bone lesions or bone loss

If the key players that are hypothesized to enhance osteogenic differentiation process in this thesis are successfully validated *in-vitro*, animal models would represent an excellent setting to investigate the potential of these targets and pave the way for future clinical trials. Indeed, a recent study has shown that inhibition of Notch signaling has enhanced the osteogenesis process from MM patients in vitro [107]. However, inhibition of Notch signaling through injection with DAPT in MM mice had no remarkable effect on tumour burden in the bone, on the number of the osteoblasts and bone formation. The authors concluded that for a more effective treatment other anti-MM drugs could be combined with Notch inhibitors. These experiments are encouraging and the models developed in this thesis could be used in a similar manner for further validations *in vitro* from patients and in animal models. The

approaches I have used could also shed light on sustained bone loss in ageing or in osteoporosis patients. However, there is a lack of transcriptomics datasets and studies from the bone marrow microenvironment of osteoporosis patients. Studies investigating the osteogenesis process in these patients and capturing molecular changes during the differentiation process would be extremely useful for the community.

For clinical trials, there could be several options including direct delivery of the therapeutic gene, or hMSCs cell based delivery combined with gene targeting. There have been several encouraging clinical trials for stimulating bone formation using hMSCs treatment. The allogenic transplantation of bone marrow derived MSCs to children with severe osteogenesis imperfecta has shown that MSCs can engraft after transplantation and improvements on growth velocity were also observed [249]. Although cell based therapy using hMSCs has very important potential for degenerative diseases and several encouraging phase I and phase II clinical trials were carried out [250], there are still concerns in the safety of their usage.

6.2.3 The challenges using hMSC cell based therapy

Some of the major concerns are based on long-term culture and genetic manipulation of hMSC cells as this might result in a malignant transformation [251]. In the usage of autologous hMSCs treatment for disease or ageing, there is a potential issue with the loss of function of the cells. For the allogenic transfer the cells could be harvested from healthy donors, but the immunomodulatory effects needs to be taken into account. Long term follow up of patients are also necessary to understand the future implications of the treatments. Safety measures should be well established before these treatments would be made available.

For the use of allogenic stem cells, stem cells banks represent an important potential. However, the limited number of cells to isolate from the bone marrow coupled with the loss of cells after storage makes this option undesirable. Adipose derived stem cells represent a

good alternative to bone marrow derived stem cells as the sample collection is less invasive, and a larger number of cells could be collected [252]. Several studies have been investigating the differences between adipose derived and bone marrow derived stem cells [253, 254]. Although adipose derived stem cells can also differentiate towards bone and cartilage, the differentiation is less effective compared with bone marrow derived stem cells [253, 254]. More research is required in this field to enhance the differentiation of adipose derived stem cells and take advantage of these features.

6.3 Cutting edge translational medicine and systems biology

This thesis demonstrates the effectiveness of systems biology approaches for studying complex human diseases. A number of independent study cases that are associated with musculoskeletal disorders have been studied. The main aim was to identify molecular mechanisms underlying the abnormalities observed in various components of the musculoskeletal system by using systems biology approaches. Understanding disease mechanisms is crucial for the development of effective treatments. Indeed, in recent years with the development of the genome-wide analysis tools, translational medicine has been a very active research area. The ultimate aim is to elucidate molecular mechanisms underlying complex diseases for the identification of biomarkers, patient classification, development of effective treatments, risk assessment of treatments, and prediction of disease progression. Application and development of system biology approaches are crucial to address these challenges by providing mechanistic models, which would allow simulation of pathological processes and predict the response to various interventions. Advances in this field will lead to optimized treatments specific to each individual.

Currently, several projects are funded for development of personalized medicine. For instance, the Virtual Physiological Human initiative (VPH-I) is aiming at developing patient

specific models for various diseases by using a multi-level integrative approach [255]. The initiative consists of a number of projects and relies on collaborative multi-disciplinary efforts involving clinicians, scientists from different backgrounds and industry partners. In the context of respiratory diseases, the project AIRPROM is aiming to developing patient-specific models to predict the impact on the airways in response to various changes occurring at the gene, cell or tissue level. To achieve this ambitious goal, they are developing a systems biology framework, which will allow integrating genomics, transcriptomics, proteomics, physiological measurements and imaging data from asthma patients. The integration of multi-level datasets is key for the success of this type of projects and the development of patient-specific medicine. With the development of next generation sequencing, more information could be measured including individual DNA sequences (SNPs etc) and more detailed RNA species including miRNAs and other small RNA species [256]. Data generation using these new technologies has until recently been very costly, however it is becoming more affordable. Certainly in the near future more studies will take advantage of these technologies. To tackle the challenge of analysing such vast amount of data of various types, development of new platforms (software), frameworks and integrative analysis strategies are required. Platforms with graphical user interface (GUI) are especially important as it provides tools for a broad range of users. There are publicly and commercially available tools such as Taverna [257] and BioXM [221] that allows integration of pipelines of analysis strategies, which could be then used by others. In recent years, efforts have also been made to integrate data analysis strategies into these tools. For instance, the strategy presented in chapter 2 and chapter 3 based on high level pathway analysis and the reverse engineering approach for integrative analysis is available in BioXM [221]. Although the novel modularization approach presented in chapter 4 is not yet integrated in these tools, a GUI implementation would provide access for an additional user base.

APPENDIX

Supplementary Material Chapter 2

PC	KEGG	SH mean	TH mean	diff	Functional Group
PC1	Focal adhesion	1.53	-3.02	4.55	tissue remodeling
PC1	Pyruvate metabolism	1.82	-2.35	4.17	Metabolism
PC1	Glycolysis / Gluconeogenesis	1.44	-2.54	3.98	Metabolism
PC1	Insulin signaling pathway	1.40	-2.29	3.69	Signaling
PC1	Adipocytokine signaling pathway	-1.39	2.17	3.56	Inflammation
PC1	Prostate cancer	0.76	-2.66	3.43	Signaling
PC1	ECM-receptor interaction	-1.77	1.56	3.33	ECM
PC1	Alanine and aspartate metabolism	1.23	-2.08	3.31	Metabolism
PC1	Cytokine-cytokine receptor interaction	0.83	-2.42	3.25	Inflammation
PC1	Wnt signaling pathway	1.34	-1.91	3.25	Signaling
PC1	Fatty acid metabolism	1.52	-1.72	3.24	Metabolism
PC1	Chronic myeloid leukemia	0.85	-2.30	3.15	
PC1	Melanoma	0.63	-2.51	3.14	Signaling
PC1	Valine, leucine and isoleucine biosynthesis	1.25	-1.77	3.02	Biosynthesis
PC1	Ribosome	1.86	-1.16	3.02	
PC1	MAPK signaling pathway	1.08	-1.89	2.97	Inflammation
PC1	ErbB signaling pathway	0.64	-2.32	2.97	tissue remodeling
PC1	GnRH signaling pathway	0.92	-2.01	2.93	tissue remodeling
PC1	Renal cell carcinoma	0.91	-1.99	2.91	
PC1	Propanoate metabolism	1.04	-1.84	2.88	Metabolism

PC1	Complement and coagulation cascades	0.90	-1.98	2.88	
PC1	Pancreatic cancer	1.09	-1.79	2.88	Signaling
PC1	Colorectal cancer	0.72	-2.09	2.81	
PC1	Phosphatidylinositol signaling system	0.80	-1.92	2.72	Signaling
PC1	Nicotinate and nicotinamide metabolism	-0.22	2.47	2.69	Metabolism
PC1	B cell receptor signaling pathway	1.00	-1.68	2.68	Inflammation
PC1	Glioma	0.56	-2.12	2.67	
PC1	Natural killer cell mediated cytotoxicity	0.79	-1.84	2.64	
PC1	Notch signaling pathway	1.15	-1.47	2.62	Signaling
PC1	T cell receptor signaling pathway	0.90	-1.59	2.49	Inflammation
PC1	Toll-like receptor signaling pathway	0.84	-1.62	2.46	Signaling
PC1	Type II diabetes mellitus	0.37	-2.09	2.46	
PC1	Melanogenesis	0.64	-1.80	2.44	
PC1	Long-term potentiation	0.83	-1.55	2.38	
PC1	VEGF signaling pathway	0.70	-1.62	2.33	Signaling
PC1	Jak-STAT signaling pathway	0.61	-1.72	2.32	Inflammation
PC1	Small cell lung cancer	1.12	-1.20	2.32	
PC1	1- and 2-Methylnaphthalene degradation	0.69	-1.59	2.28	
PC1	Acute myeloid leukemia	0.64	-1.64	2.28	
PC1	mTOR signaling pathway	0.45	-1.82	2.27	Signaling
PC1	SNARE interactions in vesicular transport	0.61	-1.66	2.27	
PC1	Gap junction	1.08	-1.16	2.24	tissue remodeling
PC1	Benzoate degradation via CoA ligation	1.03	-1.19	2.22	
PC1	Long-term depression	0.84	-1.38	2.22	
PC1	Glycerophospholipid metabolism	0.63	-1.55	2.18	Metabolism
PC1	Fc epsilon RI signaling pathway	0.60	-1.54	2.14	Signaling

PC1	Aminoacyl-tRNA biosynthesis	1.07	-1.04	2.12	Biosynthesis
PC1	Butanoate metabolism	0.90	-1.19	2.09	Metabolism
PC1	Bladder cancer	0.30	-1.76	2.05	Signaling
PC1	TGF-beta signaling pathway	-0.52	-2.56	2.04	Signaling
PC1	Arginine and proline metabolism	0.19	-1.78	1.97	Metabolism
PC1	Bile acid biosynthesis	0.54	-1.41	1.95	Biosynthesis
PC1	Non-small cell lung cancer	0.70	-1.21	1.91	
PC1	Endometrial cancer	0.90	-1.00	1.91	
PC1	Porphyrin and chlorophyll metabolism	1.12	-0.73	1.85	Metabolism
PC1	Maturity onset diabetes of the young	-0.89	0.94	1.83	
PC1	Alkaloid biosynthesis II	0.72	-1.06	1.78	Biosynthesis
PC1	3-Chloroacrylic acid degradation	0.44	-1.33	1.78	
PC1	Cysteine metabolism	0.96	-0.75	1.71	Metabolism
PC1	Huntington's disease	0.25	-1.18	1.43	
PC1	Glyoxylate and dicarboxylate metabolism	0.51	-0.89	1.40	Metabolism
PC1	Taste transduction	0.44	-0.96	1.40	
PC1	Pentose and glucuronate interconversions	0.68	-0.63	1.31	
PC1	Glutamate metabolism	0.33	-0.97	1.29	Metabolism
PC1	Galactose metabolism	-1.16	0.03	1.19	Metabolism
PC1	ABC transporters - General	0.47	-0.71	1.18	
PC1	Biosynthesis of steroids	-0.11	0.94	1.05	Biosynthesis
PC1	Parkinson's disease	-0.58	0.45	1.03	
PC2	Tight junction	-0.61	2.26	2.88	tissure remodeling
PC2	Tyrosine metabolism	1.05	-1.74	2.79	Metabolism
PC2	p53 signaling pathway	0.62	-1.81	2.43	Signaling
PC2	Ribosome	0.34	2.43	2.09	
PC2	Phenylalanine metabolism	-0.86	1.22	2.08	Metabolism
PC2	Glioma	-0.42	1.30	1.72	Signaling

PC2	Histidine metabolism	0.83	-0.75	1.58	Metabolism
PC2	Urea cycle and metabolism of amino groups	-0.02	1.48	1.51	Metabolism
PC2	Fructose and mannose metabolism	-0.29	1.14	1.43	Metabolism
PC2	Endometrial cancer	-0.09	1.23	1.33	
PC2	Hedgehog signaling pathway	0.30	-0.95	1.25	Signaling

Table A2.1. This table shows all pathways which has a difference of >=1 between sedentary healthy and trained healthy individuals (fdr<10%) using the dataset generated by Radom-Azik.

PCs	Kegg	SH mean	TH mean	diff	Groups
PC1	Cell Communication	-1.76	3.92	5.68	
PC1	ECM-receptor interaction	-1.71	3.89	5.60	ECM
PC1	Focal adhesion	-2.82	1.86	4.67	tissue remodeling
PC1	Small cell lung cancer	-1.21	2.79	4.00	
PC1	Pancreatic cancer	-1.71	1.09	2.81	Signaling
PC1	Cell adhesion molecules (CAMs)	-1.06	1.62	2.68	
PC1	TGF-beta signaling pathway	-1.02	1.57	2.59	Signaling
PC1	Axon guidance	-1.24	1.08	2.33	
PC1	Pathogenic Escherichia coli infection - EHEC	-0.83	1.37	2.19	
PC1	Pathogenic Escherichia coli infection - EPEC	-0.83	1.37	2.19	
PC1	Natural killer cell mediated cytotoxicity	-0.82	1.15	1.97	
PC1	Type I diabetes mellitus	-0.50	1.19	1.69	
PC1	PPAR signaling pathway	1.77	0.11	1.66	Signaling
PC1	Inositol phosphate metabolism	1.10	-0.56	1.65	metabolism
PC1	Benzoate degradation via CoA ligation	1.00	-0.65	1.65	
PC1	Nitrogen metabolism	0.29	-1.33	1.62	metabolism
PC1	Renal cell carcinoma	1.30	-0.31	1.60	
PC1	Prion disease	0.46	-1.06	1.52	
PC1	Endometrial cancer	-0.40	1.09	1.49	
PC1	Taste transduction	-0.86	0.61	1.47	
PC1	T cell receptor signaling pathway	-1.01	0.34	1.34	Signaling
PC1	Phosphatidylinositol signaling system	1.09	-0.23	1.32	
PC1	Oxidative phosphorylation	-0.48	0.78	1.25	
PC1	Complement and coagulation cascades	0.96	-0.28	1.24	
PC1	Pyrimidine metabolism	0.31	-0.91	1.22	metabolism
PC1	Gap junction	-0.19	0.94	1.13	tissue

					remodeling
PC1	Huntington's disease	0.79	-0.29	1.09	
PC1	Carbon fixation	-0.27	0.77	1.04	
PC1	SNARE interactions in vesicular transport	0.71	-0.30	1.01	
PC2	Focal adhesion	0.55	-4.20	4.75	tissue remodeling
PC2	Leukocyte transendothelial migration	1.30	-1.89	3.19	
PC2	Adherens junction	-1.15	1.28	2.43	Signaling
PC2	MAPK signaling pathway	-1.54	0.87	2.41	Signaling
PC2	Bladder cancer	-1.66	0.43	2.09	Signaling
PC2	mTOR signaling pathway	-1.21	0.85	2.06	Signaling
PC2	Neuroactive ligand-receptor interaction	0.81	-1.03	1.84	
PC2	Prostate cancer	0.00	-1.81	1.81	Signaling
PC2	Melanoma	0.16	-1.58	1.74	
PC2	Regulation of actin cytoskeleton	-1.13	0.56	1.69	tissue remodeling
PC2	Gap junction	-0.58	1.07	1.65	tissue remodeling
PC2	p53 signaling pathway	0.43	-1.20	1.63	Signaling
PC2	Non-small cell lung cancer	0.65	-0.92	1.57	
PC2	B cell receptor signaling pathway	0.62	-0.92	1.54	
PC2	Axon guidance	0.08	-1.36	1.44	
PC2	Long-term depression	0.23	-1.13	1.37	
PC2	Colorectal cancer	-1.00	0.31	1.32	Signaling
PC2	Glioma	0.42	-0.85	1.27	
PC2	Acute myeloid leukemia	0.66	-0.60	1.26	
PC2	Phosphatidylinositol signaling system	-0.36	0.89	1.25	Signaling
PC2	Chronic myeloid leukemia	-0.44	0.81	1.25	
PC2	PPAR signaling pathway	-0.34	0.83	1.16	Signaling
PC2	Dorso-ventral axis formation	-0.37	0.64	1.01	

Table A2.2: This table shows all pathways which has a difference of >=1 between sedentary and trained healthy individuals (fdr<10%) using the new dataset.

PC	KEGG pathway	SD mean	TD mean	diff	Group
PC1	Small cell lung cancer	1.62	-1.55	3.17	
PC1	Cytokine-cytokine receptor interaction	1.56	0.04	1.52	inflammation
PC1	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	0.56	-0.45	1.01	metabolism
PC2	MAPK signaling pathway	0.64	-1.88	2.52	inflammation
PC2	Jak-STAT signaling pathway	-0.72	1.33	2.05	inflammation
PC2	Prostate cancer	0.56	-1.32	1.89	Signaling
PC2	Cytokine-cytokine receptor interaction	0.54	-1.27	1.81	inflammation
PC2	SNARE interactions in vesicular transport	-0.55	1.18	1.73	
PC2	mTOR signaling pathway	-0.72	0.97	1.69	inflammation
PC2	Glycerophospholipid metabolism	-0.29	1.01	1.30	metabolism
PC2	Colorectal cancer	-0.35	0.94	1.30	Signaling
PC2	T cell receptor signaling pathway	-0.37	0.80	1.17	inflammation

Table A2.3. This table shows all pathways which has a difference of >=1 between sedentary COPD and trained COPD patients (fdr<10%) using the dataset generated by Radom-Azik.

PCs	Kegg	SD mean	TD mean	Groups
PC1	Aminosugars metabolism	-0.45	0.56	Metabolism
PC1	Ribosome	-0.36	0.84	
PC2	Fructose and mannose metabolism	0.34	-0.96	Metabolism

Table A2.4. This table shows all pathways which has a difference of >=1 between sedentary and trained disease individuals (fdr<10%) using our dataset.

Supplementary Material Chapter 4

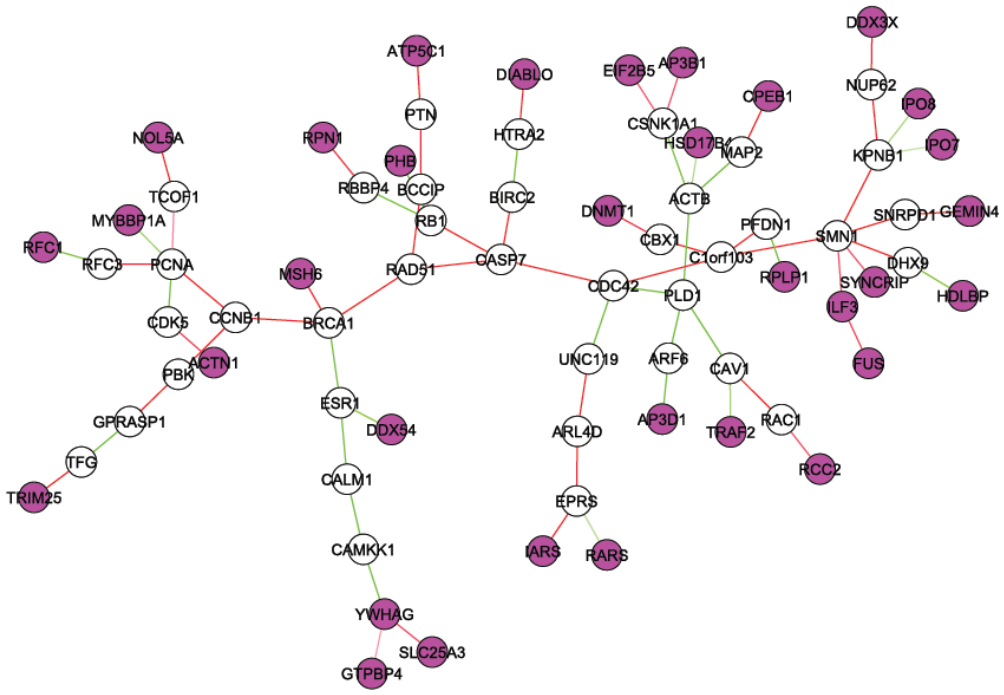


Figure A4.1. Module 2 derived from the modularization procedure.

Purple nodes represent putative targets of RhoE and edges in red represent positive co-regulation of genes in the CAM dataset, whereas green edges represent negative co-regulation.

Supplementary Material Chapter 5

Clusters	Genes
C10	HTRA1, NBL1, CTSK, FBLN2, SERPINF1, CRLF1, DDIT4, C1orf54, SEPP1, SUSD5, LGMN, PDZRN3, COL8A2, RGS3, C2orf28, DYNC1I1, CSRP2, RETSAT, FYN, CLMN, PPP1CB, COPZ2, IFI6, COX7A1, GLT25D2, WIPI1, RGL1, ABCC5, CLTCL1, CREG1, PITX2, LUM, GPNMB, GJA1, LRRN3, ZNF532, SATB2, GNG11, FDXR, HHLA3, SULT1A1, CUTC, ZNF302, TMEM135, ARL4A, ZMAT3, ARHGAP5, TMEM45A, PHYH,

	TM2D1, PTPRG, STK38L, TMEM168, PTPN13, IFI44, CCNB1IP1
C13	SCRG1, ASPN, SGCD, CA2, LRRC17, CSTA, CD200, RCAN1, USP46, KAL1, DDR2, NRIP1, TBX3, SDC2
C16	STMN2, NINJ2, IBSP, MXRA5, MGP, ALPL, MAFB, IDI1, TRIM2, HOMER2, SNCAIP, HAS1, EDNRA, MSX1, FADS2, ME1, FAM46A, POP5, PREPL, MDFI, ATP1B1, BAG2, PMEPA1
C17	FADS1, HEY1, SOCS2, WFDC1, CHRDL1, COL3A1, HMGCS1, A2M, AACS, KCNE4, RAP2A, ALDH1B1, SQLE, SPAG9, OCA2, GREM2, LIN7A, FLRT3, ATP6V1G1
C18	DCN, CCDC68, MIOS, IFNGR1, CRISPLD2, FOXO1, GPC4, FAIM, MPZL1, GPX7, C4orf18, MBNL2, CLN5, ANTXR1, CHN1, MAP3K7IP2, PON2, GBP1, RHOQ, PLSCR1, PTPLA, SLC2A10, CTNNB1, SAP30, OXR1, ZNF22, SNX2, ACSL3, SGCE, BAX
C20	HIP1, CRYAB, AMPH, PIK3R1, ACSL1, TXNIP, DLX5, FKBP5, TRIM22, IL1R1, COPS8, SNAI2, IL6ST, FOXO3, FHL1, PRKAG2, ID1, GSR, RBMS3, TWSG1, RGS2, GTF2I, KCTD12, CALD1, EPS8, GPR176, TMEM50B, ADAM10, RDX, SH3BGRL, NNT, PXMP3, THBS1, GADD45B, CD46, KPNA3, RNF146, TSPAN3, TRAM1, PCYOX1, AKIRIN1, HMGB3L1
C21	NFKBIA, FOS, ZFP36, KLF7, SCARB2, CLIC4, C19orf6, UTRN, CALU, ERLIN1, PHACTR2, LMAN1, EML4, HP1BP3, EDIL3, TMED2, VAMP3, PRRX1, YIPF5, SYPL1, UBXN4, LOX, AGPS, FERMT2, CAPRIN1, RNFT1, PTPN12, GNAS, FAS, ACTR2, C6orf62, GDE1, ENSG00000185389, SCYL2, GRAMD3, SRGN, SEC11A, OSMR, TMCO3, PRKAR1A, ENSG00000213002, RAB23, PIK3CA, NMD3, CLDND1
C25	DUSP5, PMAIP1, SPRY4, KLF6, BCAT1, EPRS, PRNP, DCBLD2, ARF6, ZNF281, HIST1H4C, SSR3
C33	WHSC1, FANCI, H2AFZ, ZWINT, NCAPH, TUBB2C, MELK, SNRPB, CSE1L, TYMS, TUBG1, UBE2C, TK1, BUB1B, SLC25A5, CCDC99, HJURP, KIF2C, KIAA0101, KIF4A, RFC5, NCAPG, TUBA1C, DDX39, ASF1B, TUBA1B, TRIP13, DBF4, AURKB, LRRC59,

	SHCBP1, TPX2, DEF8, TACC3, TTK, KDELR3, KIF15, FOXM1, KIF18A, MAD2L1, TMEM194A, H2AFX, LMNB2, CENPE, MKI67, RACGAP1, BARD1, ESPL1, ECT2, CCDC109B, RNASEH2A, HMGB2, ACOT7, CENPM, CENPI, SMC2, FGFR1OP, KIF11, EIF5A, NCAPG2, TAF5, KIF18B, KIF14, ACTL6A, RPA1, LMNB1, VANGL1, SOAT1, PSIP1, DHFR, NEIL3, ITGB3BP
C39	LYPD1, TUFT1, PSPH, RND3, DACT1, LRRC8D, ANKRD57, AMIGO2, CAMTA2, GPR177, SC65, UPP1, EML1, MRPS10, RBPJ, TULP4, C9orf91, PDGFC, INTS7, FANCL, LMO4, RABGGTB, KLF10, CEP170, COQ10B, ZCCHC8, LIMS1, AMMECR1, GALNT3, TMEM131, APOLD1, MOSPD1, CCDC59, PYROXD1, SP3, RBM16, EFHA1
C40	DUSP14, MARS, SLC1A4, SHMT2, LONP1, WARS, P4HA1, SLC3A2, GARS, YARS, TARS, EIF4EBP1, TRIB3, PHGDH, GDF15, ASNS, SLC7A5, PLAT, TFRC, CTH, SARS, G0S2, NARS, CBS, VLDLR, PCK2, FOXD1, AARS, CARS, CEBPG, HSPA9, SLC31A1, VEGFA, MOCOS, PLOD2, SLC7A11, GPR1, FHOD3, CHAC1, GOT1, MTHFD2, IER5, LEPRE1, TPBG, BCL10, SDC1, NUCB2, KDSR, CAP2, CSGALNACT1, BNIP3, LY96, HERPUD1
C45	NRG1, LAMC2, LPCAT4, DOK5, DNMBP, PTPRK, AHNAK2, ENO2, FLNC, BDNF, PLK2, RAGE, MAFF, KRT16, KRT34, DVL1, SMURF2, ARHGAP22, RGS20, INHBA, TBC1D2, SDC4, Mar-03, PLXNA1, GPRC5A, RABGEF1, TRIB1, DOK1, TEAD4, HBEGF, TACC2, ABL1, DGKI, NMNAT2, MGC4294, SH2B3, FHOD1, PTHLH, SH2D4A, FGF2, LGR4, ELOVL6, CDYL, KRT33B, TNFSF9, FADD, LARP6, CD55, CDCP1, SPTAN1, LDB2, SEC16A, TCIRG1, HGS, KIAA1644, DGCR8, LMNA, GALE, SLC14A1, DCI, ZNF365, ATF3, ZFPM2, KIAA0802, C10orf72, IL4R, CDC42BPB, MAD1L1, CLIP2, CXCL6, ESM1, PER3, AFAP1, CARM1, SF3A2
C46	PLP2, RGS10, RRAS2, PHLDA1, SERPINE2, ASAP2, ADAM12, GSTO1, KRT7, MICAL2, AXL, SLC20A1, EFHD2, SERPINE1, NT5E, SLC16A2, PHLDA2, AEBP1, WNT5A, PKM2, NTM, BLM, DRAP1, UGDH, MSN, SPP1, ATP6V0E2, FAM62A, DEK, MFAP2, NMT2, ARSJ, NMB, ERCC6, ERCC6L, CHIC2, MGAM, KIF22, HMGA1, MEX3D

C47	PRKCA, SEMA5A, C21orf7, CALB2, KCNN4, COL5A3, OPCML, PLAUR, GABBR2, MALL, TMEM158, ALDH1A3, TM4SF1, MEST, RGS4, KRT19, FJX1, ITGA3, TNFAIP3, MLPH, HMGA2, CTNNAL1, PORCN, AK5, S100A2, ITPR3, LRP8, LPXN, CA12, CCND1, FGF5, CAMK1, HPCAL1, NEK7, CDT1, ARHGAP29, DPYSL3, PRSS3, ANXA10, BCAR3, PKIA, EIF6, CDA, SERPINB2, EXO1, FGF1, RAC2, C11orf41, KRT81, PLCXD1, TGM2, DCTPP1, GPR126, SIK1, ZYX, INHBB
C5	SNED1, PC, RFTN1, C3, RABGAP1, POM121L9P, PISD, HSPB7, ABCC3, PGM1, STAT6, GSTT1, PTGIS, ST5, B4GALT5, RARRES3, HDAC5, FCGRT, C1RL, ADARB1, CCDC69, AKAP12, UBE2L6, OLFML2A, TENC1, PPL, CLEC3B, LXN
C50	S100A10, HEG1, TIMP3, CD44, CHI3L1, SYNJ2, LOXL2, TRIOBP, FXYD5, UGCG, FZD7, TJP2
C51	MMP2, CALHM2, QSOX1, SPOCK1, FHL2, STAT1, CDH13, BST1, DSP, NID2, MAP4K4, C5orf30, PRKCDBP, IGF2BP2, CBR3, ARNT2, ZNF217
C52	MMP2, S100A10, HEG1, TIMP3, IGFBP3, CD44, CHI3L1, CALHM2, QSOX1, SYNJ2, SPOCK1, FHL2, LOXL2, STAT1, TRIOBP, CDH13, FXYD5, BST1, DSP, UGCG, NID2, MAP4K4, FZD7, C5orf30, MYO1B, CD9, PRKCDBP, IGF2BP2, CBR3, C6orf145, TJP2, ARNT2, TGIF1, CTGF, FAP, ZNF217
C6	MXI1, TGFB2, MYO6, BCL6, METTL7A, SAMHD1, AOX1, IMPA2, C1orf63, CEBPD, ABCA8, RCAN2, MAP3K5, RNASE4, SORBS2, FBXO42, CFLAR, SSB, ZFP36L2, CCNG2, CEP70, SESN1, ALDH3A2, PNRC1, OPN3, NEO1, RNF141, GCNT1, TLR4, NCRNA00084, SLC26A6, AKR1C3, ZCCHC11, FBLN5, PIGG, IL10RB, ALOX15B, ABCA6, C17orf48, HSD17B11, BBX, HPS5, POLI, REV3L, SLC38A2, NUDT4, DSE, MAT2B, TWIST1, CLK4, CLASP1, HMG20B, ARMCX1, FNBP1L, ABCC4, C1orf25, SAV1, MAP4K3, PLSCR4, KRCC1, AGL
C7	AHCY, SH2B1, TCEAL1, CDKN1C, SORT1, LOC727942, SSBP2, ITGBL1, ITGA10, ALDH6A1, P2RY5, LEPROT, TCEAL4, FZD5, IFITM1, PRKD1, ID3, DAAM2, TSC22D3, WASF3, HSPA2, STON1, TNS3, GAS1, LEPR, CPM, SAT1, CAT, PPP1R3C, BDH2, HSPB2,

	LMOD1, SVEP1, STOM, NEDD9, CLIC3, NNMT, PALLD, ID2, PTK2B, NFIL3, FZD4, CYBRD1, PRRX2, ENAH, SMAD6, FAM117A, PDCD4, RRAS, RCBTB2, IRX5, CIDEC, HIST1H1C, ADH5, PYCARD, APOL2, CNN2, TCTA, PLA2G16, PIR, ECH1, GMPR2, ANGPT2, SIX1, CCDC53, MMD, ZBTB1, LMBRD1, PPAP2B, STAT2, ZFHX4, ARL6IP5, PTPLB, USP13, PGD, PAPSS1, GNAI1, DPYD
C8	TP53I3, NGFRAP1, STEAP3, FBLN1, C1S, TRIB2, SOX4, C5orf23, ID4, ANGPTL2, CFH, OLFML3, MID1, HIST1H2AC, EPHX1, GPD1L, GSN, SELENBP1, GGT5, PLAC8, SERPING1, GCHFR, CD24, IGFBP2, ANGPT1, MMP7, FRZB, IGF2, CFD, MTCH1, GAS6, PLXND1, GPX3, IFITM2, TPD52L1, ENC1, HCFC1R1, GABRE, CTSO, NELF, CYP1B1, SMPD1, SLC5A3, APOL1, GPX1, PTGDS, LDLR, PDE1A, TRAPPC6A, SPATA20, HIST1H2BK, ASAP3, BACE1, BBS1, GLS, TBC1D16, HSD11B1, SULF1, THNSL2, C5, VAMP5, SLC22A18, APOB, ARSE, SMARCD3, CTSF, SNAI1, C7orf10, SETMAR, SETBP1, RAMP1, ELANE, MYC
C9	ABAT, SFRP1, IL13RA1, CALCOCO2, LPAR1, CYB5A, GSTA4, LASS6, STAT3, SLC27A3, SSPN, TPST1, PPAP2A, NPR3, GHR, LSP1, CKB, HNMT, MCAM, C4orf31, MAOA, TNC, ACAN, LMCD1, FZD1, SRPX, RAB31, BAMBI, MME, PROS1, PRELP, ECM2, TIMP4, ZBTB16, ACTC1, CD302, COL11A1, CORIN, COMP, OLFML2B, OMD, HMOX2, EYA2, ADAMTS2, EFHD1, FBXL2, FAM13A, FAM69A, EPAS1, IL11RA, APOD, GLT8D2, CST3, FMOD, PGF, LIMCH1, C11orf67, SLC38A4, DCXR, KAZALD1, ACTA2, RANBP17, PPARG, FTO, ITM2B, TP53TG1, DECR1, IDH2, LAMA2, LAMB1, PPFIBP2, IDH1, ACOT13, COQ2, C1QTNF1, ISCU, AIM1, PAPPA, NXN, IFI44L, GMFG, MXD4, INSIG2, PDGFD, FAM3C, CNN1, USP33, MARCKS, ISG15, SUCLG2

Table A5.1. Genes in each component of the osteogenesis model.

clust #	Genes

C11	HADHB, PPP2R1B, HIST1H2AC, IDH1, ACSF2, PALMD, PLA2G16, NR1H3, SCD, GHR, RBP4, ADIPOQ, LPL, GPD1, PEX19, ALDH9A1, ALDH1L1, PPIF, ACO1, FHL1, LPIN1, MOSC1, SEMA3G, AOC2, CALCRL, LIPA, PNPLA2, ITGA7, PECR, MCCC2, NIT2, HIST1H2BD, TALDO1, SLC27A3, C20orf3, ECHDC3, UQCRC2, SPRY1, DUSP6, PIGP, HSDL2, TMEM97, C2CD2, NDUFB5, HADHA, C1orf38, MOCS1, SLC25A20, FASN, COX7C, CCNH, LRRC8B, NCOA4, MEGF9, INSIG1, ELOVL5, ASH2L, BTD, CHP, C1orf50, SAMM50, LPCAT3, NPTXR, GYS1, ADRA2A, TMED5, MYL5, LBP, APOD, CFH, APOC1, SNAPC5, SORT1, MRPS14, BTG2, ACOT13, PEBP1, LIAS, ENPEP, NECAB3, SDHD, MRPS33, ALDOC, SUOX, C21orf33, ARL6IP5, C20orf27, FMO2, C14orf139, CYB5P4, FXYD1, TMEM53, ALPL, TMEM22, EFNA1, MAOB, ETFB, FAM162A, PLEKHG6, SDCBP, KIAA0664, RPLP2, C11orf71, THTPA, RAC3, LRPPRC, ACADS, CRYL1, PEMT, PIK3CA, DDT, A2M, APOB, MOAP1, FAH, OPLAH, COX10, USP25, PEG10, NDUFAF4, UROS, HSD17B8, ECSIT, SCP2, HIST1H2BN
C12	CAT, FAM134B, SERPINF1, MAOA, CYB5A, METTL7A, MME, CHRDL1, ACSL1, CIDEC, CD302, ADFP, G0S2, DHCR24, FZD4, GPC4, MMD, MCCC1, ALDH3A2, TMEM140, STAT5A, DCXR, SDR39U1, CDC42EP4, KCNK3, GPR64, ECH1, CDO1, PCCA, CABC1, SELENBP1, FAM13A, TSPAN31, CCBL2, CCNG1, ASAH1, PIGZ, MRPL34, PHYH, GALT, TLN2, SURF1, STAR, SLC5A6, GCLC, RBPMS, PBLD, LYRM1, STK39, HIST1H2BK, GADD45A, SH3PXD2A, C14orf147, GPHN, ZNF91, TOB1, TMEM100, P2RY5, PIGH
C13	CLMN, PPARG, ATP1A2, CEBPA, ACACB, HP, CFD, ADH1B, PRKAR2B, TIMP4, PLIN, AOC3, CHI3L2, FABP4, APOE, HK2, RETSAT, TMEM135, ALDH4A1, PSAP, GBE1, ACSL5, RAP2A, ABHD5, OSBPL1A, PDCD4, CSAD, SEPP1, HNMT, MAFB, AKAP1, CITED1, PLTP, CEPT1, HPR, TMCC1, EPAS1, DCLK1, ANXA6, NRN1, AQP3, TSPAN14, PAPSS1, GPX4, PCOLCE2, CXCR7, GBP2, TMBIM1, FMO3, TSPAN6, QRSL1, RNASE4, PTPLB, GBP1, INHBB, WDR3, ACOX2, LIPT1, FCGRT, FMO4, COG2, ANG, TXNL4B, LAMA4, STXBP1, TRIB1, WFS1, CILP, LDLR, FERMT2, SYNPO, LMO4, ASPH, MUT, ECM2, EPS8, HSPB6, LBR, C8orf33, APBB1IP, RMND1, GNAI1, ZFAND1
C16	PLXND1, FBLN1, ANGPT2, LAMA2, ABCC3, SIDT2, PCSK5, GUSB, HEBP1, SCPEP1, IDUA, MARCKSL1, POM121L9P, ADA, EYA1, CTSF, C1GALT1C1, IL17RC, MEGF8, CTSC, LGALS3BP, DBC1, PLD3, TNFRSF1A, ZMAT3, CTSO, TRIM22, EIF4B, IL11RA, OSR2, CAMLG, MKL2, TPP1, PRKD1, HSPB2
C19	BCL6, CREG1, AKR1C3, SOBP, STOM, GPD1L, AKR1C2, IGFBP2, GPX3, EPHX1, NID1, ZBTB16, CCDC69, COL11A1, HSD11B1, CPM, MGP, SRPX, TBC1D16, SERPING1, GGT5, ANGPT1, ALDH6A1, RXRA, LEPR, GSTA4, PC, GLTSCR2, PPAP2A, SAT1, C2orf28, LETMD1, PIR, PISD, DAAM2, WASF3, TMEM111, GPX1, FTH1, BTG1, AACS, C1QTNF1, BBS1, TGFBR3, FOXN3, IGBP1, LMO3, TSKU, PTX3, TRAPPC6A, LEPROT, AGTR1, IFT122, TNFRSF21, ABHD14A, WDR19, PGD, SIX1, ABCA8, RNF130, ABCB6, EIF4A2, CRBN, PAPPA, LMOD1, LASS6, PEX6, G6PD, COX15, NUDT4, FOXC1, USP33, TP53TG1, LHFP, RPL22, RSL1D1, CEP68, FDXR, GSN, DHRS3, ABHD4, CRTAP, FNTA, IFI6, PUS3, MAN2B1, ROCK2, MYCBP2, CCNB1IP1, OXA1L, NELF, MYO6, RHOBTB3, FAM59A, GCHFR, COX11, STEAP1, OPN3, IGF1, LMCD1, FKBP5, ASB13, STIM1, TXNIP, THNSL2, CFLAR, SETMAR, EIF3E, CSTA, ADM, C17orf48, TM2D1, NCRNA00084, CRYAB, PABPC4, HSPA2, SNX2, TMEM168, HHLA3, PCYOX1, ENSG00000185389, NNT, AZI2, GNS
C22	TCEAL4, COPS8, SSBP2, PTPRM, REV3L, MXI1, C5orf23, IFNGR1, FRZB, MPPE1, RASSF4, HOXA10, TCEAL1, SMARCD2, NPR3, MYC, PTPRG, FAM110B, LAMB1, AKAP12, SLC6A8, TENC1, DEPDC6, SLC35A1, LOC727942, IL10RB, GSR, GLT25D2, GCLM
C25	PHF20, PHKA1, PITPNC1, SOCS2, FOS, RORA, SLC25A12, PTGER2, NXF1, UGP2, BMI1, RBMS3
C26	MT1M, CHST7, MT1E, MT1X, MT1F, IMPA2, PHC2, PANX1, C10orf26, MT4, MIOS, IL15, TBL1X, TMX4, FBN2, CRISPLD2, DLG5, MT1P2, HIP1, TPST1, ZNF395, PPP3CA, BTBD3, FYN, CMTM6, KCTD12, SRGN, WISP1, ZFHX3, CHN1, MGA, DZIP1, CCDC53, MAPKAPK2, SSFA2, MORF4L2, SLC25A37, WIPI1, NAV2, KLF7, UGCGL2, NINJ2, CTNNB1, IFI44L, PDE7B, DDIT4, MAN2A1, ENPP2, UGCGL1, IGF2BP2, ZFP36L2, PDZRN3, CLK2, HDAC4, SPATS2, C4orf18
C27	LTBP1, TBC1D1, HAS1, SH3BP4, ATP1A1, ITGA5, FARP1, GFPT2, GNG11, ABL1, ID2, ID1, ODZ3, GGH, ID3, DNAJC15, NUDT1, BCL3, FOXF2, POLD4, SLC39A14, DNAJC1, EMP2, PLSCR1, GPR1, PNPLA6, SLC38A2, ACYP1, TMEM204, FBLN5, CTSZ, C20orf149, ZMIZ1, CALR, BRD3, STEAP3, IFI44, ZC3H4, TRIP10, BICC1, UPF3A, IL13RA1, PHF21A, WDR37, CCDC28A, PGF, DIRAS3, NUP37, TWSG1, C7orf42

C3	DHRS11, HOXB6, QDPR, SCCPDH, SV2A, SERPINI1, AVPI1, SCML1, C11orf48, C14orf109, BNIP3, GNE
C30	SDC4, RGS10, CA12, COL5A1, FXYD5, LXN, PLOD3, PENK, TIMP2, RCN3, LOXL1, MVP, QSOX1, FKBP11, TMED3, DNASE1L1, AEBP1, TCIRG1, CHI3L1, PTGES, LEPREL2, ECM1, FAP, GOLM1, GTF2IRD1, ARMCX2, COL1A2, LDOC1, S100A6, FZD2, FLRT2, CAPZB, SLC10A3, TUSC3, PPIB, THY1, LOX, TAX1BP3, PDLIM2, NPDC1, TBC1D8B, CDK5RAP2, SNX7, ZYX, CDC42EP1, PRAF2, COBRA1, RBM9, FAM176B, SPARC, CDR2, IRF9, CYTSA, AMPD2, CPNE1, MAP1B, COL5A2, DCTN1, PLEKHO1, RNF34, COL16A1, FAM128A, SMAD7, SCRN1, ABR, ZDHHC7, RABAC1, EPHA2, TPM2, TXN, SLC7A11, LAMB2, ERBB2IP, BCL10, RAB11FIP5, C6orf145, TXLNA, HNRNPUL1, ZDHHC6, SETD5, CDH13, CLIP2, FSTL3, H1FX, SF3A3, PRR11, IER5, FBXL7, C14orf132, TRIM28, SGCE, INPP1, EVL, MOXD1, KDM3A, TMEM50A, CARM1, C12orf24, MICALL2, DFNA5, HS3ST3A1, MEPCE, CAD, KIAA0947
C32	AHNAK, CLSTN1, DKK1, DDOST, GANAB, ASMTL, GLG1, BACE2, RPS6KA2, MAP2K2, HECA, MIF, PLOD1, VKORC1, P4HTM, FSTL1, S100PBP, GLT8D2, UBP1, HMGN4, ZFC3H1, ARID4B
C37	CCNA2, HAPLN1, NT5E, CRIP2, DPYSL3, TPX2, PBK, TUBA1C, CDKN3, ASPM, PFKP, FOXM1, CYTL1, TFPI2, KIF20A, PRC1, KIF2C, MLF1IP, CENPF, PHLDA1, PLK1, TOP2A, NCAPD2, KIF4A, NUSAP1, CEP55, CCNB2, UBE2C, BIRC5, TBC1D19, KIAA0101, DLGAP5, GOT2, TACC3, SGK1, C7orf10, PSRC1, CCNB1, ACOT7, HRAS, GLIPR1, GPSM2, STMN1, LGALS1, RACGAP1, KIF15, COLEC12, ODC1, TK1, SCARA3, RNASEH2A, DOCK5, PLCE1, RPL39L, DDAH1, SQRDL, KLHDC3, CENPM, RPA3, PFN1, MAD2L1, PAFAH1B3, HYOU1, CALM3, KIF14, HMMR, CIT, S100A4, ARNTL2, NANS, ARF3, ANXA2, FAM158A, ANXA5, RAB27A, CKS2, CALM1, FGFR1OP, IRAK1, RPN1, STK24, LRRC15, PGK1, WEE1, ITGA3, EPB41L2, C18orf10, MYO5A, GOLT1B, OIP5, IMPA1, FAM64A, TROAP, HMGB3, GNAI2, CREB3, NPTN, DEPDC1, ADORA2B, CYR61, APPL2, UXS1, BUD31, P4HA1, TMEM48, HMGB1, PAMR1, IMPAD1, PHF10, DGKZ, SAP30L, SEMA3C, ARL4A, VASP, BEX1, LDHA, PAF1, DERL2, GUK1, CDC37, ACAN, C3orf14, CDK2AP2, AP1G1, GMNN, CDA, CLIC4
C38	WHSC1, RRM2, C5orf13, CDC2, CBX5, FANCI, CENPE, AURKB, TUBA1B, TPM1, SLC38A1, NDC80, ZWINT, H2AFX, PRKCDBP, NCAPG2, CRABP2, BDNF, ZWILCH, CENPA, RAB3B, MKI67, LEPREL1, LRRC17, WARS, TRIP13, SPC25, PGM3, TUBB3, MELK, TTK, STIL, EZR, NTM, ARMET, CDCA8, PCK2, ANP32E, NCAPG, RANGAP1, AURKA, BUB1B, SAE1, UCK2, GTSE1, KIF11, ASNS, SNRPB, CENPN, ASF1B, APOBEC3B, CTNNA1, ESPL1, CALM2, NAGK, C21orf45, KIF18A, SPAG5, RFC4, YKT6, GAS2L1, CDCA3, SHCBP1, ORC6L, LIG1, DBF4, TMEM184B, REEP4, LMNB2, HSPA13, DDX39, GNB1, H2AFZ, ECT2, PLOD2, KIF22, LIMCH1, SLC1A1, NPC1, NCAPH, CSE1L, ACTN4, SDF2L1, UBL3, DR1, DCBLD2, CTGF, TBC1D4, VRK2, PSMG1, VRK1, SMC2, EIF4EBP1, C18orf24, ARPC1A, PSMC3, CCDC109B, ELAVL1, NUP155, PLEKHB2, SIVA1, KIF18B, HNRNPA2B1, ATAD2, GADD45GIP1, HAT1, C13orf34, FANCG, BLM, RANBP9, KIAA0355, SSR1, CDC7, YAP1, EIF5A, SPP1, DYNLT3, MAPK1, MAP3K6, HNRNPH3, C13orf27, FANCL, GMFB, TAF5, RAD54B, FAM129A
C39	SLC7A5, SLC3A2, GPRC5A, KLF4, C12orf44, KIAA1279, ATP1B1, FUS, PTP4A1, SPHK1, PLEKHA1, MMP1, C1orf9, SMAD1, SGMS1, HSPA4, CRY1, PRPF4, CSTF3, POLE3
C40	GABPB1, CDR2L, DLX2, JARID2, NT5DC3, HSPB8, BAG3, EIF2AK3, ANKRD57, TBPL1, DNAJB9, B3GNT2, CTH, SIK1
C41	SH3BGRL3, AXL, EMP3, SIPA1L1, TYMS, FSCN1, ELTD1, TMED9, MAGED2, KRT7, PPP1CA, CDH2, MAGED1, PRRX2, CDK2AP1, MYH9, TUBA1A, GALNT10, PTBP1, EHD3, SEC61A1, TPD52L2, SEPW1, CAPG, CAPRIN2, ARPC5, S100A10, SSRP1, SEC61G, CAPNS1, SUMO3, HOMER3, ASAP2, MSN, RAI14, TGIF1, TBCB, DBN1, COMMD4, M6PRBP1, HMGA1, GRB10, ARMCX6, HSBP1, THBS2, ANXA11, C1GALT1, TEAD4, EIF4E2, GNG12, PRSS23, PPP4C, MLPH, CALU, MCFD2, LRRFIP1, UBA1, COPZ2, S100A11, GMPS, LRRC8D, DYRK4, NOTCH2, RBM10, VAT1, KIAA1199, RBM3, PLEC1, CARS, CBX1, CANT1, ACOT9, KIAA0746, MYOF, GALNT1, ANKLE2, ARFGAP1, RIT1, PHGDH, STT3A, HDAC1, MFSD10, TLN1, CALD1
C42	TRIOBP, SEC13, CAP1, EFHD2, PSMD2, RRBP1, ARPC2, RHOC, PRR16, EHBP1L1, RNH1, DRAP1, LMNA, SERPINH1, HK1, C11orf24, SFXN3, C16orf45, FAT1, CTBP2, PDLIM7, KLC1, PLXNA1, LSM7, NMI, ATP13A2, C21orf59, TUBA4A, SEC61B, SRM, ARHGDIA, DVL1, SRP54, TRPV2, DEF8, FAM38A, LIMS2, PRPF3
C43	KDELR3, FTSJ1, GARS, PLAT, CREB3L1, GREM1, SERPINE1, RAC2, LYPD1, SHMT2, MCM5, HPCAL1, PDLIM4, EIF1, KIAA1644, PORCN, SMAGP, PHLDA2, RRM1, NDEL1, RAB32, ITPR3, SDC1, TMEM158, FLNC, DAP,

	CLIC3, S100A2, BCAT1, SLC2A3, IL4R, XYLT2, HERPUD1, ZNF331
C44	LRRC59, YWHAH, KDELR2, SLIT2, MCM7, FAM114A1, UGDH, CCND1, EIF6, PSIP1, TULP3, SLC20A1, IGFBP3, XBP1, MYBL1, CSRP2, VEGFC, MCM6, SMC4, DSP, PTPRK, RHOA, HNRNPAB, ARF4, MORC4, MAD1L1, CHSY1, RAC1, EXT1, KCNK2, BARD1, DNMT1, EZH2, TMEM214, MCM3, RBBP8, U2AF1, CREB3L2, FAM111A, TRIB2, HIST1H4C, ACTR3, UPP1, SERTAD2, ARCN1, CBR3, BUB3, STX4, ZDHHC13, DEK, FAM57A, DONSON, SMURF2, IVNS1ABP, SEC31A, GMPPA, IQGAP1, CHIC2, PMM2, TSPAN5, KIAA1462, COPG, TULP4, CTPS, GOLGA5, SLC39A7, GLS, FZD6, PPP4R1, CBFB, NR3C1, YWHAQ, HSPA5, ARFGAP3, TES, KCTD5, TFDP1, H3F3B, ADAM12, LRRC42, PWP1, STAG1, CTCF, ARL4C, FZD7, OAT, RPS6KC1, PTGES3, TOP2B, GOLGA3, RBM15, CRIM1, PLS3, SACS, SWAP70, PCNA, CKLF, LASP1, ADSS, MTMR6, COPB2, ALDH18A1, NRIP3, ARF1, MFSD1, UTP18, SLBP, ILF2, CEP170, FKBP14, FNDC3B, NPM1, PPP1R12A, CSGALNACT2, NUP85, DDX46, SEC16A, HOMER1, SNRPA1, FIP1L1, RPS6KA3, TOPBP1, MAP4K5, ACTR2, WAPAL, NACC2, DOCK4, MEX3C, OSBPL8, NSUN5, RNF138
C5	MOSC2, PDE3B, ITIH1, SPARCL1, PDE8B, SPON1, PLA2G2A, ME1, TF, MKNK2, OSBPL11, ITIH5, ISOC1, AZGP1, ECHDC1, ATP5S, TSPAN3, NRCAM, PFKFB1, ACYP2, PEX11A, ELOVL6, SYN2, RDH5, C6, TRD@, SLC7A6, PDZD2, PCBD1, SCARB1, SNX10, FAR2, SLC19A3, OAS1, PHLDA3, KCNB1, SLC16A7, KIAA0040, ACSS3, MVD, HPGD, RNF125, PNPLA3, ARG2, ADRBK2, ACSL4, C13orf15, EIF4EBP2, CASP1, MRAS, CDR1, PDK4, ABCG1, ANKRD46, STX12, PFKFB3, ABCD2, AP3B2, ATF3, LACTB2, AGT, STMN2, ACOX1, SCAMP5, MID1IP1, ABCA1, SC5DL, LAIR1, RAMP2, CASP7, FNDC4, GAB2, SLC25A10, S100B, GBAS, CIDEA, ADRB2, PDK2, BANK1, CA5B, RRN3, WFDC1, MPDZ, ARHGEF2, UQCC, CHPT1, SLC47A1, AMBRA1, SOX9, CASQ2, CADM3, MECR, STARD8
C51	PRPS1, SMTN, KRT19, MALL, NP, FEN1, MICAL2, CCDC99, DTL, CAP2, SYNJ2, YIF1A, ADCY7, ETV5, PDGFC, GINS2, LPXN, BAALC, MCM2, AMIGO2, USP1, PLAUR, EXOSC9, GAP43, CDCA4, NETO2, KIF23, OSBPL3, PRKCA, RASA1, RAGE, NUP205, NEK7, CDC45L, CTNNAL1, POLA1, SEC24D, AHNAK2, LDB2, ATXN1, HNRNPD, MACF1, LANCL2, YARS, HJURP, SLC17A9, POLE2, RFC3, CD97, MET, FHOD3, MCM10, NMNAT2, DOCK10, BTG3, GINS1, SH2B3, MYO1B, ENO2, MAFF, EML1, TARS, RAD51AP1, KCNN4, DCK, MARCKS, EXO1, SHB, COG7, MRPL17, KRT14, SFRS2, CDC6, SH2D4A, MAPKAPK3, DOK1, PRSS3, PELO, CHMP6, SEC24A, PRRC1, UBR7, YIPF5, MGC4294, GAR1, HELLS, SLC4A2, ACBD3, PRIM1, MSH2, ANXA1, PYROXD1, FBXO5
C53	NUP93, WSB2, GALE, ABHD2, TACC2, PSMD14, PTS, C16orf61, ARL1, KIF5B, EXOSC8, POLD3, PARP2, SFRS3, SFRS1, SEC23A, SNRPD1, UCHL3, NUP107, DYNLL1, SAC3D1, NIF3L1, AKTIP, PALLD, GLOD4, EEF1E1
C54	TUBB2C, MARS, CIB1, ARPC5L, SLC1A4, LONP1, TMEM106C, CKAP2, TUBG1, BANF1, EHD4, SLC31A1, EBNA1BP2, ODZ4, EPRS, TRA2B, GRAMD3, H2AFV, NRAS, MLH1, CCDC90A, UBE2M, C11orf51, CSRP1, TDP1, TAGLN, RPA1, VANGL1, OLFML2A, HPRT1, SPRY2, THAP10
C6	PDHA1, BRP44L, DBI, NSDHL, DUSP4, GYG2, UCP2, C10orf116, DLAT, HADH, CD36, PCK1, CDKN2C, C7orf68, HSD17B12, STBD1, ACO2, AGPAT2, IDI1, HIST2H2BE, TM7SF2, MGST3, ACAT2, ACY1, ADIPOR2, PPP2R5A, ATP5G3, KCNIP2, SLC25A1, PCCB, PDHB, SLC7A10, CAMK1, DHCR7, CS, PPP1R1A, MDH1, AK3L2, LIPE, HMGCR, BRP44, PYGL, PTGER4, LSS, COX5B, RTN3, SUCLG1, AIFM1, NQO2, ALAS1, HMGCS1, CCDC56, HSD17B10, BCKDHB, LETM1, COL4A1, HEBP2, UCRC, GRPEL1, UQCRFS1, NFU1, UVRAG, GYPC, ACACA, ATP5B, COQ9, IMMT, CYP51A1, GGCT, ECHS1, FAM82A2, ETFDH, ACADVL, WDR77, ALDH1B1, COX7B, FDFT1, SMCR7L, NMB, BCAP31, FDPS, SREBF1, PGRMC1, VIPR1, MRPL46, ACAD8, ZNF117, NDUFA3, TIMM8B, GLRX5, NDUFC2, NPR1, BBOX1, HISPPD2A, PLCXD1, SPTBN1, BAI2, ATP5J, C17orf42, COL4A2, MRPS16, HIBCH, C3, SLC25A44, FMO1, RHOB, WBP2, CD46, ACTG2
C8	KANK1, MEST, LRIG1, VCAM1, CAV2, ATP9A, ACTA2, APEX2, SC4MOL, PLA2G4A, PJA1, ICT1, TJP2, FYCO1, ANP32A, SDC2, OPTN, ANAPC13, PNO1, BLCAP, ARHGAP29, MOSPD2, TSC22D1, DET1, RANBP6, SEH1L

Table A5.2. Genes in each component of the adipogenesis model.

#	Genes
C1	RASL11B, CD24, LRRC15, COL11A1, SCRG1, OMD, ASPN, MFAP4, ATP1B1, PPAP2A, APOD, MMP7, PLEKHA5, IGF1, LRRC17, LUM, SDC1, CHN1, TCF4, DNAJC12, GJA1, STEAP1, SSPN, TSPAN13, SLC5A3, PMS1, NDUFB5, RIN2, PIGH, BMI1, TRMT61B, SMAD1, ZMAT3
C11	RAB31, MAFB, FOXO1, JHDM1D, IFNGR1, MT1F, DDX3Y, RSL24D1, BDH2, SCD5, BCAT1
C12	MIF, STEAP3, SERPINA3, VKORC1, SERPING1, P4HA1, CA9, TNFRSF21, COL7A1, TRIB3, PSAP, CCNB1IP1, ACSL1, SNED1, OS9, GLTSCR2, C11orf10, SPAG4, TWIST1, CCNI, HIST1H2BK, GBE1, SLC3A2, DNAJC1, EEF1A2, EIF4B, NARF, AARS, PLOD1, OBFC1, TYRO3, OPN3, NNMT, ZNF274, PTGFR, MT1E, POLR1D, DYNLT3, TAF1D, PLTP, ZCCHC6, RASSF1, LRP1, SYNPO, ITGB3BP, H1F0, EIF1B, IFRD1, LMBRD1, OXSM, ANKMY2, FAH, SH3YL1, PLD3, TMEM135, CLK4
C13	ALDH6A1, TXNIP, SRPX, ITGA10, COPS8, FAM46A, PTER, STON1, PIK3R1, AACS, CHST1, JUN, SATB1, EIF4EBP1, TGDS, SLC46A3, BEND5, CDO1, DCP2, PHF17, ISLR, ARL4A, CUTC, NRBF2, CNIH4
C14	SAT1, PGK1, UGP2, MT1X, BNIP3, PMAIP1, ABCA1, MT1M, FOXN3, OXR1, GNL3, HSPA13
C19	NFIL3, MXI1, TSC22D3, RGS2, IL1R1, GADD45B, C7orf68, TMEM204, PNRC1, BCL6, G0S2, NOL3, SESN1, PMP22, PPP1R3C, GAS1, SORBS2, SLC25A36, COL6A3, FBXO42, CCDC28A, ANKZF1, PER1, POLR2H, ARL4C, ZNF292, INSIG2, FKBP5, HERPUD1, C10orf26, LARP6, ZNF331, PDLIM4, PGF, TBC1D1, SMARCA2, TNS3, NID2, CD9, AGPAT5, PRKAB2, AK3L2, HIP1, RBM7, TMEM87A, TBC1D16, PIK3CA
C2	COL8A2, FMOD, ANGPTL2, TPST1, S100P, SOX9, TRPS1, SLC29A1, PID1, DYNC1I1, GPX7, SPARC, SOBP, VGLL4, IVNS1ABP, PAPSS2, COL1A1, EPS8L2, PDGFRL, GPC4, COL16A1, ITM2B, CCPG1, EDNRA, STXBP6, A2M, PLA2G4A, ARHGEF6, PLOD2, GPR153, LSP1, FAM102A, STK38L, IFT20, MTHFD2, PMEPA1, BIRC2, RP6-213H19.1, GOLM1, STIM1, CTNNB1

C21	FOXO3, GFPT2, ANGPTL4, VEGFA, F3, KLF9, SLC39A14, AKAP13, SOCS2, DIXDC1, FOXF2, MARCKSL1, OSBPL9, MAP3K2
C24	COL4A2, IGFBP3, CD44, COL4A1, TMSB10, ARID5B, MAFF, TFPI2, CTSL1, LGALS3BP, BAALC, KIAA0802, PTPRG, SLC1A1, IFITM2, CAPG, WASF3, MAP4K4, AKR1B1, IER3, ARHGAP22, TAGLN2, CDK5RAP2, ST13, OSMR, PCYOX1
C25	STOM, LTBP2, IGFBP6, TGFBR3, CHI3L1, C5orf23, S100A6, CRISPLD2, FLNB, CTSA, DHRS3, MLPH, JMJD6, NPR3, PFKP, RRAGC, CALHM2, PTX3, MGAT4B, NPC2, S100A10, SCML1, CSTB, QSOX1, PECI, TRIP10, CDH13, GNS, TKT, ATP2B1, AHNAK2, YAP1, TSPAN5, F2R, SOAT1
C3	BGN, COL3A1, SERPINF1, DPT, COL10A1, COMP, HTRA1, MFAP2, ECH1, IGFBP5, CTSK, PTGDS, LPAR1, NDUFA4L2, SSR4, RNF144A, SRPX2, TMEM208, MMP13, PCOLCE2, PPP1CB, IMPA2, IGF2, BHLHE41, CSTA, SPON2, KLF11, DYRK4, DNM1, BET1, GOLGA9P, IRX5
C31	MYOF, GNG12, RAB3B, TPM1, FHL2, PPAP2B, NQO1, MEST, EFEMP1, CAP1, AXL, PENK, COTL1, GLIPR1, ACTN1, TRAM2, KIAA1199, PHLDA2, TRIOBP, NRP1, GAS6, PTRF, BST1, TAGLN, ANXA2, DAB2, MFAP5, ERLIN1, EIF6, VCL, UBE2C, ANXA11, ATP2B4, KRT7, LASP1, SSH1, SLIT2, CORO1C, SYNJ2, IPO5, ALCAM, G6PD, KRT19, MTHFD1, MYL9, S100A11, CAPNS1, CFL1, STC2, FHL1, SDC4, CALM3, LAMC1, ATP13A3, PLEKHB2, M6PRBP1, UGDH, PFN1, APPL2, MET, CUTA, ENG, ENTPD6, NTM, ATP6V1B2, TLN1, POLR2E, CKS2, DNAJB6, RACGAP1, ATP6V1H, ARNTL2, LYPD1, DRAP1, CTNNAL1, C2CD2, CLTA, ASXL2, SPAG5, TRIP6, PKM2, STMN1, INPP5A, FLNA, ASAP1, ERLIN2, PDLIM2, CYP51A1
C32	PEA15, CITED2, ADAMTS1, APLP2, UGCG, IFI30, AP2S1, IGFBP4, MVP, RPS6KA2, PLA2G16, RRAS2, TXNRD1, FZD2, FAM62A, ZDHHC7, DHCR24, SH3BGRL3, CBR3, TGFBR2, PDLIM5, BDKRB1, SMAGP, INSIG1, SNRPA, PGD, ZYX, LPP, PAFAH1B3, ETHE1, ABL1, ISG15, EFHD2, RIT1, SLC25A1, ADAMTS5, RRM1, HRAS, RAB8A
C33	CAV1, TM4SF1, SQRDL, SERPINE2, FOXD1, DSP, LGALS1, MCM7, SRGN, TFRC, MCFD2, TPM2, ARPC1A, NT5E, AUTS2, C16orf45, ANXA6, TJP2, BNC2, GPX1, ACAT1, HMGA2,

	WNT5A, C7orf10, SLBP, WNT5B, YWHAE, FZD6, GRAMD3, AHR, KIAA1462, SACS, TNFAIP8, MEIS2, SWAP70, SF3A3, IER5, TPST2, PNMA1
C34	TUBA1B, PRPS1, PSMD1, SAE1, THBS1, LIMA1, HEG1, ENDOD1, GALNT10, ACTA2, PSMD14, HSP90AA1, ACTB, SPCS3, MAPRE1, ZWINT, PDHB, EXT2, THY1, MAP2K3, HSPH1, PSMD11, TMEM50A, IDE, H2AFV, ACTR3, GMFB, AIFM1, SFRS2, YWHAB, SFRS1, ACTR2, ANGPT1, ANXA7, CRY1, NEDD4, COLEC12, LSM4, SLC20A1, HNRNPA2B1, CMTM6, DNAJA1, ANXA1, UXS1, MYL12A, ACTN4, TFDP1, EBNA1BP2, HSPA4, LIPA, WSB2, OSTM1, TULP4, PKD2, AMD1, ARHGAP17, TMEM47, LRRC8D, ZWILCH, HIF1A, UQCRFS1, NPTN, LHFP, TRIM37, MAP7D1, CLIC4, VRK1, TTPAL, OXCT1, SCP2, NDUFS1, SEH1L, PRKDC, EIF5A, BANF1, FAM3C, HARS, SLC9A6, SF3B4, ZNF706, VCAM1, UBP1, TOP2B, CHCHD3, ERBB2IP, GALNT7, LAP3
C35	CAP2, STAT1, DDAH1, EFNB2, GREM1, DLGAP5, PRC1, CSRP1, ATP6V0D1, EHD1, TPX2, PBK, RRM2, ROR1, LOXL1, POLR2L, BIRC5, CALD1, PRKCDBP, KIF4A, NEK7, BEX1, DNMT1, MELK, DDX39, NAPA, DNM1L, TXN, COPS7A, TK1, PTPLA, CDCA3, HJURP, GRB2, PHACTR2, DOCK5, TP53I3, ATP10D, ATP6V0E2
C36	CCNB2, KIF20A, TMEM106C, TPD52L1, CYP1B1, TCTA, GYG1, DPP3, PSMC3, NSF, TTK, RND3, NUTF2, RBMS3, EDIL3
C37	FSTL1, CENPF, CCNB1, TOP2A, ASPM, TYMS, CDKN3, NUSAP1, CDC2, FOXM1, OLFML3, BUB1B, C20orf3, CETN2
C39	RGS4, FGF2, CCL2, AMOTL2, FGF5, FST, HAS2, NRG1, HOXA9, CCNA2, CBX5, FAM111A, CENPE, KCTD20, CCDC99, SH2D4A, DKK1, IL7R, BARD1, RASA1, SMURF2, TUFT1, DEPDC1, MPP5, DOK5, NGF, CNN2, FAM13B, PPM2C, CDC25B, HOXC6
C4	COL5A1, COL5A2, PRELP, AOC2, SPP1, CLEC11A, CSGALNACT1, SCD, CKB, IRF7, PLCD1, WIPF1, GLT25D2, CLTCL1, PCOLCE, ALDH2, RABAC1, WIPI1, CLU, FKBP11, CST3, CTSD, ROBO1, LTBP1, BMP1, FKBP10, SLC22A17, GTPBP2, S100A13, SLC2A10
C40	ARHGAP29, CXCL12, TNFRSF11B, KCNK2, WDR1, RAGE, SH3BP4, MGLL, VEGFC, TACC3,

	CCND1, AURKA, MARCKS, GINS2, TUBB2C, SMAD3, LANCL2, ARSJ, BDNF, SOCS5, EXT1, PRR16, C1GALT1, DRAM, FDFT1, DTL, TRIP13, MLF1IP, C6orf145, LPXN, TBC1D2, PODXL, ADH5, VGLL3, MCM2, CKAP2, NRAS
C42	LDB2, SYNC, FZD7, FAM69A, STK17A, TUBB2A, GREM2, BCAR3, GCLM, CAV2, SPIN1
C43	TUBB6, TUBG1, RFTN1, ARPC5, NAV3, FEN1, IL6, ATIC, PRDX3, NACC2, TUBA1C, DYNLL1, EBP, PUS7, USP14, C21orf59, ACAT2, NUP107, SNTB2, NARG1, SSB, HDHD1A
C56	MEF2C, C1QTNF3, MXRA5, ACAN, PTH1R, RBP4, IBSP, PCK2, CXCR7, THBS3, PPFIBP2, DRD4, MATN2, COL9A2, S100A1, GTF3C1, FAM129A, SRI, EYA2, LEPRE1, MIA, CHAC1, PDGFD, ECM2, KRT16, TRPV4, WNT11, REEP1, SEMA3B, STX18, PPA1, EFHD1, HOMER2, WFDC1, UGCGL2, NES, SC65, OLFML2B, KDELC1, ENTPD3, LRRC8E, ADAM22, CHAD, SLC38A3, RNFT1, MT1G
C57	C4orf31, C13orf15, EHD3, GPC1, SLC7A5, CHST3, SDC2, LITAF, CMAH, KRT14, RAMP1, PHGDH, SLC31A1, CSPG4, CYB5R1, PAMR1, F13A1, C11orf75, MATN4
C58	C4orf31, C13orf15, EHD3, GPC1, SLC7A5, CHST3, SDC2, LITAF, CMAH, KRT14, KRT16, TRPV4, WNT11, REEP1, RAMP1, PHGDH, SEMA3B, STX18, PPA1, EFHD1, HOMER2, WFDC1, UGCGL2, SLC31A1, CSPG4, NES, CYB5R1, SC65, PAMR1, OLFML2B, F13A1, KDELC1, ENTPD3, C11orf75, LRRC8E, ADAM22, CHAD, SLC38A3, MATN4, RNFT1, MT1G
C7	SPARCL1, CHI3L2, EIF3E, EIF4A2, TNFAIP6, CCL20, RPL11, SMC6, PLAT, TMEM100, IL8
C9	NBL1, DCN, CPE, CRLF1, TGFB3, LAMA2, CSRP2, BMP2, HCFC1R1, HIST1H2AC, CGRRF1, PNMAL1, EFNA1

Table A5.3. Genes in each component of the chondrogenesis model.

BIBLIOGRAPHY

[1] Wikipedia, "Human musculoskeletal system," 2013.

[2] R. L. Lieber, *Skeletal Muscle Structure and Function*. Wiliams & Wilkins, 1992.

[3] S. T. Modules, "Structure of skeletal muscle."

[4] S. T. Modules, "Structure of skeletal muscle."

[5] P. Hopkins, *Voluntary motor systems—skeletal muscle, reflexes, and control of movement*. In: Hemmings HC & Hopkins PM, eds. Foundations of Anesthesia, 2nd edn. Mosby, London, 2005.

[6] W. Scott, J. Stevens, and S. A. Binder-Macleod, "Human skeletal muscle fiber type classifications.," *Phys Ther*, vol. 81, pp. 1810–1816, Nov 2001.

[7] E. V. Ariano MA, Armstrong RB, "Hindlimb muscle fiber populations of five mammals," *The Journal of Histochemistry and Cytochemistry*, vol. 21 (1), p. 51–5, 1973.

[8] K. J. J. E. S. B. Gollnick PD, Sjödin B, "Human soleus muscle: a comparison of fiber composition and enzyme activities with other leg muscles.," *Pflügers Archiv*, vol. 3, p. 247–55, April 1974.

[9] D. J. Glass, "Skeletal muscle hypertrophy and atrophy signaling pathways.," *Int J Biochem Cell Biol*, vol. 37, pp. 1974–1984, Oct 2005.

[10] C. Rommel, S. C. Bodine, B. A. Clarke, R. Rossman, L. Nunez, T. N. Stitt, G. D. Yancopoulos, and D. J. Glass, "Mediation of igf-1-induced skeletal myotube hypertrophy by pi(3)k/akt/mtor and pi(3)k/akt/gsk3 pathways.," *Nat Cell Biol*, vol. 3, pp. 1009–1013, Nov 2001.

[11] S. C. Bodine, T. N. Stitt, M. Gonzalez, W. O. Kline, G. L. Stover, R. Bauerlein, E. Zlotchenko, A. Scrimgeour, J. C. Lawrence, D. J. Glass, and G. D. Yancopoulos, "Akt/mtor pathway is a crucial regulator of skeletal muscle hypertrophy and can prevent muscle atrophy in vivo.," *Nat Cell Biol*, vol. 3, pp. 1014–1019, Nov 2001.

[12] S.-E. Chen, B. Jin, and Y.-P. Li, "Tnf-alpha regulates myogenesis and muscle regeneration by activating p38 mapk.," *Am J Physiol Cell Physiol*, vol. 292, pp. C1660–C1671, May 2007.

[13] R. C. J. Langen, J. L. J. V. D. Velden, A. M. W. J. Schols, M. C. J. M. Kelders, E. F. M. Wouters, and Y. M. W. Janssen-Heininger, "Tumor necrosis factor-alpha inhibits myogenic differentiation through myod protein destabilization.," *FASEB J*, vol. 18, pp. 227–237, Feb 2004.

[14] R. N. Cooney, G. O. Maish, T. Gilpin, M. L. Shumate, C. H. Lang, and T. C. Vary, "Mechanism of il-1 induced inhibition of protein synthesis in skeletal muscle.," *Shock*, vol. 11, pp. 235–241, Apr 1999.

[15] S. R. Broussard, R. H. McCusker, J. E. Novakofski, K. Strle, W. H. Shen, R. W. Johnson, R. Dantzer, and K. W. Kelley, "Il-1beta impairs insulin-like growth factor i-induced differentiation and downstream activation signals of the insulin-like growth factor i receptor in myoblasts.," *J Immunol*, vol. 172, pp. 7713–7720, Jun 2004.

[16] R. C. Langen, A. M. Schols, M. C. Kelders, E. F. Wouters, and Y. M. Janssen-Heininger, "Inflammatory cytokines inhibit myogenic differentiation through activation of nuclear factor-kappab.," *FASEB J*, vol. 15, pp. 1169–1180, May 2001.

[17] J. G. Cannon, R. A. Fielding, M. A. Fiatarone, S. F. Orencole, C. A. Dinarello, and W. J. Evans, "Increased interleukin 1 beta in human skeletal muscle after exercise.," *Am J Physiol*, vol. 257, pp. R451–R455, Aug 1989.

[18] Z. Mackiewicz, M. Hukkanen, D. Povilenaite, A. Sukura, J. E. Fonseca, I. Virtanen, and Y. T. Konttinen, "Dual effects of caspase-1, interleukin-1 beta, tumour necrosis factor-alpha and nerve growth factor receptor in inflammatory myopathies.," *Clin Exp Rheumatol*, vol. 21, no. 1, pp. 41–48, 2003.

[19] J. Lin, C. Handschin, and B. M. Spiegelman, "Metabolic control through the pgc-1 family of transcription coactivators.," *Cell Metab*, vol. 1, pp. 361–370, Jun 2005.

[20] J. Lin, H. Wu, P. T. Tarr, C.-Y. Zhang, Z. Wu, O. Boss, L. F. Michael, P. Puigserver, E. Isotani, E. N. Olson, B. B. Lowell, R. Bassel-Duby, and B. M. Spiegelman, "Transcriptional co-activator pgc-1 alpha drives the formation of slow-twitch muscle fibres.," *Nature*, vol. 418, pp. 797–801, Aug 2002.

[21] Z. Gerhart-Hines, J. T. Rodgers, O. Bare, C. Lerin, S.-H. Kim, R. Mostoslavsky, F. W. Alt, Z. Wu, and P. Puigserver, "Metabolic control of muscle mitochondrial function and fatty acid oxidation through sirt1/pgc-1alpha.," *EMBO J*, vol. 26, pp. 1913–1923, Apr 2007.

[22] S. L. McGee and M. Hargreaves, "Histone modifications and skeletal muscle metabolic gene expression.," *Clin Exp Pharmacol Physiol*, vol. 37, pp. 392–396, Mar 2010.

[23] S. L. McGee, B. J. W. van Denderen, K. F. Howlett, J. Mollica, J. D. Schertzer, B. E. Kemp, and M. Hargreaves, "Amp-activated protein kinase regulates glut4 transcription by phosphorylating histone deacetylase 5.," *Diabetes*, vol. 57, pp. 860–867, Apr 2008.

[24] F. P. C. L. J. Julie C. Crockett, Michael J.Rogers and M. H. Helfrich, "Bone remodelling at a glance," *Journal of Cell Science*, vol. 124, pp. 991–998, 2011.

[25] A. J. Friedenstein, K. V. Petrakova, A. I. Kurolesova, and G. P. Frolova, "Heterotopic of bone marrow. analysis of precursor cells for osteogenic and hematopoietic tissues.," *Transplantation*, vol. 6, pp. 230–247, Mar 1968.

[26] P. A. Zuk, M. Zhu, H. Mizuno, J. Huang, J. W. Futrell, A. J. Katz, P. Benhaim, H. P. Lorenz, and M. H. Hedrick, "Multilineage cells from human adipose tissue: implications for cell-based therapies.," *Tissue Eng*, vol. 7, pp. 211–228, Apr 2001.

[27] J. G. Toma, I. A. McKenzie, D. Bagli, and F. D. Miller, "Isolation and characterization of multipotent skin-derived precursors from human skin.," *Stem Cells*, vol. 23, no. 6, pp. 727–737, 2005.

[28] C. Campagnoli, I. A. Roberts, S. Kumar, P. R. Bennett, I. Bellantuono, and N. M. Fisk, "Identification of mesenchymal stem/progenitor cells in human first-trimester fetal blood, liver, and bone marrow.," *Blood*, vol. 98, pp. 2396–2402, Oct 2001.

[29] C. Rosada, J. Justesen, D. Melsvik, P. Ebbesen, and M. Kassem, "The human umbilical cord blood: a potential source for osteoblast progenitor cells.," *Calcif Tissue Int*, vol. 72, pp. 135–142, Feb 2003.

[30] B. C. Perry, D. Zhou, X. Wu, F.-C. Yang, M. A. Byers, T.-M. G. Chu, J. J. Hockema, E. J. Woods, and W. S. Goebel, "Collection, cryopreservation, and characterization of human dental pulp-derived mesenchymal stem cells for banking and clinical use.," *Tissue Eng Part C Methods*, vol. 14, pp. 149–156, Jun 2008.

[31] K. L. Seeberger, A. Eshpeter, and G. S. Korbutt, "Isolation and culture of human multipotent stromal cells from the pancreas.," *Methods Mol Biol*, vol. 698, pp. 123–140, 2011.

[32] M. F. Pittenger, A. M. Mackay, S. C. Beck, R. K. Jaiswal, R. Douglas, J. D. Mosca, M. A. Moorman, D. W. Simonetti, S. Craig, and D. R. Marshak, "Multilineage potential of adult human mesenchymal stem cells.," *Science*, vol. 284, pp. 143–147, Apr 1999.

[33] M. Dominici, K. L. Blanc, I. Mueller, I. Slaper-Cortenbach, F. Marini, D. Krause, R. Deans, A. Keating, D. Prockop, and E. Horwitz, "Minimal criteria for defining multipotent mesenchymal stromal cells. the international society for cellular therapy position statement.," *Cytotherapy*, vol. 8, no. 4, pp. 315–317, 2006.

[34] M. A. Haniffa, M. P. Collin, C. D. Buckley, and F. Dazzi, "Mesenchymal stem cells: the fibroblasts' new clothes?," *Haematologica*, vol. 94, pp. 258–263, Feb 2009.

[35] OSTEOCORD.

[36] A. I. Caplan and S. P. Bruder, "Mesenchymal stem cells: building blocks for molecular medicine in the 21st century.," *Trends Mol Med*, vol. 7, pp. 259–264, Jun 2001.

[37] N. R. Jørgensen, Z. Henriksen, O. H. Sørensen, and R. Civitelli, "Dexamethasone, bmp-2, and 1,25-dihydroxyvitamin d enhance a more differentiated osteoblast phenotype: validation of an in vitro model for human bone marrow-derived primary osteoblasts.," *Steroids*, vol. 69, pp. 219–226, Apr 2004.

[38] E. Piek, L. S. Sleumer, E. P. van Someren, L. Heuver, J. R. de Haan, I. de Grijs, C. Gilissen, J. M. Hendriks, R. I. van Ravestein-van Os, S. Bauerschmidt, K. J. Dechering, and E. J. van Zoelen, "Osteo-transcriptomics of human mesenchymal stem cells: accelerated gene expression and osteoblast differentiation induced by vitamin d reveals c-myc as an enhancer of bmp2-induced osteogenesis.," *Bone*, vol. 46, pp. 613–627, Mar 2010.

[39] L. Ling, V. Nurcombe, and S. M. Cool, "Wnt signaling controls the fate of mesenchymal stem cells.," *Gene*, vol. 433, pp. 1–7, Mar 2009.

[40] G. L. Lin and K. D. Hankenson, "Integration of bmp, wnt, and notch signaling pathways in osteoblast differentiation.," *J Cell Biochem*, vol. 112, pp. 3491–3501, Dec 2011.

[41] C. N. Bennett, K. A. Longo, W. S. Wright, L. J. Suva, T. F. Lane, K. D. Hankenson, and O. A. MacDougald, "Regulation of osteoblastogenesis and bone mass by wnt10b.," *Proc Natl Acad Sci U S A*, vol. 102, pp. 3324–3329, Mar 2005.

[42] C. N. Bennett, H. Ouyang, Y. L. Ma, Q. Zeng, I. Gerin, K. M. Sousa, T. F. Lane, V. Krishnan, K. D. Hankenson, and O. A. MacDougald, "Wnt10b increases postnatal bone formation by enhancing osteoblast differentiation.," *J Bone Miner Res*, vol. 22, pp. 1924–1932, Dec 2007.

[43] P. V. N. Bodine, W. Zhao, Y. P. Kharode, F. J. Bex, A.-J. Lambert, M. B. Goad, T. Gaur, G. S. Stein, J. B. Lian, and B. S. Komm, "The wnt antagonist secreted frizzled-related protein-1 is a negative regulator of trabecular bone formation in adult mice.," *Mol Endocrinol*, vol. 18, pp. 1222–1237, May 2004.

[44] J. de Boer, R. Siddappa, C. Gaspar, A. van Apeldoorn, R. Fodde, and C. van Blitterswijk, "Wnt signaling inhibits osteogenic differentiation of human mesenchymal stem cells.," *Bone*, vol. 34, pp. 818–826, May 2004.

[45] Y. Zhao and S. Ding, "A high-throughput sirna library screen identifies osteogenic suppressors in human mesenchymal stem cells.," *Proc Natl Acad Sci U S A*, vol. 104, pp. 9673–9678, Jun 2007.

[46] N. Indrawattana, G. Chen, M. Tadokoro, L. H. Shann, H. Ohgushi, T. Tateishi, J. Tanaka, and A. Bunyaratvej, "Growth factor combination for chondrogenic induction from human mesenchymal stem cell.," *Biochem Biophys Res Commun*, vol. 320, pp. 914–919, Jul 2004.

[47] L. Longobardi, L. O'Rear, S. Aakula, B. Johnstone, K. Shimer, A. Chytil, W. A. Horton, H. L. Moses, and A. Spagnoli, "Effect of igf-i in the chondrogenesis of bone marrow mesenchymal stem cells in the presence or absence of tgf-beta signaling.," *J Bone Miner Res*, vol. 21, pp. 626–636, Apr 2006.

[48] W. Bi, J. M. Deng, Z. Zhang, R. R. Behringer, and B. de Crombrugghe, "Sox9 is required for cartilage formation.," *Nat Genet*, vol. 22, pp. 85–89, May 1999.

[49] S. Murakami, M. Kan, W. L. McKeehan, and B. de Crombrugghe, "Up-regulation of the chondrogenic sox9 gene by fibroblast growth factors is mediated by the mitogen-activated protein kinase pathway.," *Proc Natl Acad Sci U S A*, vol. 97, pp. 1113–1118, Feb 2000.

[50] I. Sekiya, B. L. Larson, J. T. Vuoristo, R. L. Reger, and D. J. Prockop, "Comparison of effect of bmp-2, -4, and -6 on in vitro cartilage formation of human adult stem cells from bone marrow stroma.," *Cell Tissue Res*, vol. 320, pp. 269–276, May 2005.

[51] B. Johnstone, T. M. Hering, A. I. Caplan, V. M. Goldberg, and J. U. Yoo, "In vitro chondrogenesis of bone marrow-derived mesenchymal progenitor cells.," *Exp Cell Res*, vol. 238, pp. 265–272, Jan 1998.

[52] C. A. Hellingman, E. N. B. Davidson, W. Koevoet, E. L. Vitters, W. B. van den Berg, G. J. V. M. van Osch, and P. M. van der Kraan, "Smad signaling determines chondrogenic differentiation of bone-marrow-derived mesenchymal stem cells: inhibition of smad1/5/8p prevents terminal differentiation and calcification.," *Tissue Eng Part A*, vol. 17, pp. 1157–1167, Apr 2011.

[53] T. E. Hardingham, R. A. Oldershaw, and S. R. Tew, "Cartilage, sox9 and notch signals in chondrogenesis.," *J Anat*, vol. 209, pp. 469–480, Oct 2006.

[54] G. Nalesso, J. Sherwood, J. Bertrand, T. Pap, M. Ramachandran, C. De Bari, C. Pitzalis, and F. Dell'accio, "Wnt-3a modulates articular chondrocyte phenotype by activating both canonical and noncanonical pathways.," *J Cell Biol*, vol. 193, pp. 551–564, May 2011.

[55] F. Zaucke, R. Dinser, P. Maurer, and M. Paulsson, "Cartilage oligomeric matrix protein (comp) and collagen ix are sensitive markers for the differentiation state of articular primary chondrocytes.," *Biochem J*, vol. 358, pp. 17–24, Aug 2001.

[56] S. P. Grogan, T. Olee, K. Hiraoka, and M. K. Lotz, "Repression of chondrogenesis through binding of notch signaling proteins hes-1 and hey-1 to n-box domains in the col2a1 enhancer site.," *Arthritis Rheum*, vol. 58, pp. 2754–2763, Sep 2008.

[57] T. J. Mead and K. E. Yutzey, "Notch pathway regulation of chondrocyte differentiation and proliferation during appendicular and axial skeleton development.," *Proc Natl Acad Sci U S A*, vol. 106, pp. 14420–14425, Aug 2009.

[58] A. G. N. Agustí, "Systemic effects of chronic obstructive pulmonary disease.," *Proc Am Thorac Soc*, vol. 2, no. 4, pp. 367–70; discussion 371–2, 2005.

[59] WHO, *World Health Statistics*. WHO, 2008.

[60] R. C. I. Wüst and H. Degens, "Factors contributing to muscle wasting and dysfunction in copd patients.," *Int J Chron Obstruct Pulmon Dis*, vol. 2, no. 3, pp. 289–300, 2007.

[61] J. Garcia-Aymerich, P. Lange, M. Benet, P. Schnohr, and J. M. Antó, "Regular physical activity reduces hospital admission and mortality in chronic obstructive pulmonary disease: a population based cohort study.," *Thorax*, vol. 61, pp. 772–778, Sep 2006.

[62] E. B. Swallow, D. Reyes, N. S. Hopkinson, W. D.-C. Man, R. Porcher, E. J. Cetti, A. J. Moore, J. Moxham, and M. I. Polkey, "Quadriceps strength predicts mortality in patients with moderate to severe chronic obstructive pulmonary disease.," *Thorax*, vol. 62, pp. 115–120, Feb 2007.

[63] J. M. Seymour, M. A. Spruit, N. S. Hopkinson, S. A. Natanek, W. D.-C. Man, A. Jackson, H. R. Gosker, A. M. W. J. Schols, J. Moxham, M. I. Polkey, and E. F. M. Wouters, "The prevalence of quadriceps weakness in copd and the relationship with disease severity.," *Eur Respir J*, vol. 36, pp. 81–88, Jul 2010.

[64] S. Bernard, P. LeBlanc, F. Whittom, G. Carrier, J. Jobin, R. Belleau, and F. Maltais, "Peripheral muscle weakness in patients with chronic obstructive pulmonary disease.," *Am J Respir Crit Care Med*, vol. 158, pp. 629–634, Aug 1998.

[65] F. Whittom, J. Jobin, P. M. Simard, P. Leblanc, C. Simard, S. Bernard, R. Belleau, and F. Maltais, "Histochemical and morphological characteristics of the vastus lateralis muscle in patients with chronic obstructive pulmonary disease.," *Med Sci Sports Exerc*, vol. 30, pp. 1467–1474, Oct 1998.

[66] H. R. Gosker, B. Kubat, G. Schaart, G. J. van der Vusse, E. F. M. Wouters, and A. M. W. J. Schols, "Myopathological features in skeletal muscle of patients with chronic obstructive pulmonary disease.," *Eur Respir J*, vol. 22, pp. 280–285, Aug 2003.

[67] E. Barreiro, A. M. W. J. Schols, M. I. Polkey, J. B. Galdiz, H. R. Gosker, E. B. Swallow, C. Coronell, J. Gea, and E. N. I. G. M. A. in COPD project, "Cytokine profile in quadriceps muscles of patients with severe copd.," *Thorax*, vol. 63, pp. 100–107, Feb 2008.

[68] F. Maltais, P. LeBlanc, F. Whittom, C. Simard, K. Marquis, M. Bélanger, M. J. Breton, and J. Jobin, "Oxidative enzyme activities of the vastus lateralis muscle and the functional status in patients with copd.," *Thorax*, vol. 55, pp. 848–853, Oct 2000.

[69] H. R. Gosker, M. K. C. Hesselink, H. Duimel, K. A. Ward, and A. M. W. J. Schols, "Reduced mitochondrial density in the vastus lateralis muscle of patients with copd.," *Eur Respir J*, vol. 30, pp. 73–79, Jul 2007.

[70] E. M. Pouw, A. M. Schols, G. J. van der Vusse, and E. F. Wouters, "Elevated inosine monophosphate levels in resting muscle of patients with stable chronic obstructive pulmonary disease.," *Am J Respir Crit Care Med*, vol. 157, pp. 453–457, Feb 1998.

[71] M. C. Steiner, R. Evans, S. J. Deacon, S. J. Singh, P. Patel, J. Fox, P. L. Greenhaff, and M. D. L. Morgan, "Adenine nucleotide loss in the skeletal muscles during exercise in chronic obstructive pulmonary disease.," *Thorax*, vol. 60, pp. 932–936, Nov 2005.

[72] N. J. Sinden and R. A. Stockley, "Systemic inflammation and comorbidity in copd: a result of 'overspill' of inflammatory mediators from the lungs? review of the evidence.," *Thorax*, vol. 65, pp. 930–936, Oct 2010.

[73] W. Q. Gan, S. F. P. Man, A. Senthilselvan, and D. D. Sin, "Association between chronic obstructive pulmonary disease and systemic inflammation: a systematic review and a meta-analysis.," *Thorax*, vol. 59, pp. 574–580, Jul 2004.

[74] R. C. J. Langen, A. M. W. J. Schols, M. C. J. M. Kelders, J. L. J. van der Velden, E. F. M. Wouters, and Y. M. W. Janssen-Heininger, "Muscle wasting and impaired muscle regeneration in a murine model of chronic pulmonary inflammation.," *Am J Respir Cell Mol Biol*, vol. 35, pp. 689–696, Dec 2006.

[75] M. Sandri, C. Sandri, A. Gilbert, C. Skurk, E. Calabria, A. Picard, K. Walsh, S. Schiaffino, S. H. Lecker, and A. L. Goldberg, "Foxo transcription factors induce the atrophy-related ubiquitin ligase atrogin-1 and cause skeletal muscle atrophy.," *Cell*, vol. 117, pp. 399–412, Apr 2004.

[76] S. C. Bodine, E. Latres, S. Baumhueter, V. K. Lai, L. Nunez, B. A. Clarke, W. T. Poueymirou, F. J. Panaro, E. Na, K. Dharmarajan, Z. Q. Pan, D. M. Valenzuela, T. M. DeChiara, T. N. Stitt, G. D. Yancopoulos, and D. J. Glass, "Identification of ubiquitin ligases required for skeletal muscle atrophy.," *Science*, vol. 294, pp. 1704–1708, Nov 2001.

[77] Y.-P. Li, Y. Chen, J. John, J. Moylan, B. Jin, D. L. Mann, and M. B. Reid, "Tnf-alpha acts via p38 mapk to stimulate expression of the ubiquitin ligase atrogin1/mafbx in skeletal muscle.," *FASEB J*, vol. 19, pp. 362–370, Mar 2005.

[78] J. Kim, K.-J. Won, H. M. Lee, B.-Y. Hwang, Y.-M. Bae, W. S. Choi, H. Song, K. W. Lim, C.-K. Lee, and B. Kim, "p38 mapk participates in muscle-specific ring finger 1-mediated atrophy in cast-immobilized rat gastrocnemius muscle.," *Korean J Physiol Pharmacol*, vol. 13, pp. 491–496, Dec 2009.

[79] D. C. Guttridge, M. W. Mayo, L. V. Madrid, C. Y. Wang, and A. S. Baldwin, “Nf-kappab-induced loss of myod messenger rna: possible role in muscle decay and cachexia.,” *Science*, vol. 289, pp. 2363–2366, Sep 2000.

[80] J. Riddoch-Contreras, T. George, S. A. Natanek, G. S. Marsh, N. S. Hopkinson, R. Tal-Singer, P. Kemp, and M. I. Polkey, “p38 mitogen-activated protein kinase is not activated in the quadriceps of patients with stable chronic obstructive pulmonary disease.,” *COPD*, vol. 9, pp. 142–150, Apr 2012.

[81] E. M. Mercken, G. J. Hageman, R. C. Langen, E. F. Wouters, and A. M. Schols, “Decreased exercise-induced expression of nuclear factor-?b-regulated genes in muscle of patients with copd.,” *Chest*, vol. 139, pp. 337–346, Feb 2011.

[82] F. B. Favier, F. Costes, A. Defour, R. Bonnefoy, E. Lefai, S. Baugé, A. Peinnequin, H. Benoit, and D. Freyssenet, “Downregulation of akt/mammalian target of rapamycin pathway in skeletal muscle is associated with increased redd1 expression in response to chronic hypoxia.,” *Am J Physiol Regul Integr Comp Physiol*, vol. 298, pp. R1659–R1666, Jun 2010.

[83] G. A. Nader, “Molecular determinants of skeletal muscle mass: getting the "akt" together.,” *Int J Biochem Cell Biol*, vol. 37, pp. 1985–1996, Oct 2005.

[84] G. Ferretti, H. Hauser, and P. E. di Prampero, “Maximal muscular power before and after exposure to chronic hypoxia.,” *Int J Sports Med*, vol. 11 Suppl 1, pp. S31–S34, Feb 1990.

[85] B. Kayser, H. Hoppeler, H. Claassen, and P. Cerretelli, “Muscle structure and performance capacity of himalayan sherpas.,” *J Appl Physiol*, vol. 70, pp. 1938–1942, May 1991.

[86] H. Hoppeler, E. Kleinert, C. Schlegel, H. Claassen, H. Howald, S. R. Kayar, and P. Cerretelli, “Morphological adaptations of human skeletal muscle to chronic hypoxia.,” *Int J Sports Med*, vol. 11 Suppl 1, pp. S3–S9, Feb 1990.

[87] M. B. Reid and Y. P. Li, “Cytokines and oxidative signalling in skeletal muscle.,” *Acta Physiol Scand*, vol. 171, pp. 225–232, Mar 2001.

[88] G. Williams, T. Brown, L. Becker, M. Prager, and B. P. Giroir, “Cytokine-induced expression of nitric oxide synthase in c2c12 skeletal muscle myocytes.,” *Am J Physiol*, vol. 267, pp. R1020–R1025, Oct 1994.

[89] R. C. J. Langen, A. M. W. J. Schols, M. C. J. M. Kelders, J. L. J. V. D. Velden, E. F. M. Wouters, and Y. M. W. Janssen-Heininger, “Tumor necrosis factor-alpha inhibits myogenesis through redox-dependent and -independent pathways.,” *Am J Physiol Cell Physiol*, vol. 283, pp. C714–C721, Sep 2002.

[90] A. Couillard, F. Maltais, D. Saey, R. Debigaré, A. Michaud, C. Koechlin, P. LeBlanc, and C. Préfaut, “Exercise-induced quadriceps oxidative stress and peripheral muscle dysfunction in patients with chronic obstructive pulmonary disease.,” *Am J Respir Crit Care Med*, vol. 167, pp. 1664–1669, Jun 2003.

[91] C. Koechlin, F. Maltais, D. Saey, A. Michaud, P. LeBlanc, M. Hayot, and C. Préfaut, "Hypoxaemia enhances peripheral muscle oxidative stress in chronic obstructive pulmonary disease.," *Thorax*, vol. 60, pp. 834–841, Oct 2005.

[92] R. Debigaré and F. Maltais, *Physiologic basis of respiratory disease, chapter 49: Peripheral Muscle Dysfunction in COPD*. BC Decker Inc, 2005.

[93] V. Kim, J. O. Benditt, R. A. Wise, and A. Sharafkhaneh, "Oxygen therapy in chronic obstructive pulmonary disease.," *Proc Am Thorac Soc*, vol. 5, pp. 513–518, May 2008.

[94] R. A. Rabinovich, E. Ardite, T. Troosters, N. Carbó, J. Alonso, J. M. G. de Suso, J. Vilaró, J. A. Barberà, M. F. Polo, J. M. Argilés, J. C. Fernandez-Checa, and J. Roca, "Reduced muscle redox capacity after endurance training in patients with chronic obstructive pulmonary disease.," *Am J Respir Crit Care Med*, vol. 164, pp. 1114–1118, Oct 2001.

[95] R. Casaburi, S. Bhasin, L. Cosentino, J. Porszasz, A. Somfay, M. I. Lewis, M. Fournier, and T. W. Storer, "Effects of testosterone and resistance training in men with chronic obstructive pulmonary disease.," *Am J Respir Crit Care Med*, vol. 170, pp. 870–878, Oct 2004.

[96] I. M. Ferreira, D. Brooks, Y. Lacasse, R. S. Goldstein, and J. White, "Nutritional supplementation for stable chronic obstructive pulmonary disease.," *Cochrane Database Syst Rev*, no. 2, p. CD000998, 2005.

[97] G. Faager, K. Söderlund, C. M. Sköld, S. Rundgren, A. Tollbäck, and P. Jakobsson, "Creatine supplementation and physical training in patients with copd: a double blind, placebo-controlled study.," *Int J Chron Obstruct Pulmon Dis*, vol. 1, no. 4, pp. 445–453, 2006.

[98] S. Marwood, S. Jack, M. Patel, P. Walker, J. Bowtell, and P. Calverley, "No effect of glutamine ingestion on indices of oxidative metabolism in stable copd.," *Respir Physiol Neurobiol*, vol. 177, pp. 41–46, Jun 2011.

[99] P. R. Greipp, J. S. Miguel, B. G. M. Durie, J. J. Crowley, B. Barlogie, J. Bladé, M. Boccadoro, J. A. Child, H. Avet-Loiseau, J.-L. Harousseau, R. A. Kyle, J. J. Lahuerta, H. Ludwig, G. Morgan, R. Powles, K. Shimizu, C. Shustik, P. Sonneveld, P. Tosi, I. Turesson, and J. Westin, "International staging system for multiple myeloma.," *J Clin Oncol*, vol. 23, pp. 3412–3420, May 2005.

[100] R. A. Kyle and S. V. Rajkumar, "Criteria for diagnosis, staging, risk stratification and response assessment of multiple myeloma.," *Leukemia*, vol. 23, pp. 3–9, Jan 2009.

[101] P. P. Carbone, L. E. Kellerhouse, and E. A. Gehan, "Plasmacytic myeloma. a study of the relationship of survival to various clinical manifestations and anomalous protein type in 112 patients.," *Am J Med*, vol. 42, pp. 937–948, Jun 1967.

[102] A. D. Woolf and B. Pfleger, "Burden of major musculoskeletal conditions.," *Bull World Health Organ*, vol. 81, no. 9, pp. 646–656, 2003.

[103] J. Justesen, K. Stenderup, E. N. Ebbesen, L. Mosekilde, T. Steiniche, and M. Kassem, "Adipocyte tissue volume in bone marrow is increased with aging and in patients with osteoporosis.," *Biogerontology*, vol. 2, no. 3, pp. 165–171, 2001.

[104] E. J. Moerman, K. Teng, D. A. Lipschitz, and B. Lecka-Czernik, "Aging activates adipogenic and suppresses osteogenic programs in mesenchymal marrow stroma/stem cells: the role of ppar-gamma2 transcription factor and tgf-beta/bmp signaling pathways.," *Aging Cell*, vol. 3, pp. 379–389, Dec 2004.

[105] Y. Wan, L.-W. Chong, and R. M. Evans, "Ppar-gamma regulates osteoclastogenesis in mice.," *Nat Med*, vol. 13, pp. 1496–1503, Dec 2007.

[106] J. Corre, K. Mahtouk, M. Attal, M. Gadelorge, A. Huynh, S. Fleury-Cappellesso, C. Danho, P. Laharrague, B. Klein, T. Rème, and P. Bourin, "Bone marrow mesenchymal stem cells are abnormal in multiple myeloma.," *Leukemia*, vol. 21, pp. 1079–1088, May 2007.

[107] S. Xu, H. Evans, C. Buckle, K. D. Veirman, J. Hu, D. Xu, E. Menu, A. D. Becker, I. V. Broek, X. Leleu, B. V. Camp, P. Croucher, K. Vanderkerken, and I. V. Riet, "Impaired osteogenic differentiation of mesenchymal stem cells derived from multiple myeloma patients is associated with a blockade in the deactivation of the notch signaling pathway.," *Leukemia*, May 2012.

[108] G. R. Mundy, L. G. Raisz, R. A. Cooper, G. P. Schechter, and S. E. Salmon, "Evidence for the secretion of an osteoclast stimulating factor in myeloma.," *N Engl J Med*, vol. 291, pp. 1041–1046, Nov 1974.

[109] F. Silvestris, P. Cafforio, N. Calvani, and F. Dammacco, "Impaired osteoblastogenesis in myeloma bone disease: role of upregulated apoptosis by cytokines and malignant plasma cells.," *Br J Haematol*, vol. 126, pp. 475–486, Aug 2004.

[110] D. O. Gradaigh, D. Ireland, S. Bord, and J. E. Compston, "Joint erosion in rheumatoid arthritis: interactions between tumour necrosis factor alpha, interleukin 1, and receptor activator of nuclear factor kappab ligand (rankl) regulate osteoclasts.," *Ann Rheum Dis*, vol. 63, pp. 354–359, Apr 2004.

[111] Y. Hirashima, N. Ishiguro, S. Kondo, and H. Iwata, "Osteoclast induction from bone marrow cells is due to pro-inflammatory mediators from macrophages exposed to polyethylene particles: a possible mechanism of osteolysis in failed tha.," *J Biomed Mater Res*, vol. 56, pp. 177–183, Aug 2001.

[112] S. Wei, H. Kitaura, P. Zhou, F. P. Ross, and S. L. Teitelbaum, "Il-1 mediates tnf-induced osteoclastogenesis.," *J Clin Invest*, vol. 115, pp. 282–290, Feb 2005.

[113] C. T. Ritchlin, S. A. Haas-Smith, P. Li, D. G. Hicks, and E. M. Schwarz, "Mechanisms of tnf-alpha- and rankl-mediated osteoclastogenesis and bone resorption in psoriatic arthritis.," *J Clin Invest*, vol. 111, pp. 821–831, Mar 2003.

[114] T. Hirayama, L. Danks, A. Sabokbar, and N. A. Athanasou, "Osteoclast formation and activity in the pathogenesis of osteoporosis in rheumatoid arthritis.," *Rheumatology (Oxford)*, vol. 41, pp. 1232–1239, Nov 2002.

[115] M. Tsuboi, A. Kawakami, T. Nakashima, N. Matsuoka, S. Urayama, Y. Kawabe, K. Fujiyama, T. Kiriyama, T. Aoyagi, K. Maeda, and K. Eguchi, "Tumor necrosis factor-alpha and interleukin-1beta increase the fas-mediated apoptosis of human osteoblasts.," *J Lab Clin Med*, vol. 134, pp. 222–231, Sep 1999.

[116] J. M. Murphy, K. Dixon, S. Beck, D. Fabian, A. Feldman, and F. Barry, "Reduced chondrogenic and adipogenic activity of mesenchymal stem cells from patients with advanced osteoarthritis.," *Arthritis Rheum*, vol. 46, pp. 704–713, Mar 2002.

[117] D. Morimoto, S. Kuroda, T. Kizawa, K. Nomura, C. Higuchi, H. Yoshikawa, and T. Tomita, "Equivalent osteoblastic differentiation function of human mesenchymal stem cells from rheumatoid arthritis in comparison with osteoarthritis.," *Rheumatology (Oxford)*, vol. 48, pp. 643–649, Jun 2009.

[118] M.-C. Kastrinaki, P. Sidiropoulos, S. Roche, J. Ringe, S. Lehmann, H. Kritikos, V.-M. Vlahava, B. Delorme, G. D. Eliopoulos, C. Jorgensen, P. Charbord, T. Häupl, D. T. Boumpas, and H. A. Papadaki, "Functional, molecular and proteomic characterisation of bone marrow mesenchymal stem cells in rheumatoid arthritis.," *Ann Rheum Dis*, vol. 67, pp. 741–749, Jun 2008.

[119] M. Kassem and B. M. Abdallah, "Human bone-marrow-derived mesenchymal stem cells: biological characteristics and potential role in therapy of degenerative diseases.," *Cell Tissue Res*, vol. 331, pp. 157–163, Jan 2008.

[120] N. Kimelman, G. Pelled, G. A. Helm, J. Huard, E. M. Schwarz, and D. Gazit, "Review: gene- and stem cell-based therapeutics for bone regeneration and repair.," *Tissue Eng*, vol. 13, pp. 1135–1150, Jun 2007.

[121] K. Pelttari, E. Steck, and W. Richter, "The use of mesenchymal stem cells for chondrogenesis.," *Injury*, vol. 39 Suppl 1, pp. S58–S65, Apr 2008.

[122] R. Quarto, M. Mastrogiacomo, R. Cancedda, S. M. Kutepov, V. Mukhachev, A. Lavroukov, E. Kon, and M. Marcacci, "Repair of large bone defects with the use of autologous bone marrow stromal cells.," *N Engl J Med*, vol. 344, pp. 385–386, Feb 2001.

[123] E. M. Horwitz, D. J. Prockop, P. L. Gordon, W. W. Koo, L. A. Fitzpatrick, M. D. Neel, M. E. McCarville, P. J. Orchard, R. E. Pyeritz, and M. K. Brenner, "Clinical responses to bone marrow transplantation in children with severe osteogenesis imperfecta.," *Blood*, vol. 97, pp. 1227–1231, Mar 2001.

[124] S. Wakitani, K. Imoto, T. Yamamoto, M. Saito, N. Murata, and M. Yoneda, "Human autologous culture expanded bone marrow mesenchymal cell transplantation for repair of cartilage defects in osteoarthritic knees.," *Osteoarthritis Cartilage*, vol. 10, pp. 199–206, Mar 2002.

[125] J. Quackenbush, "Microarray data normalization and transformation.," *Nat Genet*, vol. 32 Suppl, pp. 496–501, Dec 2002.

[126] Z. Wu, "A review of statistical methods for preprocessing oligonucleotide microarrays.," *Stat Methods Med Res*, vol. 18, pp. 533–541, Dec 2009.

[127] R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed, "Exploration, normalization, and summaries of high density oligonucleotide array probe level data.," *Biostatistics*, vol. 4, pp. 249–264, Apr 2003.

[128] G. R. M.-M. F. S. F. Wu Z, Irizarry R, "A model-based background adjustment for oligonucleotide expression arrays.," *Journal of the American Statistical Association*, vol. 99(468), p. 909–917, 2004;.

[129] C. Li and W. H. Wong, "Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection.," *Proc Natl Acad Sci U S A*, vol. 98, pp. 31–36, Jan 2001.

[130] L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry, "affy–analysis of affymetrix genechip data at the probe level.," *Bioinformatics*, vol. 20, pp. 307–315, Feb 2004.

[131] I. Medina, J. Carbonell, L. Pulido, S. C. Madeira, S. Goetz, A. Conesa, J. Tárraga, A. Pascual-Montano, R. Nogales-Cadenas, J. Santoyo, F. García, M. Marbà, D. Montaner, and J. Dopazo, "Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling.," *Nucleic Acids Res*, vol. 38, pp. W210–W213, Jul 2010.

[132] M. J. van der Laan and K. S. Pollard, "A new algorithm for hybridhierarchical clustering with visualization and the bootstrap," *Journal of Statistical Planning and Inference*, vol. 117, pp. 275–303, 2001.

[133] J. Handl, J. Knowles, and D. B. Kell, "Computational cluster validation in post-genomic data analysis.," *Bioinformatics*, vol. 21, pp. 3201–3212, Aug 2005.

[134] R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.

[135] A. I. Saeed, N. K. Bhagabati, J. C. Braisted, W. Liang, V. Sharov, E. A. Howe, J. Li, M. Thiagarajan, J. A. White, and J. Quackenbush, "Tm4 microarray software suite.," *Methods Enzymol*, vol. 411, pp. 134–193, 2006.

[136] G. K. Seo Young Kim Research Institute for Basic Science, Chonnam National University and J. W. Lee, "Comparison of various statistical methods for identifying differential gene expression in replicated microarray data," *Statistical Methods in Medical Research*, vol. 15, pp. 3–20, 2006.

[137] Y. Hochberg and Y. Benjamini, "More powerful procedures for multiple significance testing.," *Stat Med*, vol. 9, pp. 811–818, Jul 1990.

[138] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.," *Nucleic Acids Res*, vol. 37, pp. 1–13, Jan 2009.

[139] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using david bioinformatics resources.," *Nat Protoc*, vol. 4, no. 1, pp. 44–57, 2009.

[140] F. Ortega, K. Sameith, N. Turan, R. Compton, V. Trevino, M. Vannucci, and F. Falciani, "Models and computational strategies linking physiological response to molecular networks from large-scale data.," *Philos Transact A Math Phys Eng Sci*, vol. 366, pp. 3067–3089, Sep 2008.

[141] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease.," *Nat Rev Genet*, vol. 12, pp. 56–68, Jan 2011.

[142] M. Bansal, V. Belcastro, A. Ambesi-Impiombato, and D. di Bernardo, "How to infer gene networks from expression profiles.," *Mol Syst Biol*, vol. 3, p. 78, 2007.

[143] J. Yu, V. A. Smith, P. P. Wang, A. J. Hartemink, and E. D. Jarvis, "Advances to bayesian network inference for generating causal networks from observational biological data.," *Bioinformatics*, vol. 20, pp. 3594–3603, Dec 2004.

[144] D. D. D. W. Susanne Toepfer, Reinhard Guthke and M. Pfaff, "The net gene rator algorithm: Reconstruction of gene regulatory networks," *KNOWLEDGE DISCOVERY AND EMERGENT COMPLEXITY IN BIOINFORMATICS*, vol. 4366, pp. 119–130, 2007.

[145] M. Bansal, G. D. Gatta, and D. di Bernardo, "Inference of gene regulatory networks and compound mode of action from time course gene expression profiles.," *Bioinformatics*, vol. 22, pp. 815–822, Apr 2006.

[146] R. Gupta, A. Stincone, P. Antczak, S. Durant, R. Bicknell, A. Bikfalvi, and F. Falciani, "A computational framework for gene regulatory network inference that combines multiple methods and datasets.," *BMC Syst Biol*, vol. 5, p. 52, 2011.

[147] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera, and A. Califano, "Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context.," *BMC Bioinformatics*, vol. 7 Suppl 1, p. S7, 2006.

[148] I. Avila-Campillo, K. Drew, J. Lin, D. J. Reiss, and R. Bonneau, "Bionetbuilder: automatic integration of biological networks.," *Bioinformatics*, vol. 23, pp. 392–393, Feb 2007.

[149] J. Gao, A. S. Ade, V. G. Tarcea, T. E. Weymouth, B. R. Mirel, H. V. Jagadish, and D. J. States, "Integrating and annotating the interactome using the mimi plugin for cytoscape.," *Bioinformatics*, vol. 25, pp. 137–138, Jan 2009.

[150] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: a software environment for integrated models of biomolecular interaction networks.," *Genome Res*, vol. 13, pp. 2498–2504, Nov 2003.

[151] T. Ideker, O. Ozier, B. Schwikowski, and A. F. Siegel, "Discovering regulatory and signalling circuits in molecular interaction networks.," *Bioinformatics*, vol. 18 Suppl 1, pp. S233–S240, 2002.

[152] G. D. Bader and C. W. V. Hogue, "An automated method for finding molecular complexes in large protein interaction networks.," *BMC Bioinformatics*, vol. 4, p. 2, Jan 2003.

[153] X. Li, M. Wu, C.-K. Kwoh, and S.-K. Ng, "Computational approaches for detecting protein complexes from protein interaction networks: a survey.," *BMC Genomics*, vol. 11 Suppl 1, p. S3, 2010.

[154] K. Sameith, P. Antczak, E. Marston, N. Turan, D. Maier, T. Stankovic, and F. Falciani, "Functional modules integrating essential cellular functions are predictive of the response of leukaemia cells to dna damage.," *Bioinformatics*, vol. 24, pp. 2602–2607, Nov 2008.

[155] I. A. Maraziotis, K. Dimitrakopoulou, and A. Bezerianos, "Growing functional modules from a seed protein via integration of protein interaction and gene expression data.," *BMC Bioinformatics*, vol. 8, p. 408, 2007.

[156] A. M. Schols, “Nutrition in chronic obstructive pulmonary disease.,” *Curr Opin Pulm Med*, vol. 6, pp. 110–115, Mar 2000.

[157] R. A. Pauwels, A. S. Buist, P. M. Calverley, C. R. Jenkins, S. S. Hurd, and G. O. L. D. S. Committee, “Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. nhlbi/who global initiative for chronic obstructive lung disease (gold) workshop summary.,” *Am J Respir Crit Care Med*, vol. 163, pp. 1256–1276, Apr 2001.

[158] F. Maltais, P. LeBlanc, C. Simard, J. Jobin, C. Bérubé, J. Bruneau, L. Carrier, and R. Belleau, “Skeletal muscle adaptation to endurance training in patients with chronic obstructive pulmonary disease.,” *Am J Respir Crit Care Med*, vol. 154, pp. 442–447, Aug 1996.

[159] E. Sala, J. Roca, R. M. Marrades, J. Alonso, J. M. G. D. Suso, A. Moreno, J. A. Barberá, J. Nadal, L. de Jover, R. Rodriguez-Roisin, and P. D. Wagner, “Effects of endurance training on skeletal muscle bioenergetics in chronic obstructive pulmonary disease.,” *Am J Respir Crit Care Med*, vol. 159, pp. 1726–1734, Jun 1999.

[160] E. Barreiro, R. Rabinovich, J. Marin-Corral, J. A. Barberà, J. Gea, and J. Roca, “Chronic endurance exercise induces quadriceps nitrosative stress in patients with severe copd.,” *Thorax*, vol. 64, pp. 13–19, Jan 2009.

[161] R. Debigaré, F. Maltais, C. H. Côté, A. Michaud, M.-A. Caron, M. Mofarrahi, P. Leblanc, and S. N. A. Hussain, “Profiling of mrna expression in quadriceps of patients with copd and muscle wasting.,” *COPD*, vol. 5, pp. 75–84, Apr 2008.

[162] S. Radom-Aizik, N. Kaminski, S. Hayek, H. Halkin, D. M. Cooper, and I. Ben-Dov, “Effects of exercise training on quadriceps muscle gene expression in chronic obstructive pulmonary disease.,” *J Appl Physiol*, vol. 102, pp. 1976–1984, May 2007.

[163] G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki, “David: Database for annotation, visualization, and integrated discovery.,” *Genome Biol*, vol. 4, no. 5, p. P3, 2003.

[164] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa, “Kegg: Kyoto encyclopedia of genes and genomes.,” *Nucleic Acids Res*, vol. 27, pp. 29–34, Jan 1999.

[165] K. Pearson, *On lines and planes of closest fit to a system of points in space*, 1901.

[166] Y. H. Yoav Benjamini, *On lines and planes of closest fit to a system of points in space*, 1995.

[167] R. S. Richardson, H. Wagner, S. R. Mudaliar, R. Henry, E. A. Noyszewski, and P. D. Wagner, “Human vegf gene expression in skeletal muscle: effect of acute normoxic and hypoxic exercise.,” *Am J Physiol*, vol. 277, pp. H2247–H2252, Dec 1999.

[168] A. Mal, M. Sturniolo, R. L. Schiltz, M. K. Ghosh, and M. L. Harter, “A role for histone deacetylase hdac1 in modulating the transcriptional activity of myod: inhibition of the myogenic program.,” *EMBO J*, vol. 20, pp. 1739–1753, Apr 2001.

[169] I. Rahman, “Regulation of glutathione in inflammation and chronic lung diseases.,” *Mutat Res*, vol. 579, pp. 58–80, Nov 2005.

[170] M. P. Engelen, A. M. Schols, J. D. Does, N. E. Deutz, and E. F. Wouters, "Altered glutamate metabolism is associated with reduced muscle glutathione levels in patients with emphysema.," *Am J Respir Crit Care Med*, vol. 161, pp. 98–103, Jan 2000.

[171] A. Jain, J. Mårtensson, E. Stole, P. A. Auld, and A. Meister, "Glutathione deficiency leads to mitochondrial damage in brain.," *Proc Natl Acad Sci U S A*, vol. 88, pp. 1913–1917, Mar 1991.

[172] H. R. Gosker, A. Bast, G. R. M. M. Haenen, M. A. J. G. Fischer, G. J. van der Vusse, E. F. M. Wouters, and A. M. W. J. Schols, "Altered antioxidant status in peripheral skeletal muscle of patients with copd.," *Respir Med*, vol. 99, pp. 118–125, Jan 2005.

[173] D. Maier, P. Krubasik, S. Losko, M. Hernandez, and J. Villa i Freixa, "Knowledge management for systems biology and translational medicine. experiences from the eu biobridge project," in *Proc. IEEE Int. Conf. Bioinformatics and Biomedicine Workshop BIBMW 2009*, 2009.

[174] N. Turan, S. Kalko, A. Stincone, K. Clarke, A. Sabah, K. Howlett, S. J. Curnow, D. A. Rodriguez, M. Cascante, L. O'Neill, S. Egginton, J. Roca, and F. Falciani, "A systems biology approach identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease.," *PLoS Comput Biol*, vol. 7, p. e1002129, Sep 2011.

[175] D. A. Vignali, "Multiplexed particle-based flow cytometric assays.," *J Immunol Methods*, vol. 243, pp. 243–255, Sep 2000.

[176] V. Pinto-Plata, J. Toso, K. Lee, D. Park, J. Bilello, H. Mullerova, M. M. D. Souza, R. Vessey, and B. Celli, "Profiling serum biomarkers in patients with copd: associations with clinical parameters.," *Thorax*, vol. 62, pp. 595–601, Jul 2007.

[177] T. Boes and M. Neuhäuser, "Normalization for affymetrix genechips.," *Methods Inf Med*, vol. 44, no. 3, pp. 414–417, 2005.

[178] T. D. Fletcher, *QuantPsyc: Quantitative Psychology Tools*, 2010. R package version 1.4.

[179] K. Basso, A. A. Margolin, G. Stolovitzky, U. Klein, R. Dalla-Favera, and A. Califano, "Reverse engineering of regulatory networks in human b cells.," *Nat Genet*, vol. 37, pp. 382–390, Apr 2005.

[180] A. A. Margolin, K. Wang, W. K. Lim, M. Kustagi, I. Nemenman, and A. Califano, "Reverse engineering cellular networks.," *Nat Protoc*, vol. 1, no. 2, pp. 662–671, 2006.

[181] M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P.-L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader, "Integration of biological networks and gene expression data using cytoscape.," *Nat Protoc*, vol. 2, no. 10, pp. 2366–2382, 2007.

[182] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical bayes methods.," *Biostatistics*, vol. 8, pp. 118–127, Jan 2007.

[183] V. K. Mootha, C. M. Lindgren, K.-F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstråle, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, and L. C. Groop, "Pgc-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.," *Nat Genet*, vol. 34, pp. 267–273, Jul 2003.

[184] M. Bakay, Z. Wang, G. Melcon, L. Schiltz, J. Xuan, P. Zhao, V. Sartorelli, J. Seo, E. Pegoraro, C. Angelini, B. Shneiderman, D. Escolar, Y.-W. Chen, S. T. Winokur, L. M. Pachman, C. Fan, R. Mandler, Y. Nevo, E. Gordon, Y. Zhu, Y. Dong, Y. Wang, and E. P. Hoffman, "Nuclear envelope dystrophies show a transcriptional fingerprint suggesting disruption of rb-myod pathways in muscle regeneration.," *Brain*, vol. 129, pp. 996–1013, Apr 2006.

[185] B. S. B. S. W. M. F. Q. Q. H. M. M. I. R. L. S. K. T. Budak MT, Willmann G, "Hypoxic gene regulation on mice quadriceps muscle."

[186] V. G. Tusher, R. Tibshirani, and G. Chu, "Significance analysis of microarrays applied to the ionizing radiation response.," *Proc Natl Acad Sci U S A*, vol. 98, pp. 5116–5121, Apr 2001.

[187] R. A. Fielding, T. J. Manfredi, W. Ding, M. A. Fiatarone, W. J. Evans, and J. G. Cannon, "Acute phase response in exercise. iii. neutrophil and il-1 beta accumulation in skeletal muscle.," *Am J Physiol*, vol. 265, pp. R166–R172, Jul 1993.

[188] H. R. Gosker, M. P. Zeegers, E. F. M. Wouters, and A. M. W. J. Schols, "Muscle fibre type shifting in the vastus lateralis of patients with copd is associated with disease severity: a systematic review and meta-analysis.," *Thorax*, vol. 62, pp. 944–949, Nov 2007.

[189] M. J. Mador, T. J. Kufel, and L. Pineda, "Quadriceps fatigue after cycle exercise in patients with chronic obstructive pulmonary disease.," *Am J Respir Crit Care Med*, vol. 161, pp. 447–453, Feb 2000.

[190] R. López-Alemany, M. Suelves, and P. Muñoz-Cánoves, "Plasmin generation dependent on alpha-enolase-type plasminogen receptor is required for myogenesis.," *Thromb Haemost*, vol. 90, pp. 724–733, Oct 2003.

[191] P. Bonaldo, P. Braghetta, M. Zanetti, S. Piccolo, D. Volpin, and G. M. Bressan, "Collagen vi deficiency induces early onset myopathy in the mouse: an animal model for bethlem myopathy.," *Hum Mol Genet*, vol. 7, pp. 2135–2140, Dec 1998.

[192] L. Eklund, J. Piuhola, J. Komulainen, R. Sormunen, C. Ongvarrasopone, R. Fässler, A. Muona, M. Ilves, H. Ruskoaho, T. E. Takala, and T. Pihlajaniemi, "Lack of type xv collagen causes a skeletal myopathy and cardiovascular defects in mice.," *Proc Natl Acad Sci U S A*, vol. 98, pp. 1194–1199, Jan 2001.

[193] B. K. Pedersen and A. D. Toft, "Effects of exercise on lymphocytes and cytokines.," *Br J Sports Med*, vol. 34, pp. 246–251, Aug 2000.

[194] O. Ochoa, D. Sun, S. M. Reyes-Reyna, L. L. Waite, J. E. Michalek, L. M. McManus, and P. K. Shireman, "Delayed angiogenesis and vegf production in ccr2-/- mice during impaired skeletal muscle regeneration.," *Am J Physiol Regul Integr Comp Physiol*, vol. 293, pp. R651–R661, Aug 2007.

[195] A. Germani, A. D. Carlo, A. Mangoni, S. Straino, C. Giacinti, P. Turrini, P. Biglioli, and M. C. Capogrossi, "Vascular endothelial growth factor modulates skeletal myoblast function.," *Am J Pathol*, vol. 163, pp. 1417–1428, Oct 2003.

[196] N. Arsic, S. Zacchigna, L. Zentilin, G. Ramirez-Correa, L. Pattarini, A. Salvi, G. Sinagra, and M. Giacca, "Vascular endothelial growth factor stimulates skeletal muscle regeneration in vivo.," *Mol Ther*, vol. 10, pp. 844–854, Nov 2004.

[197] W. A. Irwin, N. Bergamin, P. Sabatelli, C. Reggiani, A. Megighian, L. Merlini, P. Braghetta, M. Columbaro, D. Volpin, G. M. Bressan, P. Bernardi, and P. Bonaldo, "Mitochondrial dysfunction and apoptosis in myopathic mice with collagen vi deficiency.," *Nat Genet*, vol. 35, pp. 367–371, Dec 2003.

[198] G. G. Lavery, E. A. Walker, N. Turan, D. Rogoff, J. W. Ryder, J. M. Shelton, J. A. Richardson, F. Falciani, P. C. White, P. M. Stewart, K. L. Parker, and D. R. McMillan, "Deletion of hexose-6-phosphate dehydrogenase activates the unfolded protein response pathway and induces skeletal myopathy.," *J Biol Chem*, vol. 283, pp. 8453–8461, Mar 2008.

[199] T. A. McKinsey, C. L. Zhang, and E. N. Olson, "Signaling chromatin to make muscle.," *Curr Opin Cell Biol*, vol. 14, pp. 763–772, Dec 2002.

[200] M. Haberland, M. A. Arnold, J. McAnally, D. Phan, Y. Kim, and E. N. Olson, "Regulation of hdac9 gene expression by mef2 establishes a negative-feedback loop in the transcriptional circuitry of muscle differentiation.," *Mol Cell Biol*, vol. 27, pp. 518–525, Jan 2007.

[201] T. J. Cohen, D. S. Waddell, T. Barrientos, Z. Lu, G. Feng, G. A. Cox, S. C. Bodine, and T.-P. Yao, "The histone deacetylase hdac4 connects neural activity to muscle transcriptional reprogramming.," *J Biol Chem*, vol. 282, pp. 33752–33759, Nov 2007.

[202] V. Moresi, A. H. Williams, E. Meadows, J. M. Flynn, M. J. Potthoff, J. McAnally, J. M. Shelton, J. Backs, W. H. Klein, J. A. Richardson, R. Bassel-Duby, and E. N. Olson, "Myogenin and class ii hdacs control neurogenic muscle atrophy by inducing e3 ubiquitin ligases.," *Cell*, vol. 143, pp. 35–45, Oct 2010.

[203] M. Fulco, R. L. Schiltz, S. Iezzi, M. T. King, P. Zhao, Y. Kashiwaya, E. Hoffman, R. L. Veech, and V. Sartorelli, "Sir2 regulates skeletal muscle differentiation as a potential sensor of the redox state.," *Mol Cell*, vol. 12, pp. 51–62, Jul 2003.

[204] A. Couillard, C. Koechlin, J. P. Cristol, A. Varray, and C. Prefaut, "Evidence of local exercise-induced systemic oxidative stress in chronic obstructive pulmonary disease patients.," *Eur Respir J*, vol. 20, pp. 1123–1129, Nov 2002.

[205] H. Zhang, Y.-M. Go, and D. P. Jones, "Mitochondrial thioredoxin-2/peroxiredoxin-3 system functions in parallel with mitochondrial gsh system in protection against oxidative stress.," *Arch Biochem Biophys*, vol. 465, pp. 119–126, Sep 2007.

[206] M. Holgado-Madruga and A. J. Wong, "Gab1 is an integrator of cell death versus cell survival signals in oxidative stress.," *Mol Cell Biol*, vol. 23, pp. 4471–4484, Jul 2003.

[207] I. Rahman, J. Marwick, and P. Kirkham, "Redox modulation of chromatin remodeling: impact on histone acetylation and deacetylation, nf-kappab and pro-inflammatory gene expression.," *Biochem Pharmacol*, vol. 68, pp. 1255–1267, Sep 2004.

[208] B. Schwer, J. Bunkenborg, R. O. Verdin, J. S. Andersen, and E. Verdin, "Reversible lysine acetylation controls the activity of the mitochondrial enzyme acetyl-coa synthetase 2.," *Proc Natl Acad Sci U S A*, vol. 103, pp. 10224–10229, Jul 2006.

[209] W. C. Hallows, S. Lee, and J. M. Denu, "Sirtuins deacetylate and activate mammalian acetyl-coa synthetases.," *Proc Natl Acad Sci U S A*, vol. 103, pp. 10230–10235, Jul 2006.

[210] D. B. Lombard, F. W. Alt, H.-L. Cheng, J. Bunkenborg, R. S. Streeper, R. Mostoslavsky, J. Kim, G. Yancopoulos, D. Valenzuela, A. Murphy, Y. Yang, Y. Chen, M. D. Hirschey, R. T. Bronson, M. Haigis, L. P. Guarente, R. V. Farese, S. Weissman, E. Verdin, and B. Schwer, "Mammalian sir2 homolog sirt3 regulates global mitochondrial lysine acetylation.," *Mol Cell Biol*, vol. 27, pp. 8807–8814, Dec 2007.

[211] D. Z. Qian, X. Wang, S. K. Kachhap, Y. Kato, Y. Wei, L. Zhang, P. Atadja, and R. Pili, "The histone deacetylase inhibitor nvp-laq824 inhibits angiogenesis and has a greater antitumor effect in combination with the vascular endothelial growth factor receptor tyrosine kinase inhibitor ptk787/zk222584.," *Cancer Res*, vol. 64, pp. 6626–6634, Sep 2004.

[212] K. Jatta, G. Eliason, G. M. Portela-Gomes, L. Grimelius, O. Caro, L. Nilholm, A. Sirjsö, K. Piehl-Aulin, and S. M. Abdel-Halim, "Overexpression of von hippel-lindau protein in skeletal muscles of patients with chronic obstructive pulmonary disease.," *J Clin Pathol*, vol. 62, pp. 70–76, Jan 2009.

[213] L. Huang, "Targeting histone deacetylases for the treatment of cancer and inflammatory diseases.," *J Cell Physiol*, vol. 209, pp. 611–616, Dec 2006.

[214] W. K. Kelly and P. A. Marks, "Drug insight: Histone deacetylase inhibitors–development of the new targeted anticancer agent suberoylanilide hydroxamic acid.," *Nat Clin Pract Oncol*, vol. 2, pp. 150–157, Mar 2005.

[215] H.-Y. Lin, C.-S. Chen, S.-P. Lin, J.-R. Weng, and C.-S. Chen, "Targeting histone deacetylase in cancer therapy.," *Med Res Rev*, vol. 26, pp. 397–413, Jul 2006.

[216] M. T. Dittrich, G. W. Klau, A. Rosenwald, T. Dandekar, and T. Müller, "Identifying functional modules in protein-protein interaction networks: an integrated exact approach.," *Bioinformatics*, vol. 24, pp. i223–i231, Jul 2008.

[217] M. S. Scott, T. Perkins, S. Bunnell, F. Pepin, D. Y. Thomas, and M. Hallett, "Identifying regulatory subnetworks for a set of genes.," *Mol Cell Proteomics*, vol. 4, pp. 683–692, May 2005.

[218] P. Klein and R. Ravi., "A nearly best-possible approximation algorithm for node-weighted steiner trees.," *J. Algorithms*, vol. 19, p. 104– 115, 1995.

[219] U. P. G. W. K. P. M. M. F. Ivana Ljubic, Ren´e Weiskircher, "An algorithmic framework for the exact solution of the prize-collecting steiner tree problem," *Math. Program., Ser. B*, vol. 105, pp. 427–449, 2006.

[220] B. Lehne and T. Schlitt, "Protein-protein interaction databases: keeping up with growing interactomes.," *Hum Genomics*, vol. 3, pp. 291–297, Apr 2009.

[221] D. Maier, W. Kalus, M. Wolff, S. G. Kalko, J. Roca, I. M. de Mas, N. Turan, M. Cascante, F. Falciani, M. Hernandez, J. Villà-Freixa, and S. Losko, "Knowledge

management for systems biology a general and visually driven framework applied to translational medicine.," *BMC Syst Biol*, vol. 5, p. 38, 2011.

[222] S. Peri, J. D. Navarro, R. Amanchy, T. Z. Kristiansen, C. K. Jonnalagadda, V. Surendranath, V. Niranjan, B. Muthusamy, T. K. B. Gandhi, M. Gronborg, N. Ibarrola, N. Deshpande, K. Shanker, H. N. Shivashankar, B. P. Rashmi, M. A. Ramya, Z. Zhao, K. N. Chandrika, N. Padma, H. C. Harsha, A. J. Yatish, M. P. Kavitha, M. Menezes, D. R. Choudhury, S. Suresh, N. Ghosh, R. Saravana, S. Chandran, S. Krishna, M. Joy, S. K. Anand, V. Madavan, A. Joseph, G. W. Wong, W. P. Schiemann, S. N. Constantinescu, L. Huang, R. Khosravi-Far, H. Steen, M. Tewari, S. Ghaffari, G. C. Blobe, C. V. Dang, J. G. N. Garcia, J. Pevsner, O. N. Jensen, P. Roepstorff, K. S. Deshpande, A. M. Chinnaiyan, A. Hamosh, A. Chakravarti, and A. Pandey, "Development of human protein reference database as an initial platform for approaching systems biology in humans.," *Genome Res*, vol. 13, pp. 2363–2371, Oct 2003.

[223] G. R. Mishra, M. Suresh, K. Kumaran, N. Kannabiran, S. Suresh, P. Bala, K. Shivakumar, N. Anuradha, R. Reddy, T. M. Raghavan, S. Menon, G. Hanumanthu, M. Gupta, S. Upendran, S. Gupta, M. Mahesh, B. Jacob, P. Mathew, P. Chatterjee, K. S. Arun, S. Sharma, K. N. Chandrika, N. Deshpande, K. Palvankar, R. Raghavnath, R. Krishnakanth, H. Karathia, B. Rekha, R. Nayak, G. Vishnupriya, H. G. M. Kumar, M. Nagini, G. S. S. Kumar, R. Jose, P. Deepthi, S. S. Mohan, T. K. B. Gandhi, H. C. Harsha, K. S. Deshpande, M. Sarker, T. S. K. Prasad, and A. Pandey, "Human protein reference database–2006 update.," *Nucleic Acids Res*, vol. 34, pp. D411–D414, Jan 2006.

[224] M. Hagedorn, L. Zilberberg, J. Wilting, X. Canron, G. Carrabba, C. Giussani, M. Pluderi, L. Bello, and A. Bikfalvi, "Domain swapping in a cooh-terminal fragment of platelet factor 4 generates potent angiogenesis inhibitors.," *Cancer Res*, vol. 62, pp. 6884–6890, Dec 2002.

[225] F. Z. H. S. e. a. Freije WA, Castro-Vargas FE, "freij-affy-human-91666," Mar 08 2006.

[226] W. A. Freije, F. E. Castro-Vargas, Z. Fang, S. Horvath, T. Cloughesy, L. M. Liau, P. S. Mischel, and S. F. Nelson, "Gene expression profiling of gliomas strongly predicts survival.," *Cancer Res*, vol. 64, pp. 6503–6510, Sep 2004.

[227] E. Poch, R. Miñambres, E. Mocholí, C. Ivorra, A. Pérez-Aragó, C. Guerri, I. Pérez-Roger, and R. M. Guasch, "Rhoe interferes with rb inactivation and regulates the proliferation and survival of the u87 human glioblastoma cell line.," *Exp Cell Res*, vol. 313, pp. 719–731, Feb 2007.

[228] P. Jia, S. Zheng, J. Long, W. Zheng, and Z. Zhao, "dmgwas: dense module searching for genome-wide association studies in protein-protein interaction networks.," *Bioinformatics*, vol. 27, pp. 95–102, Jan 2011.

[229] A. Neve, A. Corrado, and F. P. Cantatore, "Osteoblast physiology in normal and pathological conditions.," *Cell Tissue Res*, vol. 343, pp. 289–302, Feb 2011.

[230] N. C. Walsh, S. Reinwald, C. A. Manning, K. W. Condon, K. Iwata, D. B. Burr, and E. M. Gravallese, "Osteoblast function is compromised at sites of focal bone erosion in inflammatory arthritis.," *J Bone Miner Res*, vol. 24, pp. 1572–1585, Sep 2009.

[231] D. J. Shealy, P. H. Wooley, E. Emmell, A. Volk, A. Rosenberg, G. Treacy, C. L. Wagner, L. Mayton, D. E. Griswold, and X.-Y. R. Song, "Anti-tnf-alpha antibody allows healing of joint damage in polyarthritic transgenic mice.," *Arthritis Res*, vol. 4, no. 5, p. R7, 2002.

[232] A. R. Pettit, H. Ji, D. von Stechow, R. Müller, S. R. Goldring, Y. Choi, C. Benoist, and E. M. Gravallese, "Trance/rankl knockout mice are protected from bone erosion in a serum transfer model of arthritis.," *Am J Pathol*, vol. 159, pp. 1689–1699, Nov 2001.

[233] D. Granchi, G. Ochoa, E. Leonardi, V. Devescovi, S. R. Baglìo, L. Osaba, N. Baldini, and G. Ciapetti, "Gene expression patterns related to osteogenic differentiation of bone marrow-derived mesenchymal stem cells during ex vivo expansion.," *Tissue Eng Part C Methods*, vol. 16, pp. 511–524, Jun 2010.

[234] H. J. Yoo, S. S. Yoon, S. Y. Park, E. Y. Lee, E. B. Lee, J. H. Kim, and Y. W. Song, "Gene expression profile during chondrogenesis in human bone marrow derived mesenchymal stem cells using a cdna microarray.," *J Korean Med Sci*, vol. 26, pp. 851–858, Jul 2011.

[235] W. Wagner, S. Bork, P. Horn, D. Krunic, T. Walenda, A. Diehlmann, V. Benes, J. Blake, F.-X. Huber, V. Eckstein, P. Boukamp, and A. D. Ho, "Aging and replicative senescence have related effects on human stem and progenitor cells.," *PLoS One*, vol. 4, no. 6, p. e5846, 2009.

[236] B. S. Carvalho, "Preprocessing affymetrix exon st and gene st arrays," *Biocund*.

[237] H. E. Lockstone, "Exon array data analysis using affymetrix power tools and r statistical software," *Briefings in Bioinformatics*, pp. 1–2, 2010.

[238] J. Herrero, A. Valencia, and J. Dopazo, "A hierarchical unsupervised growing neural network for clustering gene expression patterns.," *Bioinformatics*, vol. 17, pp. 126–136, Feb 2001.

[239] P. Zoppoli, S. Morganella, and M. Ceccarelli, "Timedelay-aracne: Reverse engineering of gene networks from time-course data by an information theoretic approach.," *BMC Bioinformatics*, vol. 11, p. 154, 2010.

[240] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.," *Proc Natl Acad Sci U S A*, vol. 102, pp. 15545–15550, Oct 2005.

[241] M. GeneGo.

[242] J. Zhu, E. Shimizu, X. Zhang, N. C. Partridge, and L. Qin, "Egfr signaling suppresses osteoblast differentiation and inhibits expression of master osteoblastic transcription factors runx2 and osterix.," *J Cell Biochem*, vol. 112, pp. 1749–1760, Jul 2011.

[243] H. W. A. V. W. M. C. J. S. G. S. J. L. Jonathan Gordon, Mohammad Hassan, "Sirt2 is a protein deacetylase involved in the regulation of osteoblastogenesis by inhibiting adipogenesis," *ASBMR Annual Meeting. Session: Concurrent Oral Session 37: Osteoblasts: Insights on Osteoblast Gene Regulation.*, 2010.

[244] S. Peng, G. Zhou, K. D. K. Luk, K. M. C. Cheung, Z. Li, W. M. Lam, Z. Zhou, and W. W. Lu, "Strontium promotes osteogenic differentiation of mesenchymal stem cells through the ras/mapk signaling pathway.," *Cell Physiol Biochem*, vol. 23, no. 1-3, pp. 165–174, 2009.

[245] S. Zanotti, A. Smerdel-Ramoya, L. Stadmeyer, D. Durant, F. Radtke, and E. Canalis, "Notch inhibits osteoblast differentiation and causes osteopenia.," *Endocrinology*, vol. 149, pp. 3890–3899, Aug 2008.

[246] S. Zanotti and E. Canalis, "Notch and the skeleton.," *Mol Cell Biol*, vol. 30, pp. 886–896, Feb 2010.

[247] C. Higuchi, A. Myoui, N. Hashimoto, K. Kuriyama, K. Yoshioka, H. Yoshikawa, and K. Itoh, "Continuous inhibition of mapk signaling promotes the early osteoblastic differentiation and mineralization of the extracellular matrix.," *J Bone Miner Res*, vol. 17, pp. 1785–1794, Oct 2002.

[248] P. B. Gurpur, J. Liu, D. J. Burkin, and S. J. Kaufman, "Valproic acid activates the pi3k/akt/mtor pathway in muscle and ameliorates pathology in a mouse model of duchenne muscular dystrophy.," *Am J Pathol*, vol. 174, pp. 999–1008, Mar 2009.

[249] E. M. Horwitz, P. L. Gordon, W. K. K. Koo, J. C. Marx, M. D. Neel, R. Y. McNall, L. Muul, and T. Hofmann, "Isolated allogeneic bone marrow-derived mesenchymal cells engraft and stimulate growth in children with osteogenesis imperfecta: Implications for cell therapy of bone.," *Proc Natl Acad Sci U S A*, vol. 99, pp. 8932–8937, Jun 2002.

[250] A. Trounson, R. G. Thakar, G. Lomax, and D. Gibbons, "Clinical trials for stem cell therapies.," *BMC Med*, vol. 9, p. 52, 2011.

[251] G. V. Røsland, A. Svendsen, A. Torsvik, E. Sobala, E. McCormack, H. Immervoll, J. Mysliwietz, J.-C. Tonn, R. Goldbrunner, P. E. Lønning, R. Bjerkvig, and C. Schichor, "Long-term cultures of bone marrow-derived human mesenchymal stem cells frequently undergo spontaneous malignant transformation.," *Cancer Res*, vol. 69, pp. 5331–5339, Jul 2009.

[252] J. W. Kuhbier, B. Weyand, C. Radtke, P. M. Vogt, C. Kasper, and K. Reimers, "Isolation, characterization, differentiation, and application of adipose-derived stem cells.," *Adv Biochem Eng Biotechnol*, vol. 123, pp. 55–105, 2010.

[253] S. Kern, H. Eichler, J. Stoeve, H. Klüter, and K. Bieback, "Comparative analysis of mesenchymal stem cells from bone marrow, umbilical cord blood, or adipose tissue.," *Stem Cells*, vol. 24, pp. 1294–1301, May 2006.

[254] A. Shafiee, E. Seyedjafari, M. Soleimani, N. Ahmadbeigi, P. Dinarvand, and N. Ghaemi, "A comparison between osteogenic differentiation of human unrestricted somatic stem cells and mesenchymal stem cells from bone marrow and adipose tissue.," *Biotechnol Lett*, vol. 33, pp. 1257–1264, Jun 2011.

[255] Virtual physiological human initiative, http://www.vph-noe.eu/

[256] E. R. Mardis, "Next-generation dna sequencing methods.," *Annu Rev Genomics Hum Genet*, vol. 9, pp. 387–402, 2008.

[257] T. Oinn, M. Addis, J. Ferris, D. Marvin, M. Senger, M. Greenwood, T. Carver, K. Glover, M. R. Pocock, A. Wipat, and P. Li, "Taverna: a tool for the composition and enactment of bioinformatics workflows.," *Bioinformatics*, vol. 20, pp. 3045–3054, Nov 2004.