MOLECULAR AND COMPUTATIONAL APPROACHES TO UNDERSTANDING KELOID SCARRING

OOI NICK SERN, BRANDON
(B. Eng. (Hons.), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAMME IN BIOENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2010

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest gratitude to the Graduate Programme in Bioengineering for giving me the opportunity to pursue my PhD studies. By having access to the wonderful resources at the National University of Singapore, I have been greatly facilitated in my quest for knowledge and learning.

I would also like to thank my supervisors, Prof Phan Toan Thang and Prof Thiagarajan for their invaluable advice and enthusiastic support throughout my candidature. I have learnt a great deal from my interactions with them and I believe that this will stand me in good stead for my future endeavors. A big thank you also goes out to my Thesis Steering Committee comprising of Prof Bay Boon Huat and Dr Martin Lindsay Buist for their constructive advice and helpful suggestions. Their comments have helped shape this thesis in more ways than one.

In all areas of work, colleagues play an immense role in the learning and development of any project. Here I am indebted to Dr Anandaroop Mukhopadhyay, Dr Masilamani Jeyakumar, Ms Audrey Khoo, Mr Ong Chee Tian, Ms Zhou Yue and Mr Do Dang Vinh from the Wound Healing and Stem Cell Research Group, and to Dr Geoffrey Koh and Mr Liu Bing from the Computational Systems Biology Group. It is through them that I have learnt the *in vitro* and *in silico* techniques that were essential to my project. Their companionship has also been most welcome during the long years of my candidature. Special thanks also go to Dr Lim Cheh Peng from the Institute of Molecular and Cell Biology for checking my paper manuscripts, and also for the helpful support in microarray work and analysis.

Last but definitely not least, I would like to thank my family and countless friends who have supported and encouraged me throughout my PhD years. It is with their unwavering support and with God's grace that I now stand at the brink of completion of this project.

TABLE OF CONTENTS

SUMMARY

Keloid scars are aberrations in the wound healing process, resulting in the appearance of protrusive crab like extensions growing into normal tissue. They do not subside with time, and may develop over the most minor of skin wounds, such as insect bites or acne. Aside from being an aesthetic impediment, keloids are frequently associated with itchiness, pain and, when involving the skin overlying a joint, restricted range of motion. To date, none of the known treatment modalities have proven optimal.

In recent years, a systems approach to understanding biology has gained eminence, in part due to the limitations of a purely reductionist approach in explaining biological phenomena. However, there are merits to the reductionist approach; much of what we know of biology today can be attributed to the work of molecular biologists of the past. In this dissertation, we will adopt both these approaches to tackling the keloid problem.

In the first part of this thesis, we examined the role played by a novel growth factor, the hepatoma-derived growth factor (HDGF), in keloid pathogenesis. Using a combination of immunohistochemical staining and Western blots, we found that secreted HDGF is increased in the keloid condition and its secretion is modulated by epithelial–mesenchymal interactions. Furthermore, exogenous HDGF exerts a proliferative effect on keloid fibroblasts and increases the production of the angiogenic factor VEGF, indicating that it plays some role in the process of angiogenesis.

With the advent of high throughput technology, researchers are no longer confined to the study of individual molecules. In the second part of this dissertation, we utilized the microarray platform to assess the global transcriptional differences between

keloid and normal fibroblasts under serum free conditions. Many of the genes that have been found to be differentially expressed in previous studies were reconfirmed in this study. In addition, some interesting and novel genes not previously reported were also discovered. Gene Ontology terms that were found to be significantly enriched include those relating to immune response, antigen processing and presentation, chemokine and cytokine activity, extracellular matrix and ribosomal proteins.

In the third part of this thesis, we attempted to reverse engineer gene networks from microarray expression profiles of keloid and normal fibroblasts. Using a physical approach to model transcription factor interactions, we discovered some of the binding motifs that were active in the keloid condition. Furthermore, we used the influence approach to reverse engineer some of the networks that were found to be significantly enriched from the second part of this dissertation. Our results indicate that transcriptional networks were better suited for this process compared to cytokine receptor interactions and intracellular signaling networks. We also found that the NFKB transcriptional network that was inferred from normal fibroblast data was more accurate compared to that inferred from keloid data, suggesting a more robust network in the keloid condition. This would mean that targeting NFKB alone may not be sufficient to reduce its transcriptional products in keloid fibroblasts.

The work done in this thesis, utilizing both molecular and computational approaches, has advanced our understanding by shedding light on some of the important players and key networks in keloid scarring. In addition, the results from this study has generated new and promising future areas of research, and is a small step forward to finding a solution to this condition.

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

α-SMA	Alpha-smooth muscle actin
ADAM12	A disintegrin and metalloprotease 12
ANOVA	Analysis of variance
AVEN	Apoptosis caspase activation inhibitor
ATXN1	Ataxin-1
BANJO	Bayesian Network Inference with Java Objects
BDe	Bayesian metric with Dirichlet priors and equivalence
C3	Complement component 3
CADM1	Cell adhesion molecule 1
CDF	Chip definition file
cDNA	Complementary DNA
CFB	Complement factor B
COL1A1	Collagen type I alpha 1
COL5A1	Collagen type V alpha 1
COL5A3	Collagen type V alpha 3
COL15A1	Collagen type XV alpha 1
COL17A1	Collagen type XVII alpha 1
cRNA	Complementary RNA
CTGF	Connective tissue growth factor
CXCL1	Chemokine ligand 1
CXCL2	Chemokine ligand 2
CXCL6	Chemokine ligand 6

DAVID	Database for Annotation, Visualization and Integrated Discovery
DNA	Deoxyribonucleic acid
DPI	Data Processing Inequality
ECM	Extracellular matrix
ECM1	Extracellular matrix protein 1
EPHB4	Ephrin type-B receptor 4
fREDUCE	fast-Regulatory Element Detection Using Correlation with Expression
GEO	Gene Expression Omnibus
G0S2	G0/G1switch 2
GO	Gene Ontology
GPR137B	G-protein-coupled receptor 137B
GPR153	G-protein-coupled receptor 153
GPSM2	G-protein signaling modulator 2
GRB10	Growth factor receptor-bound protein 10
HATH	Homologous to amino terminus of HDGF
HDGF	Hepatoma-derived growth factor
HIF-1α	Hypoxia induced factor-1α
HLA	Human leukocyte antigen
HOXA11	Homeobox A11
HOXD10	Homeobox D10
HRP	HDGF-related protein

HSD11B1	Hydroxysteroid (11-beta) dehydrogenase 1
IFIT1	Interferon-induced protein with tetratricopeptide repeats 1
IFIT3	interferon-induced protein with tetratricopeptide repeats 3
IGF-1	Insulin-like growth factor-1
IGFBP	Insulin-like growth factor binding protein
IGFBP3	Insulin-like growth factor binding protein 3
IHC	Immunohistochemical staining
IL6	Interleukin 6
IL8	Interleukin 8
IL32	Interleukin 32
IVT	*In vitro* transcription
KEGG	Kyoto Encyclopedia of Genes and Genomes
KF	Keloid fibroblasts
KK	Keloid keratinocytes
KRT19	Keratin 19
LAMA2	Laminin alpha 2
LEDGF	Lens epithelium-derived growth factor
MEMO1	Mediator of cell motility 1
MAPK	Mitogen-activated protein kinase
MAS	Microarray suite
MHC	Major histocompatibility complex
MI	Mutual Information
MMP	Matrix metalloproteinase

mRNA	Messenger RNA
mTOR	Mammalian target of rapamycin
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MYO1D	Myosin 1D
MYO19	Myosin 19
NF	Normal fibroblasts
NFKB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGF	Nerve growth factor
NK	Normal keratinocytes
OAS1	2',5'-oligoadenylate synthetase 1
PAI-1	Plasminogen activator inhibitor-1
PCNA	Proliferating cell nuclear antigen
PCA	Principal components analysis
PCR	Polymerase chain reaction
PDGF	Platelet-derived growth factor
PDGFRB	Platelet-derived growth factor receptor beta
PI3-K	Phosphatidylinositol 3-kinase
POSTN	Periostin
PPV	Positive Predicted Value
PTGES	Prostaglandin E synthase
PTX3	Pentraxin-related gene

RAC2	Ras-related C3 botulinum toxin substrate 2
REDUCE	Regulatory Element Detection Using Correlation with Expression
RMA	Robust multichip analysis
RNA	Ribonucleic acid
RNAP	RNA polymerase
RPS	Ribosomal protein
RSAD2	Radical S-adenosyl methionine domain containing 2
SEM	Standard error of the mean
SEM5A	Semophorin-5A
SFRP1	Secreted frizzled-related protein 1
SLC39A8	Solute carrier family 39 member 8
SOS2	Son of sevenless homolog 2
TAP	Transporter associated with antigen processing
TF	Transcription factor
THBS1	Thrombospondin-1
TNFAIP3	Tumor necrosis factor alpha-induced protein 3
TNFAIP6	Tumor necrosis factor alpha-induced protein 6
TNFSF10	Tumor necrosis factor superfamily member 10
TGF-β	Transforming growth factor beta
VEGF	Vascular endothelial growth factor
WNT5A	Wingless-type MMTV integration site family member 5A

LIST OF PRESENTATIONS AND PUBLICATIONS

Biostar 2006: 2nd International Congress on Regenerative Biology, Stuttgart. Hepatoma-derived growth factor contributes to keloid pathogenesis via epithelial mesenchymal interactions and secretion into conditioned media [Poster presentation]

J Cell Mol Med. 2010 Jun; 14(6A):1328-37. Hepatoma-derived growth factor and its role in keloid pathogenesis. Ooi BN, Mukhopadhyay A, Masilamani J, Do DV, Lim CP, Cao XM, Lim IJ, Mao L, Ren HN, Nakamura H, Phan TT. [Published]

Burns. Microarray analysis of serum starved keloid and normal fibroblasts suggest a role for immunological factors and ribosomal proteins in keloid pathogenesis. BNS Ooi, CP Lim, XM Cao, TT Phan [Submitted]

Theor Biol Med Model. 2011 May 2; 8:13. Insights gained from the reverse engineering of gene networks in keloid fibroblasts. Ooi BN, Phan TT. [Published]

CHAPTER ONE

INTRODUCTION

"The availability of genome sequence is just the beginning. Scientists now want to understand the genes and the role they play in the prevention, diagnosis and treatment of disease."

– Dr Randy Scott, President of Incyte

1.1 Background and motivations for the thesis

Since the discovery of deoxyribonucleic acid (DNA) in the 1950s by Watson and Crick, biology has moved at a rapid rate. Thanks to the Human Genome Project, we now have in our possession the complete genome of the human species. Future biomedical research would involve the application of this knowledge to the understanding of various biological processes in the hopes of uncovering new methods of treating the numerous diseases and medical conditions afflicting the human race.

Among the many diseases to beset mankind, keloids do not rank very highly in the hall of fame. However, the appearance of these large protruding claw-like scars is bound to elicit shock and distress in most observers due to their unsightly nature. Furthermore, aside from causing emotional trauma, keloid scars can be painful or itchy, and may restrict mobility if formed over a joint (Lee et al. 2004). In a study assessing the quality of life of patients with keloid and hypertrophic scarring, it was demonstrated for the first time that the quality of life of these patients was reduced due to physical and/or psychological effect (Bock et al. 2006). The problem is further exacerbated by the fact

that there is no particularly effective treatment to date (Tuan & Nichter 1998; Louw 2007). These scars also have a propensity to recur after surgery and have been considered as benign tumours (English & Shenefelt 1999).

For all the reasons stated above, it would be beneficial if we could discover some effective method of treating these scars. To this end, an understanding of the molecular etiology of keloids would be useful. Furthermore, since keloid formation is generally considered to be a form of abnormal wound healing, any insights gained from this endeavour would also increase our understanding of the wound healing process.

1.2 Approach and methodology

We have decided to adopt both top down as well as reductionist approaches to understanding the mechanisms underlying keloid pathology. In the first part of this dissertation, we investigated the role played by a novel protein in the keloid condition using molecular biology techniques. While it was found that this molecule, the hepatoma-derived growth factor (HDGF) is significantly expressed in keloids, our data also suggests that it is unlikely that this growth factor is able to induce keloid formation on its own. Therefore, while a reductionist, in depth study of this molecule would certainly increase our understanding about keloids, the knowledge gained would only be a small fraction of the complex mechanisms underlying keloid pathology.

Researchers today are no longer confined to studying one molecule at a time thanks to the development of various high throughput techniques. These technologies enable us to have a snapshot of the thousands of molecules present in the cell at any one

time. In the second part of this dissertation, we utilized one of these technologies, specifically the Affymetrix microarray platform, to gain insights into some of the system level differences between keloid and normal cells. Based on this approach, we would be able to identify all genes that are significantly different between the two conditions. The data generated from this study can then be utilized for further research by using a reductionist approach to study the genes individually, or by extracting biological meaning through a computational approach.

One way of increasing our biological knowledge is to learn how the different molecules in the cell are connected. In the third part of the dissertation, we attempt to reconstruct gene networks using a combination of probabilistic and regression techniques. There are two general strategies for reverse engineering gene networks – a physical approach where physical interactions between transcription factors and their promoters are modeled, and an influence approach where the mechanistic process is abstracted out as a black box. For the physical approach, we will use the entire microarray data set for modeling, but for the influence approach, we will focus on small networks of genes that have been found to be differentially expressed from the second part of this dissertation. Most attempts at modeling biological networks have been done using simulated data; our work would highlight some of the issues involved in working with experimental data. Furthermore, it is hoped that insights gained from this endeavor would provide some clues about the different transcriptional regulatory mechanisms present in keloid and normal fibroblasts.

1.3 Contributions of the thesis

We first discovered increased expression of HDGF in keloid scars compared to normal skin. An in vitro study of the role of HDGF using keloid and normal derived cells suggest that epidermal mesenchymal interactions govern the increased secretion of this growth factor in the keloid condition. Furthermore, HDGF was found to increase the proliferation of keloid fibroblasts and was also found to increase the production of the vascular endothelial growth factor (VEGF). However, one of the hallmarks of keloids is an increased extracellular matrix production, and HDGF did not seem to contribute to this aspect of keloid formation.

In the second part of this dissertation, we used the microarray platform in an attempt to identify groups of genes that can be implicated in the formation of keloids. While other groups have utilized this technology previously, none had surveyed the global transcriptional landscape in serum starvation conditions. Furthermore, there was very little overlap in many of the microarray studies done, and it is hoped that our study would help identify some of the more consistent differentially expressed genes. Our results indicate some consistency with previous studies done on keloid fibroblasts. We also uncovered differentially expressed genes that have not been reported previously, and enrichment analysis indicate that processes such as immune response, antigen processing and presentation, chemokine and cytokine activity, extracellular matrix and ribosomal proteins are among those that are affected in the keloid condition.

In the third part of this dissertation, we attempted to reverse engineer gene networks using the microarray data that was generated in the second part of the thesis, as well as any publicly available microarray data on keloid and normal fibroblasts that we

could find in the literature. Using the physical approach of correlating expression values to binding motifs, we found some consensus sequences that were active in the keloid condition, as well as some sequences that were responsive to steroids, one of the commonly used treatments for keloids. These consensus sequences are possible transcription factor binding sites and could be explored for developing future keloid treatments or to improve the efficacy of current steroid treatments. We also compared different normalization methods and influence approaches on the reconstruction of known gene networks taken from the KEGG database that were found to be statistically enriched in our microarray data. We found that the combination of the Bayesian algorithm, RMA normalization and transcriptional networks gave the best reconstruction results and this could serve as a guide for future influence approaches dealing with experimental data.

Fig. 1.1 summarizes the three main approaches taken in this thesis. An in depth single factor study of HDGF was first undertaken based on previous results from our lab. The limitations of this approach however was evident from the results of this study, motivating a more global approach to understanding keloid scarring through the use of microarray technology. Results from the microarray experiments were then analyzed computationally to provide further insights into the molecular mechanisms underpinning this condition.

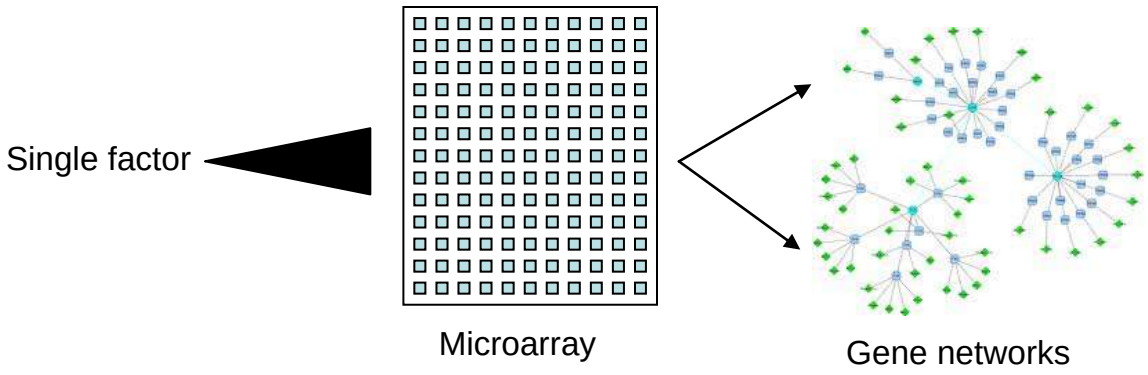


Figure 1.1: Summary of the three main approaches used in this study. An in depth analysis of the role of HDGF was first conducted. This was followed by a global overview of transcriptional differences between keloid and normal fibroblasts using the microarray platform. The microarray data was then reverse engineered to identify some of the important players and key networks in keloid formation.

1.4 Organization of the thesis

The rest of the thesis is organized as follows. In Chapter Two, background information on wound healing and keloid scarring is presented. Chapter Three describes the materials and methods used in both molecular and computational approaches employed in this study. In Chapter Four, the importance of HDGF in keloid formation is studied using a combination of cell and molecular techniques. Chapter Five examines the global transcriptional differences between keloid and normal skin fibroblasts by utilizing the Affymetrix microarray platform. In Chapter Six, insights obtained from the reverse engineering of keloid and normal fibroblast gene networks are discussed. Conclusions from the thesis are presented in Chapter Seven.

CHAPTER TWO

LITERATURE REVIEW

2.1 Wound healing

To understand the underlying mechanisms involved in pathologic conditions such as scarring and fibrosis, it is useful to first review what is known about normal tissue response to injury. Upon wounding, an orderly series of events is triggered, with the final desired outcome being the restoration of anatomical structure and function. These events can be grouped into four distinct but overlapping phases, hemostasis, inflammation, proliferation and remodeling (Mast 1992).

2.1.1 Hemostasis and inflammation

The healing cascade starts with the aggregation of platelets at the wound site and the release of clotting factors. This results in the formation of a fibrin clot to plug the wound (Clark 2001). At the same time, a cocktail of growth factors and cytokines are released from the serum of the disrupted blood vessels and degranulating platelets (Werner & Grose 2003). Following hemostasis, neutrophils infiltrate into the wound site and monocytes are activated to become wound macrophages. These inflammatory cells serve two purposes: firstly as a means of removing foreign material, bacteria and damaged matrix components by phagocytosis, and secondly as a source of growth factors that are required to initiate the next phase of the healing process (Sylvia 2003; Diegelmann, Cohen & Kaplan 1981).

2.1.2 Proliferation

In the proliferative phase, the predominant cell in the wound site is the dermal fibroblast (Stadelmann, Digenis & Tobin 1998). This cell is responsible for producing collagen and other extracellular matrix components needed to restore structure and function to the injured tissue. At least 23 different types of collagen have been identified but type I is predominant in the scar tissue of skin (Prockop & Kivirikko 1995). Also during this phase, keratinocytes in the epidermis proliferate and migrate from the wound edge leading to the process of reepithelialization (Santoro & Gaudino 2005). In addition, local factors in the wound microenvironment such as low pH and reduced oxygen tension initiate the release of angiogenic factors leading to the migration and proliferation of endothelial cells (Knighton et al. 1983). Massive angiogenesis leads to the formation of new blood vessels, and the resulting wound connective tissue is known as granulation tissue because of the granular appearance of the numerous capillaries (Werner & Grose 2003). Around a week after the wounding has taken place, fibroblasts have differentiated into myofibroblasts and the wound begins to contract. Myofibroblasts contain the same kind of actin as found in smooth muscle cells, alpha-smooth muscle actin (α-SMA) to produce more force during contracture (Hinz 2006).

2.1.3 Remodeling

In the final stage, collagen undergoes cross-linking to improve its strength and stability. This stage is characterized by continued collagen synthesis and collagen catabolism finally resulting in a normal scar (Parks 1999). This process requires a balance between matrix biosynthesis and matrix degradation. A disruption in this balance either due to

excessive matrix deposition or decreased matrix degradation leads to keloid and hypertrophic scars (Raghow 1994).

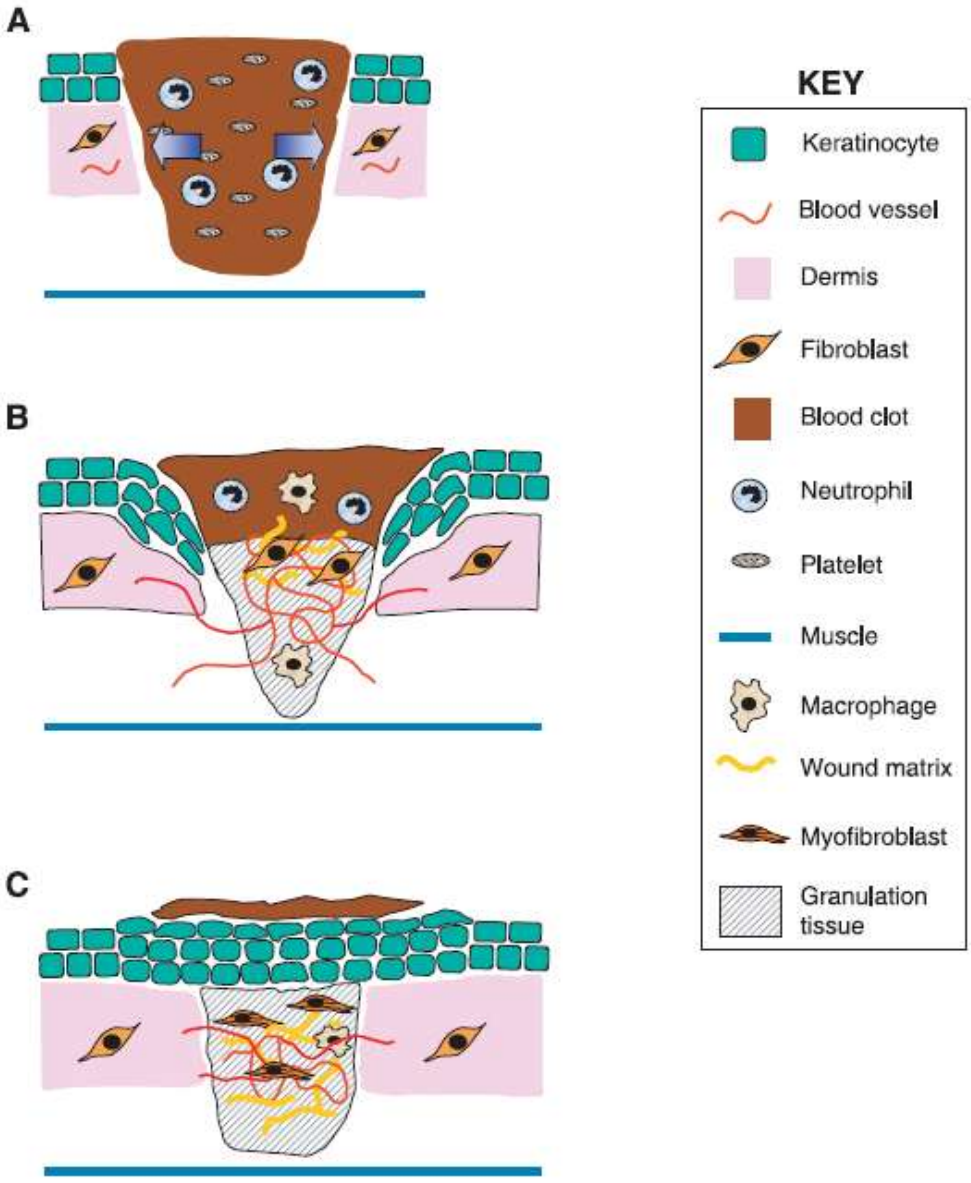


Figure 2.1: Schematic representation of different stages of wound repair (Werner & Grose 2003). *A*: 12–24 h after injury the wounded area is filled with a blood clot. Neutrophils invade into the clot. *B*: at days 3–7 after injury, macrophages are abundant in the wound tissue. Endothelial cells migrate into the clot; they proliferate and form new blood vessels. Fibroblasts migrate into the wound tissue, where they proliferate and deposit extracellular matrix. Keratinocytes proliferate at the wound edge and migrate above the provisional matrix. *C*: 1–2 wk after injury the wound is completely filled with granulation tissue. The wound is completely covered with a neoepidermis.

2.2 Keloid scarring

2.2.1 Keloid versus hypertrophic scars

The term cheloide was coined in 1802 to describe the lateral extensions often observed in these scars, which resemble the legs of a crab growing into normal tissue (Urioste, Arndt & Dover 1999). Keloids are commonly compared with hypertrophic scars, and the two share some similarities such as increased collagen secretion and a similar gross appearance. However, unlike hypertrophic scars that are confined to the area of injury, keloids may extend well beyond the confines of the original wound. Furthermore, hypertrophic scars usually subside with time, whereas keloids continue to evolve over time, without a quiescent or regressive phase (Nemeth 1993). While hypertrophic scars usually develop within a few weeks after skin injury, keloids normally show a delayed onset, normally forming months after skin trauma (Marneros & Krieg 2004).

2.2.2 Epidemiology

It is not well documented how commonly keloids occur in the general population but the reported incidence range from a high of 16% among adults in Zaire to a low of less than 1% among adults in England (English & Shenefelt 1999). It is widely accepted that dark-skinned populations have a higher occurrence of keloids than light-skinned populations, with the reported incidence ratio between the two groups ranging from 2:1 to 19:1 (Atiyeh, Costagliola & Hayek 2005). Among Asians, keloid incidence appears to be more common in Chinese (Shaffer, Taylor & Cook-Bolden 2002). Both autosomal dominant

and autosomal recessive genetic inheritance have been proposed but not confirmed and some data suggest familial occurrence (Bloom 1956; Omo-Dare 1975).

A difference in occurrence of keloids based on gender has not been demonstrated convincingly (Marneros & Krieg 2004). However, most reported cases have occurred in individuals between 10 and 30 years of age (Rockwell, Cohen & Ehrlich 1989). Hormone levels are high at this age, indicating that they may have some influence on keloid formation. This hypothesis is supported by data showing an increased binding of androgens in keloid tissue (Ford et al. 1983; Schierle, Scholz & Lemperle 1997). Furthermore, some reports suggest that keloids appear more often in puberty, enlarge during pregnancy, and decrease in size after menopause (Moustafa, Abdel-Fattah & Abdel-Fattah 1975). However, other explanations such as increased neo-angiogenesis during pregnancy are also possible (Seifert & Mrowietz 2009).

2.2.3 Clinical presentation

Keloids are generally considered to be a result of excessive wound healing, although some also believe these scars to be a type of benign fibrous tumor (Slemp & Kirschner 2006). They are characterized by an overgrowth of dense fibrous tissue coupled with excessive deposition of extracellular matrix (ECM) components such as collagen and fibronectin (Rockwell, Cohen & Ehrlich 1989; Babu, Diegelmann & Oliver 1989). They uniquely affect only humans, and may develop even after the most minor of skin wounds, such as insect bites or acne (Urioste, Arndt & Dover 1999). Keloids are frequently associated with itchiness, pain and, when involving the skin overlying a joint, restricted range of motion (Lee et al. 2004). For unknown reasons, keloids occur more frequently

on the chest, shoulders, upper back, back of the neck, and earlobes (Bayat et al. 2004). Corneal keloidal scarring has also been observed (Shukla, Arora & Arora 1975).

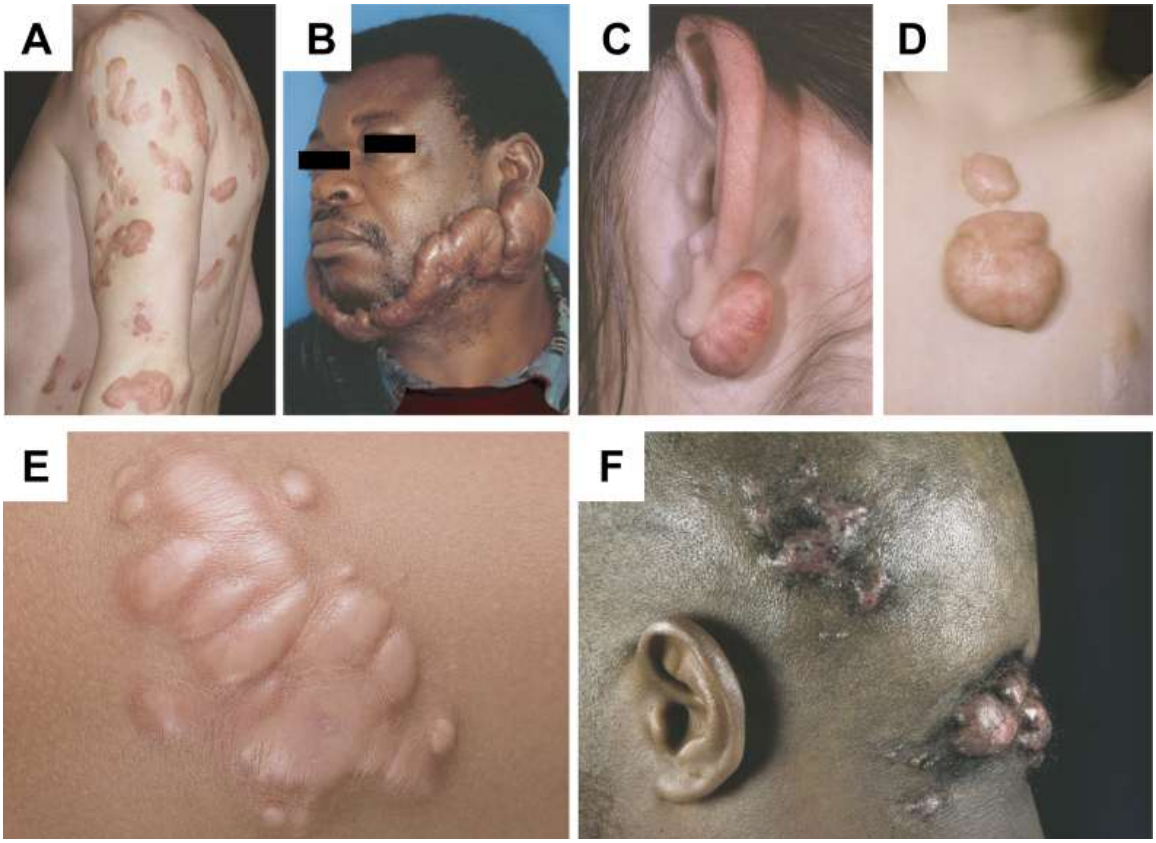


Figure 2.2: Keloid formation in different parts of the body and in different patients (Marneros & Krieg 2004).

2.2.4 Histopathology

Normal skin contains distinct collagen bundles that run parallel to the epithelial surface. In hypertrophic scars, collagen bundles are flatter, less demarcated, and are arranged in a wavy pattern. In keloids, the collagen bundles are thick and are randomly oriented as swirls and whorls (Rockwell, Cohen & Ehrlich 1989). Keloid formations are characterized by active angiogenesis and hypoxia (Appleton, Brown & Willoughby 1996). Occlusion of some microvessels by excessive endothelial cells may lead to local hypoxic conditions and apoptosis (Kischer 1992).

2.2.5 Etiology

Several etiological factors for keloids have been proposed, with skin injury being the most obvious. Spontaneous occurrence of keloids in the absence of trauma is rare although a few cases have been reported (Shaffer, Taylor & Cook-Bolden 2002). However, such spontaneous occurrence could be the result of a minor, overlooked trauma to the skin (Marneros & Krieg 2004). Increased skin tension has also been postulated to play some role in keloid formation. However, soles and palms which are sites of high skin tension are rarely sites of keloid formation, and the most affected site reported, the earlobe, is under minimal tension (Seifert & Mrowietz 2009).

The role of immunologic factors in keloid formation has not been studied in detail and remains to be elucidated. Immune cell infiltrate in keloids include T lymphocytes and denditric cells (Santucci et al. 2001) and an increased number of macrophages, epidermal Langerhans cells and mast cells have been noted as well (Niessen et al. 2004; Smith, Smith & Finn 1987). Some authors have reported an association with cell membrane proteins, such as HLA-DRB-16, B-14, and BMW-16 (Datubo-Brown 1990), elevated tissue levels of IgG, IgA, and IgM (Kischer et al. 1983), and abnormal immune response to sebum (Yagi, Dafalla & Osman 1979). The sebum hypothesis provides an explanation for the absence of keloids on anatomical sites lacking sebaceous glands, such as palms and soles (Seifert & Mrowietz 2009). Dermal injury exposes the pilosebaceous unit to the systemic circulation, initiating a cell-mediated immune response in persons who retain T lymphocytes sensitive to sebum. Subsequent release of cytokines, including various interleukins and TGF-beta, stimulates chemotaxis of mast cells and production of collagen by fibroblasts. This hypothesis also gives a plausible reason as to why only

human beings, the only mammals with true sebaceous glands, are affected by keloid scarring (Al-Attar et al. 2006).

Several studies have shown that many different cytokines and growth factors are involved in the formation of keloids. Some of the important molecules that were elevated in keloids include transforming growth factor beta (TGF-β) (Lee et al. 1999), interleukin-6 (IL-6) (Ghazizadeh 2007) and vascular endothelial growth factor (VEGF) (Ong et al. 2007). Keloid fibroblasts were also more responsive in mitogenic assays to platelet-derived growth factor (PDGF) (Haisa, Okochi & Grotendorst 1994).

Another possible factor underlying the growth and formation of keloids is their resistance to apoptosis. Keloid fibroblasts was found to be more resistant to Fas mediated apoptosis (Chodon et al. 2000) and the overexpression of insulin-like growth factor-1 (IGF-1) receptor inhibited ceramid-induced apoptosis (Ishihara et al. 2000). Furthermore, decreased expression of proapoptotic genes (Sayah et al. 1999) and increased expression of apoptotic inhibitors (Messadi et al. 2004) have also been observed in keloid fibroblasts.

Tissue hypoxia could be another contributory factor to pathogenesis. An increased level of hypoxia marker, hypoxia induced factor-1α (HIF-1α) was detected in keloid tissues and hypoxia appears to elevate the expression of plasminogen activator inhibitor-1 (PAI-1) (Zhang et al. 2003). Increased PAI-1 activity correlated with an elevated collagen expression in fibrin gel cultures of keloid fibroblasts (Tuan et al. 2003). Hypoxia-driven VEGF is also increased in keloids (Wu et al. 2004).

While most *in vitro* studies focus on keloid fibroblasts, recent evidence points to altered interactions between keratinocytes and fibroblasts in keloids. To examine

epithelial-mesenchymal cross-talk in skin, experiments using normal or keloid keratinocytes co-cultured with normal or keloid fibroblasts have been conducted. In such co-culture systems, keloid keratinocytes promoted the proliferation of keloid fibroblasts to a greater extent than normal keratinocytes, while the least proliferation was seen in keloid fibroblasts cultured without any keratinoctyes (Lim et al. 2001; Funayama et al. 2003). Furthermore, co-culturing normal or keloid fibroblasts with keloid keratinocytes resulted in an increased expression of collagen I and III compared to the non co-cultured condition (Lim et al. 2002). These data suggest that epithelial-mesenchymal interactions could contribute to keloid pathogenesis.

2.2.6 Treatment

Like many other diseases, the best treatment for keloids is prevention. Although many different treatment modalities have been proposed, none have proven to be optimal. Surgical excision of a keloid is associated with a high recurrence rate and therefore has to be combined with some other adjunctive therapy. These include compression therapy, silicone sheeting, cryotherapy, radiation or laser therapy (Slemp & Kirschner 2006; Louw 2007).

Unfortunately, there are drawbacks to many of these methods. Compression therapy is ultimately limited by the ability to adequately fit the garment to the wounded area and patient discomfort frequently reduces compliance (Cheng et al. 1984). The success of silicone sheeting is also limited by patient compliance, and silicone products may cause adverse effects, including skin maceration and excoriation (Slemp & Kirschner 2006). Cryotherapy could lead to permanent hypopigmentation resulting from

cold sensitivity of melanocytes and is therefore less desirable in patients with darker skin (Louw 2007). On the other hand, radiation therapy causes hyperpigmentation and carries the theoretical risk of radiation induced malignancy (Wolfram et al. 2009). The efficacy of laser treatment has been low with a recurrence rate of 50% (Apfelberg et al. 1989).

Other pharmacologic therapies for reducing the recurrence rate exist, with the application of corticosteroids being the most well known. Potential side effects of corticosteroid injections include pain, skin atrophy, telangiectasia formation, depigmentation, and infection (Urioste, Arndt & Dover 1999). Treatment with interferons, which are cytokines secreted by T-helper cells, may help to reduce fibrosis, but treatment has also been met with some success, but has severe side effects including fever, chills, night sweats, fatigue, myalgia and headache (Wolfram et al. 2009). 5-Fluorouracil is another compound that has been used successfully as an antiproliferative agent. The injection can be painful however, and purpura and ulcers have been documented (Wolfram et al. 2009).

The side effects of the above treatments notwithstanding, ultimately, none of the above methods are completely effective in preventing the recurrence of keloids. Many attempts have been made to find successful alternatives, with the ultimate direction of research geared toward understanding scarring at the molecular level in the hope of obtaining a permanent solution to this problem.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Media and chemicals

Dulbecco's modified eagle medium (DMEM), Hanks balanced salt solution (HBSS), fetal calf serum (FCS), streptomycin, penicillin, gentamicin and fungizone were purchased from Gibco. Keratinocyte growth medium (KGM) was purchased from Clonetics (USA). Phosphate buffered saline without Ca^{2+} and Mg^{2+} (PBS), epidermal growth factor (EGF), cholera toxin and hydrocortisone were purchased from Sigma Chemical Co (USA). Dispase II was purchased from Boehringer Mannheim (USA). Rhodamine counter stain was obtained from Difco (USA). Tris base was purchased from J.T Baker. Triton X-100, ethylenediaminetetraacetic acid (EDTA), 30% acrylamide/bis solution (37.5:1 2.6%C) and glycine were purchased from Biorad. Sodium Chloride (NaCl), nonidet P-40 (NP-40), sodium dodecyl sulphate, hydrogen peroxide (H_2O_2), bovine serum albumin (BSA), tween-20, potassium chloride (KCL), potassium phosphate (K_3PO_4), magnesium chloride ($MgCl_2$), MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide], N,N – dimethylformamide (DMF) and paraformaldehyde were all purchased from Sigma Chemical Co (USA). Methanol and acetic acid were purchased from Lab-Scan. RNeasy kit was bought from Qiagen (Germany) while the GeneChip Eukaryotic Target Labeling and Control Reagents and arrays were bought from Affymetrix (USA).

3.2 Cell isolation

3.2.1 Keloid keratinocyte and fibroblast database

Keratinocytes and fibroblasts were randomly selected from a specimen bank of keratinocyte/fibroblast strains derived from excised keloid specimens. All patients had received no previous treatment for the keloids before surgical excision. A full history was taken and an examination was performed, complete with coloured slide photographic documentation, before taking informed consent prior to excision. Approval by the National University of Singapore (NUS) Institutional Review Board (NUS-IRB) was sought before excision of human tissue and collection of cells.

3.2.2 Keratinocyte culture from keloid scar and normal skin

Excised keloid scar and normal skin specimens were repetitively washed in PBS containing 150 μg/ml gentamicin and 7.5 μg fungizone, until the washing solution became clear. The tissue was then divided into pieces of approximately 5mm × 10mm and the epidermis was scored. Dispase 5mg/ml in HBSS was added and skin was incubated overnight at 4ºC. The epidermis was carefully scraped off with a scalpel the next day and placed in trypsin 0.25%/Glucose 0.1%/EDTA 0.02% for 10 min in the incubator. Trypsin action was quenched by DMEM/10% FCS. The suspended cells were transferred into tubes and centrifuged at 1000 rpm for 8 min. The cells were seeded in Keratinocyte Culture Medium (80 ml DMEM supplemented with 20 ml FCS, EGF 10 ng/ml, cholera toxin 1×10^{-9} M and hydrocortisone 0.4 μg/ml) at 1×10^{5} cells/cm^2 for 24 hrs before changing to Keratinocyte Growth Medium (KGM). The cell strains were

maintained and stored at -150ºC. Only cells from second and third passages were used for the experiments.

3.2.3 Fibroblast culture from keloid scar and normal skin

Remnant dermis from the keloid scar and normal skin were either minced or incubated in a solution of collagenase type 1 (0.5 mg/ml) and trypsin (0.2 mg/ml) at 37ºC for 6 hrs. Cells were pelleted and grown in tissue culture flasks. Alternatively the skin tissue samples were chopped into pieces of 1-2 mm^2. The pieces were then transferred to a 100mm tissue culture dish previously coated with a thin layer of DMEM/10%FCS. Culture medium enough to cover the explants were then added and topped up after 2-3 days. After 4-7 days the fibroblasts outgrew from the tissue. Fibroblast cell strains were maintained and stored at -150ºC until use. Only cells from the second and third passages were used for the experiments.

3.2.4 Cell counting

Before cells were seeded into culture flasks for experiments, aliquots of the cell suspension were mixed with trypan blue in a ratio of 1:4 and counted in a Neubauer's haemocytometer. Non-viable cells will be stained blue while viable cells remain opaque. Viable cells in the four corner squares of the haemocytometer were counted. Since the volume of each square is 10^{-4} cm^3 the following formula can be used to calculate the number of cells in the cell suspension:

Cells per ml = the average count per square x 4 (dilution factor) x 10^4

Total cell number = cells per ml x the original volume of fluid from which cell sample was removed

3.3 HDGF experiments

3.3.1 Immunohistochemistry

Paraffin sections of keloid and normal tissue were dewaxed or deparaffinized in two changes of xylene followed by rehydration in 100%, 95% and 70% ethanol gradient. Antigens were then retrieved by immersing the slides in 0.01 M citrate buffer, pH 6.0, heating in a microwave oven (high for 2.5 min., low for 5 min.), cooling at 4°C for 20 min. and washing in water for 5 min. Endogenous peroxidase was blocked in 3% H_2O_2 and non-specific binding was blocked for 1 hr (CAS block; Zymed Laboratories, South San Francisco, CA, USA). The sections were incubated with antibodies specific for HDGF, diluted 1:1000 for 1 hr. After washing, the slides were incubated in anti-mouse IgG-peroxidase (Zymed) or anti-rabbit IgG-peroxidase (Zymed), diluted 1:500 for 2 hrs, for HDGF primary antibodies, respectively. The slides were washed in Tris-buffered NaCl (TBS) or 0.05% Tween-20, pH 7.5, and then with MilliQ H_2O (Millipore Corp, Billerica, MA, USA). The reaction product was developed with 3,3'-diaminobenzidine tetrahydrochloride substrate kit (Zymed), and the sections were counterstained with hematoxylin. All wash steps were carried out in TBS/0.05% Tween-20. The antibodies were diluted in 1% bovine serum albumin (BSA)/TBS. Non-immune mouse/rabbit antibody of the appropriate immunoglobulin isotype was used for negative controls.

3.3.2 Serum stimulation of fibroblasts

Fibroblasts were seeded in six-well plates at a density of 1 x 10^4 cells/ml in 10% FCS for 24 hrs and subsequently starved in a serum-free medium for another 48 hrs. After 48 hrs, the fibroblasts were stimulated by exposure to either 10% FCS or DMEM for 5 days before being harvested.

3.3.3 Keratinocyte-fibroblast co-culture

Keloid keratinocytes (KK) and normal keratinocytes (NK) obtained from randomly selected keloid and normal strains were seeded at a density of 1 x 10^5 cells/cm^2 on Transwell clear polyester membrane inserts with 0.4-μm pore size and 4.5 cm^2 (Corning Incorporated Life Sciences, Acton, MA, USA) area. The cells were maintained for 4 days in EpiLife medium (Cascade Biologics, OR, USA) until 100% confluent. The medium was then changed to EpiLife supplemented with increased calcium concentration and the cells were exposed to the air–liquid interface for another 3 days, allowing the keratinocytes to stratify and reach terminal differentiation. Keloid fibroblasts (KF) and normal fibroblasts (NF) were seeded in six well plates at a density of 1 x 10^5 cells/well in DMEM/10% FCS for 48 hrs to 80% confluency. Keratinocytes on membrane inserts and plated fibroblasts were washed twice with PBS before the inserts were placed into the six-well plates containing fibroblast cultures to initiate KK/KF or NK/NF co-cultures in fresh serum-free DMEM. Whole-cell extracts and conditioned media were harvested and analysed separately.

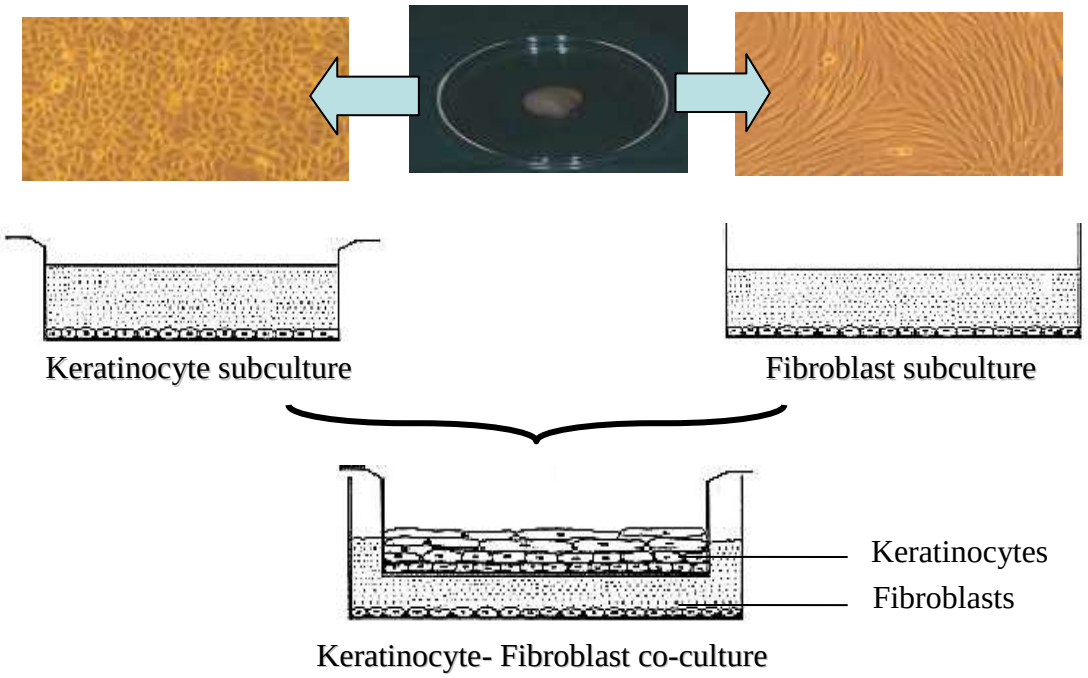


Figure 3.1: Co-culture of epidermal keratinocytes and dermal fibroblasts as an *in vitro* model to study epithelial-mesenchymal interactions. Figure courtesy of Dr Anandaroop Mukhopadhyay.

3.3.4 Treatment of fibroblasts with HDGF

KF and NF cells were seeded in 6-well plates at a density of 1×10^4 cells/ml in DMEM/10% FCS for 24 hrs and then in serum-free medium for another 48 hrs.. The cells were subsequently treated with 250 ng/ml of recombinant HDGF. Cells without treatment were used as controls. Whole-cell extracts and conditioned media were harvested at different time points and subjected to Western blot analysis for different molecular targets.

3.3.5 Treatment of keloid co-cultures with inhibitors

Keloid co-cultures were established as previously mentioned. The cells were subsequently treated with varying concentrations of various inhibitors: Rapamycin (0.01 and 2 µM; Calbiochem, CA), WP631 (0.05, 0.1, and 0.2 µM; Calbiochem) and

Mitoxantrone (0.05, 0.1 and 0.2 µM; Calbiochem) to investigate their effects on HDGF expression. Cells without inhibitor treatment were used as controls. After 72 h, conditioned media from the co-cultures was collected and subjected to Western blot analysis for HDGF expression.

3.3.6 Smad-null and Smad-overexpression cell assay

This assay was used to study the effect of Smad1, 2 and 3 on HDGF expression. Mouse embryo fibroblasts of wild-type (MEF-wt), Smad $1^{-/-}$ ($S1^{-/-}$), Smad $1^{+/+}$ ($S1^{+/+}$), Smad $2^{-/-}$ ($S2^{-/-}$) and Smad $3^{-/-}$ ($S3^{-/-}$) were kindly provided by Dr Rik Derynck, UCSF. Cells were seeded in six-well plates for 24 h, followed by serum-free DMEM starvation for 48 h before co-culture with keloid keratinocytes. Whole cell extracts and conditioned media were assayed for HDGF by Western blot.

3.3.7 MTT assay

The MTT assay is a commonly used colorimetric assay to quantify cell numbers. It is widely used in studies involving cell proliferation or cell toxicity. MTT [3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium] is cleaved by an active succinate-tetrazolium reductase system present in the mitochondrial respiratory chain of a living and metabolizing cell into blue formazon crystals which can be solubilized and their absorbance measured. The relationship between cell number and absorbance is linear. In our experimental set up, normal and keloid fibroblasts were seeded in 96 well plates. The cells were divided into control and treatment groups and were incubated with 10 ul of MTT (5mg/ml) in 100ul of DMEM to give a final concentration of 0.5mg/ml in each well

for about 2hrs. The medium was removed and the blue crystals were solubilized by Hansen's method (Hansen, Nielsen & Berg 1989) using 20% w/v SDS in a solution of DMF: water (1:1 v/v) and shaking in an orbital shaker. The absorbance of the solution was then measured directly by using a plate reader at 570nm.

3.3.8 Western blotting

Frozen tissue specimens or cultured fibroblasts under different experimental conditions were lysed in cell lysis buffer containing 20 mM Tris-HCl (pH 7.5), 1% Triton X-100, 100 mM NaCl, 0.5% Nonidet P-40 and 1 mg/ml protease inhibitor cocktail (Boehringer Mannheim, Mannheim, Germany). This was followed by centrifugation at 13000 x g for 10 min. Supernatant was collected while the pellet was discarded. Protein concentration of the tissue extracts were determined by Bradford method. Proteins were then subjected to Western blot analysis. In total, 50 µg of whole-cell extract was separated by 14% or 8% SDS-PAGE under reducing conditions and electroblotted onto a nitrocellulose membrane. Blots were incubated with numerous antibodies including mouse and rabbit anti-HDGF (a gift from Dr Ren Henning, MD Anderson Cancer Centre and Dr H Nakamura, Hyogo College of Medicine, Japan), mouse anti-PCNA (Santa Cruz, CA, USA), mouse anti-VEGF, anti-connective tissue growth factor (CTGF), rabbit anti-p44/p42 mitogen-activated protein kinase (MAPK), mouse anti-phospho p44/p42 MAPK, rabbit anti-Akt, rabbit anti-phospho Akt (Ser473) (Cell Signaling Technology Inc, USA), mouse anti-collagen I (Monosan), mouse anti-α-smooth muscle actin (SMA) (Sigma) and mouse anti-fibronectin (BD Transduction Laboratories). The blots were visualized with a chemiluminescence-based photoblot system (Amersham Biosciences, Buckinghamshire,

UK). For the analysis involving conditioned media, 4 ml of the conditioned media was concentrated using a Centricon centrifugal filter units (Millipore Corp., MA, USA) and then subjected to Western blotting.

3.3.9 Quantification of Western blot and statistical analysis

A Bio-Rad gel scanner and densitometer program (Gel-Pro Analyzer ver. 4.5; MediaCybernetics, Bethesda, MD, USA) was utilized to assess concentrations of the bands obtained by Western blots. These were measured as total density units. The paired Student's t-test or the Welch's t-test was used for all analyses where appropriate. A value of $P < 0.05$ was considered to be statistically significant. The error bars denote the standard error of the mean (SEM). All statistical analyses were done using Microsoft Excel 2003 (Redmond, WA, USA).

3.4 Microarray experiments

3.4.1 Cell culture

Three different keloid fibroblast samples and three different normal fibroblast samples that were previously maintained and stored at -150ºC were thawed and used for the experiments. Fibroblasts were seeded in 15cm dishes at a density of 1×10^4 cells/ml in 10% FCS until confluency and subsequently starved in a serum-free medium for 48 hrs. After 48 hrs, the serum free medium was replaced and fibroblasts were harvested after another 24 hrs (day 1), 72 hrs (day 3) and 120 hrs (day 5). Cells were grown and

processed in three batches. Each batch consisted of one keloid and one normal sample harvested at the three different time points. KF1, NF1, KF2 and NF2 were samples from different patients while KF3 and NF3 were samples from the same patient.

3.4.2 RNA extraction

RNA was extracted using the RNeasy-kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Cell culture medium was completely aspirated and 1 ml of Buffer RLT was added directly to each plate. Plates were scraped and the cell lysate collected with a cell scraper. Lysate was collected into a microcentrifuge tube and vortexed for 10 seconds. Lysate was then passed 5 times through a blunt 20-gauge needle fitted to an RNAse free syringe. 70% ethanol was added to the homogenized lysate in a 1:1 ratio and mixed by pipetting. 700 µl of the sample, including any precipitate that may have formed, was transferred to an RNeasy spin column placed in a collection tube and centrifuged for 15 s at $\geq$ 8000 x g. Successive aliquots of any excess of the sample were centrifuged in the same RNeasy spin column. Flow-through was discarded after each centrifugation. After all the sample has been loaded, 700 µl of Buffer RW1 was added to the RNeasy spin column and centrifuged for 15 s at $\geq$ 8000 x g to wash the spin column membrane. Flow-through was discarded and 500 µl of Buffer RPE was added to the RNeasy spin column and centrifuged for 15 s at $\geq$ 8000 x g. Flow-through was discarded and this step was repeated for a longer centrifugation time of 2 min. RNeasy spin column was then placed in a new 1.5 ml collection tube and 50 µl of RNase free water was added directly to the membrane. The column was centrifuged for 1 min at $\geq$ 8000 x g to elute the RNA. This step was repeated using another 50 µl of RNase free water. All steps were

performed at room temperature. Purified RNA was quantified by UV absorbance at 260 and 280 nm on a ND1000 spectrophotometer (Nanodrop^{TM}, ThermoScientific).

3.4.3 cRNA preparation and labeling

Labeled complementary RNA (cRNA) was produced from 15 µg of total RNA using the GeneChip One-Cycle Eukaryotic Target Labeling and Control Reagents (Affymetrix, Santa Clara, USA) according to the manufacturer's protocol. Briefly, 15 µg of total RNA was first reverse transcribed using a T7-Oligo(dT) Promoter Primer in the first-strand cDNA synthesis reaction. Following RNase H-mediated second-strand cDNA synthesis, the double-stranded cDNA was purified using the cDNA Cleanup Spin Column and served as a template in the subsequent *in vitro* transcription (IVT) reaction. The IVT reaction was carried out in the presence of T7 RNA Polymerase and a biotinylated nucleotide analog/ribonucleotide mix for cRNA amplification and biotin labeling. Biotin labeled cRNA was then purified using the IVT cRNA Cleanup Spin Column, quantified by UV absorbance at 260 nm on the ND1000 spectrophotometer and fragmented using the Affymetrix Fragmentation Buffer. All reagents were obtained from Affymetrix.

3.4.4 Affymetrix chip hybridization and scanning

Fragmented cRNA was then hybridized to preequilibrated Affymetrix GeneChip U133A arrays at 45 °C for 15 hours. The cocktails were removed after hybridization and the chips were washed and stained using Affymetrix wash buffers and stain cocktails in an automated fluidic station. The chips were then scanned in a Hewlett-Packard ChipScanner (Affymetrix, Santa Clara, USA) to detect hybridization signals.

3.4.5 Data analysis

Following data collection, preliminary analysis and visualization was done using the Affymetrix GeneChip Operating Software for an assessment of the quality of the data. Further statistical analysis was done using Version 10.0.2 of the Genespring GX software (Agilent, Palo Alto, CA). Normalization and summarization of arrays was done using both the Microarray Suite (MAS) 5.0 algorithm (default Affymetrix approach; Affymetrix Users Guide, www.affymetrix.com) and the Robust Multichip Analysis (RMA) (Bolstad et al. 2003) approach. The MAS 5.0 algorithm is the most widely used analysis method for GeneChips. The RMA algorithm is an alternative analysis procedure that is more robust than MAS 5.0 for data with normal errors or long-tailed symmetric errors. The data was first analyzed by Two-Way Analysis of Variance (ANOVA) to assess the individual influence of time point (day 1, day 3 or day 5) and cell type (NF or KF) on gene expression, as well as their net interactive effect. The Welch's t-test was then used to identify genes that were significantly different in keloid compared to normal fibroblasts. All statistical tests utilized the Benjamini Hochberg method to correct for multiple testing. Genes that were significantly different ($P < 0.05$) under both MAS 5.0 summarization and RMA summarization were used for further analysis. Hierarchical clustering and principal components analysis (PCA) were used to visually verify the ability of the genes selected to distinguish between keloid and normal cells. Finally, the list of genes that were significantly different was processed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Gene Functional Classification Tool (Dennis et al. 2003; Huang, Sherman & Lempicki 2009) to identify biological function associated with these genes.

3.5 Reverse engineering

3.5.1 Preparation of additional microarray samples

Two different keloid and two different normal fibroblast samples (all from different patients) were grown, serum-starved and harvested for RNA after day 1, day 3 and day 5 as in the previous study. In addition, one keloid fibroblast sample was grown, treated with HDGF and harvested for RNA after 6 hours, day 1 and day 2. All RNA were reverse transcribed, amplified and labeled using the GeneChip Two-Cycle Eukaryotic Target Labeling and Control Reagants (Affymetrix, Santa Clara, USA) according to the manufacturer's protocol. Unlike in the previous study where the One-Cycle protocol was used, the Two-Cycle protocol allowed for smaller amounts of starting RNA at the expense of longer running time in the form of a second reverse transcription step. As we had also run out of Affymetrix Genechip U133A arrays, the cheaper and newer Affymetrix Genechip U133 Plus 2.0 arrays were used to hybridize to the labeled cRNA. The chips were scanned in a Hewlett-Packard ChipScanner (Affymetrix, Santa Clara, USA) to detect hybridization signals. Raw microarray data in the form of .CELS files from Smith et al's experiments were also downloaded from the GEO database (Smith et al. 2008).

3.5.2 Data preprocessing

Following data collection, RMA and MAS 5.0 normalization and summarization were done using the R Bioconductor package. The four different datasets (original dataset using U133A arrays, new dataset using U133 Plus 2.0 arrays, HDGF dataset using U133

Plus 2.0 arrays and Smith's dataset using U133 Plus 2.0 arrays) were normalized and summarized independently. Two different custom Chip Definition Files (CDF) were used (Dai et al. 2005). The first CDF was based on the Ensembl Gene database for analysis with fREDUCE as it is easy to obtain the upstream sequence which is required by fREDUCE from the Ensembl database. The second was based on the Entrez Gene database for influence based reverse engineering methods such as BANJO and ARACNE as these probe mappings allow one to ignore any differential signal due to multiple probesets and gives a single value for a given gene. In addition, two lists were produced. In the first list, no filtering was done while in the second list, 25% of the lowly expressed genes were filtered.

3.5.3 Application of the fREDUCE algorithm

Human genomic sequences 1000 base pairs upstream from the transcriptional start site if known, or from the initiation codon, were extracted from the Ensembl database (Curwen et al. 2004). As fREDUCE requires only a single expression dataset and makes use of the entire genomic dataset (both signal and background), the datasets were compared as follows: A: Keloid versus normal fibroblasts under serum starvation conditions (only KF1, KF2, NF1 and NF2 were used to keep the number of samples close to the other conditions), B: Keloid versus normal fibroblasts under serum conditions, C: Keloid treated with steroid versus serum induced keloid fibroblasts, D: Normal treated with steroid versus serum induced normal fibroblasts, E: Keloid versus normal fibroblasts both treated with steroid and F: Keloid treated with HDGF versus untreated keloid fibroblasts. The expression value for each gene is represented as the following t-statistic:

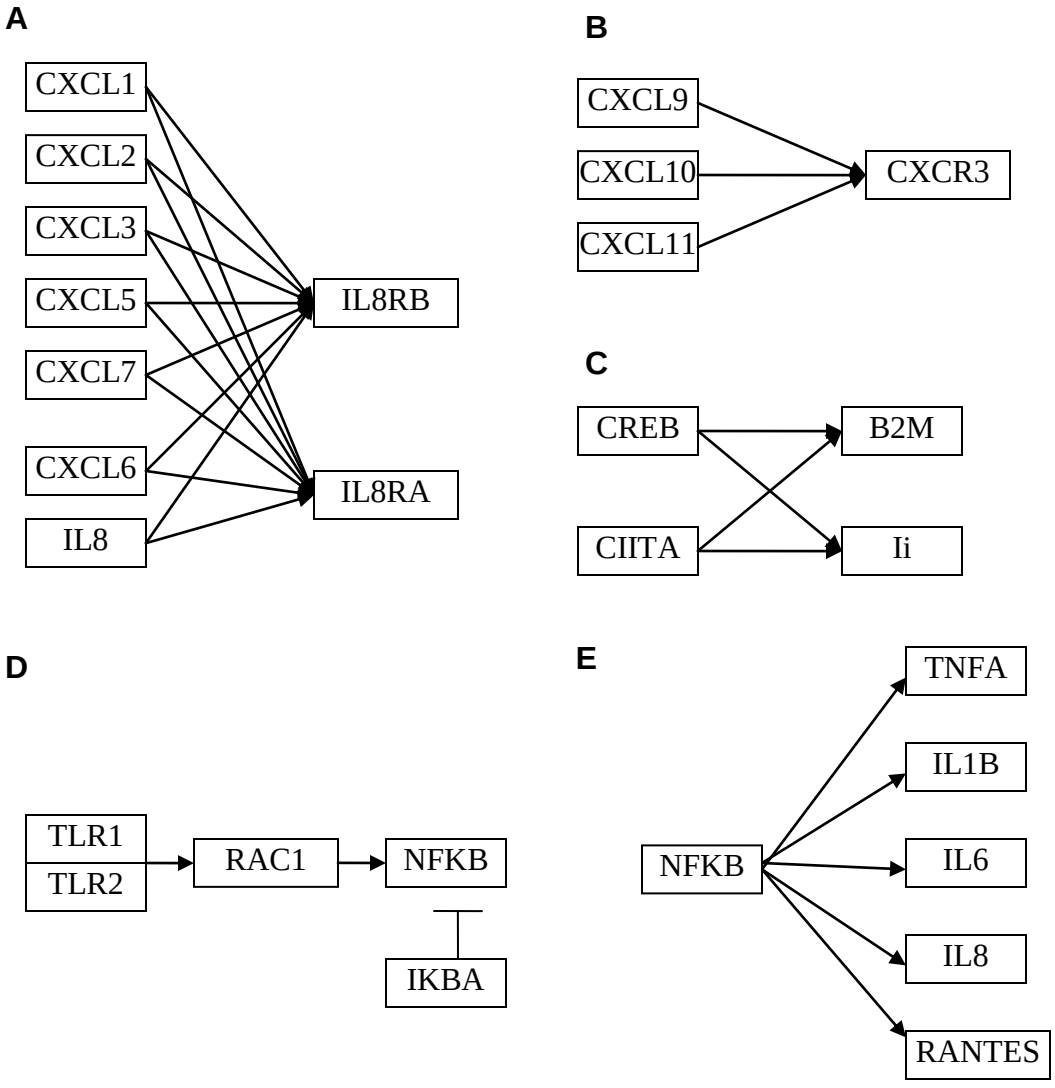
$$t^g = \frac{\mu_e{}^g - \mu_c{}^g}{\sqrt{\frac{Var_e{}^g}{n_e} + \frac{Var_c{}^g}{n_c}}}$$

where g is the index over genes, $\mu_e{}^g$ is the mean value of gene g under our condition of interest, $\mu_c{}^g$ os the mean value of gene g under control conditions, $Var_e{}^g$ is the variance of gene g under our condition of interest, $Var_c{}^g$ is the variance of gene g under control conditions, and n_e and n_c are the number of samples under our condition of interest and under control conditions respectively. This statisitic is similar to the z-statistic used by the fREDUCE creators (Wu et al. 2007). We then ran fREDUCE on the t-statistic for RMA normalized and MAS 5.0 normalized as well as unfiltered and filtered gene lists on the basis that a higher t-statistic translates to higher expression. Four different sets of parameters were run on each replicate: length 6 with 0 IUPAC substitutions, length 6 with 1 IUPAC substitution, length 7 with 0 IUPAC substitutions and length 7 with 1 IUPAC substitution. Top and consistent binding sequences obtained from fREDUCE above were then searched through the TRANSFAC database (Matys et al. 2003) for possible gene targets and their corresponding transcription factors. Only gene targets identified from Homo sapiens were collected, and binding sites for all these targets were reconfirmed to be located within the 1000 base pair upstream sequences collected from the Ensemble database previously.

3.5.4 Pathways selected for influence approach

KEGG pathways that were enriched from the previous study were used for the influence approach. These were the antigen presentation and processing pathway, cytokine-cytokine receptor interaction and toll-like receptor signaling pathway. The ribosome

pathway was not used as it would be a fully connected graph. Genes that were used as nodes for modeling were chosen on the basis that there is only one gene representing that particular node, all other genes will be assumed to be hidden nodes. The following 5 pathways in Figure 3.2 were eventually selected for the influence approach. Pathways were chosen such that cytokine receptor signaling networks, transcription factor networks and intracellular signaling networks were represented separately as the physical interactions in these networks are different.



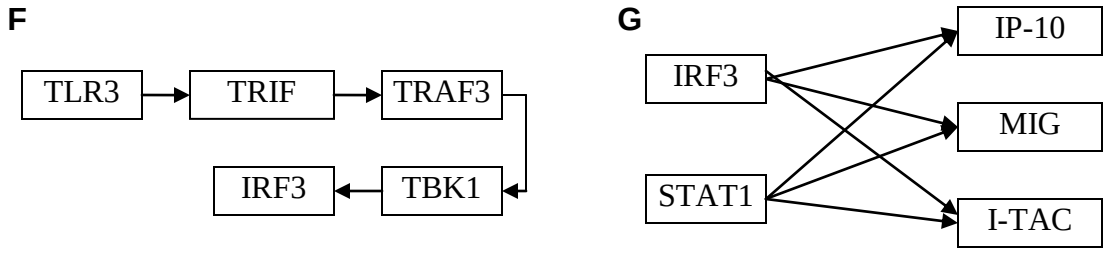


Figure 3.2: KEGG pathways used for the influence approach. (A and B) Pathways taken from the cytokine-cytokine receptor interaction map. (C) Transcriptional pathway taken from antigen processing and presentation map. (D, E, F and G) Pathways taken from the toll-like receptor signaling map. Pathways were also chosen such that A and B represent cytokine receptor interactions, C, E and G represent transcriptional networks and D and F represent intracellular signaling.

3.5.5 Application of the ARACNE and BANJO algorithms

The full set of data was used for the influence approach. To enable comparison between the different data sets, gene expression for all the relevant nodes were normalized using the average of GAPDH and B-actin expression. GAPDH and B-actin were first plotted to determine their correlation and outliers were removed from the dataset. Three keloid experiments from the serum starvation U133A dataset did not meet these criteria and was removed giving a total of 28 keloid experiments and 24 normal experiments. We ran ARACNE and BANJO on the keloid and normal inputs separately, and also on the MAS 5 and RMA normalized expression values separately. All parameters were left at their default values. For ARACNE, kernel width and number of bins were automatically detected by the software while DPI tolerance to remove false positives was set at 0.15. For BANJO, the Proposer/Searcher strategies were chosen as random local move and simulated annealing, respectively, and the amount of time BANJO uses to explore the Bayesian Network space was set to one minute. All the other parameters such as reannealingTemperature, coolingFactor, and so on, were left with their default values.

Parameter values were selected as best values (in terms of network inference accuracy) as shown by Bansal et al (Bansal et al. 2007). In order to estimate the joint probability distribution of all variables in the network, BANJO requires discrete data. The data was therefore discretized into 7 discrete states using the quantile discretization procedure in the software. Furthermore, as the simulated annealing algorithm in BANJO does not guarantee a global maximum, the runs were repeated three times and the result with the highest maximum score was taken.

3.5.6 Estimation of the performance of the algorithms

In order to assess the inference performances we computed the Positive Predicted Value (PPV) and the Sensitivity scores as described by Bansal et al (Bansal et al. 2007). The following definitions were used:

TP = Number of True Positives = number of edges in the real network that are correctly inferred;

FP = Number of False Positives = number of inferred edges that are not in the real network;

FN = Number of False Negatives = number of edges in the real network that are not inferred.

The following were then computed:

$$PPV = \frac{TP}{TP + FP}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

In order to compute the random PPV we considered the expected value of a

hypergeometrically distributed random variable whose distribution function and expected value are, respectively:

$$P_x = \frac{{}^{M}C_{X}\,{}^{N-M}C_{n-x}}{{}^{N}C_{x}}$$

$$E[x] = M\,\frac{{}^{N-1}C_{n-1}}{{}^{N}C_{n}} = M\,\frac{n}{N}$$

where N = number of possible edges in the network, M = number of true edges and n = number of predicted edges. Then,

$$PPV_{rand} = \frac{TP_{rand}}{TP + FP} = \frac{E[x]}{n} = \frac{M}{N}$$

All statistical tests are done using the one tailed paired t-test.

CHAPTER FOUR

THE ROLE OF HEPATOMA-DERIVED GROWTH FACTOR IN KELOID PATHOGENESIS

4.1 Introduction

Hepatoma-derived growth factor (HDGF) is a novel heparin-binding protein that was originally purified from the conditioned media of HuH-7 hepatoma cells (Nakamura et al. 1989). This growth factor was found to have mitogenic activity for a wide variety of cells, including fibroblasts (Abouzied et al. 2005), endothelial cells (Everett et al. 2004), renal (Kishima et al. 2002) and lung epithelial cells (Mori et al. 2004), vascular smooth muscle cells (Everett, Stoops & McNamara 2001) and fetal hepatocytes (Enomoto et al. 2002). A growing number of studies report a possible role of HDGF in the development of different types of cancers. In particular, it has been implicated in esophageal cancer (Matsuyama et al. 2001; Yamamoto et al. 2007), pancreatic cancer (Uyama et al. 2006), hepatocellular carcinoma (Yoshida et al. 2003), melanoma (Bernard et al. 2003), lung cancer (Ren et al. 2004) and gastric carcinoma (Chang et al. 2007).

HDGF is the first member of the HDGF family of proteins that was discovered to contain a well conserved N-terminal amino acid sequence, which is called the homologous to amino terminus of HDGF (HATH) region (Izumoto et al. 1997). Subsequently, five related proteins have been identified, four of which are named HDGF-related protein (HRP) -1 to HRP-4 and the fifth of which is named lens epithelium-

derived growth factor (LEDGF). Except for their growth factor activity, the functions of these proteins are largely unknown (Abouzied et al. 2004).

Although the mitogenic effect of HDGF has been proven, the pathway by which it exerts this proliferative activity is still unclear. Two different pathways have been proposed. Despite lacking the secretory sequence present in most secretory proteins (Nakamura et al. 1994), it has been shown that exogenous HDGF could possibly act by binding to an as yet unknown cell surface receptor, triggering signaling events downstream that result in increased proliferation (Abouzied et al. 2005). Others have shown that nuclear localization is required for the mitogenic activity of HDGF (Everett, Stoops & McNamara 2001; Kishima et al. 2002).

In terms of wound repair, HDGF has been found to be involved in lung remodeling after injury by promoting the growth of lung epithelial cells (Mori et al. 2004). Among the growth factors responding to vascular wall injury, HDGF is unique, in that it is not expressed in the vascular wall until injury occurs (Everett et al. 2000). HDGF gene expression was also increased during retinal pigment epithelial wound repair (Singh et al. 2001). Furthermore, HDGF has been found to be up-regulated in human dermal fibroblasts subjected to mechanical stimulation from stressed collagen lattices (Kessler et al. 2001). Abnormal scarring has been correlated to regions of the body with higher mechanical force than others (Wang et al. 2006; Aarabi et al. 2007). Therefore, these results, combined with the fact that HDGF has been implicated in the aberrant growth of tumours, lead us to speculate that it could also play some role in the formation of keloids. In unpublished data from microarray experiments done by my supervisor, Prof

Phan at Stanford University, HDGF was one of the genes found to be up-regulated in keloid tissue.

In this study, we investigated the expression and localization of HDGF by performing immunohistochemical staining (IHC) and Western blot analysis on keloid and normal skin tissue. We further studied the expression of HDGF using in vitro models of normal and keloid fibroblasts subjected to serum stimulation. To examine the effect of epithelial-mesenchymal interactions on the expression of HDGF, we employed a two chamber serum free system where keratinocytes on membrane inserts were co-cultured with the fibroblasts.

In a second set of experiments, we examined the effect of exogenous recombinant HDGF on the keloid and normal fibroblasts. Cells treated with recombinant HDGF were assessed for increased proliferation by the MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] assay and by quantifying proliferating cell nuclear antigen (PCNA) expression. Western blotting was also performed to identify some of the downstream signaling targets of exogenously applied HDGF.

Finally, to identify some of the upstream signals regulating the expression of HDGF, we investigated the effect of Sp1 and mammalian target of rapamycin (mTOR) inhibitors on the secretion of HDGF from fibroblast keratinocyte co-cultures. The effect of TGF-β signaling on HDGF expression was also examined by assaying Smad-null and Smad-overexpressing mouse embryo fibroblasts.

4.2 Results

4.2.1 HDGF expression is increased in keloid scar dermis

Immunohistochemical labelling showed that HDGF was present in both the epidermis (Fig. 4.1A) and the dermis (Fig. 4.1B) of normal and keloid tissue. Epidermal staining intensity was irregular and sample-dependent, with some keloid samples exhibiting stronger staining, while others exhibiting equal or weaker staining when compared with their normal counterparts. However, HDGF expression in the dermis was found to be higher in all keloid samples. This can be more clearly seen at a lower magnification (Fig. 4.1C). In the keloid tissues, almost the whole dermis was stained brown compared with a significantly smaller area in normal skin. Western blot results from the keloid and normal whole-tissue extracts reconfirmed these observations. The keloid tissue samples had a significantly higher expression of HDGF compared with the normal tissue samples ($P < 0.05$; Fig. 4.2).

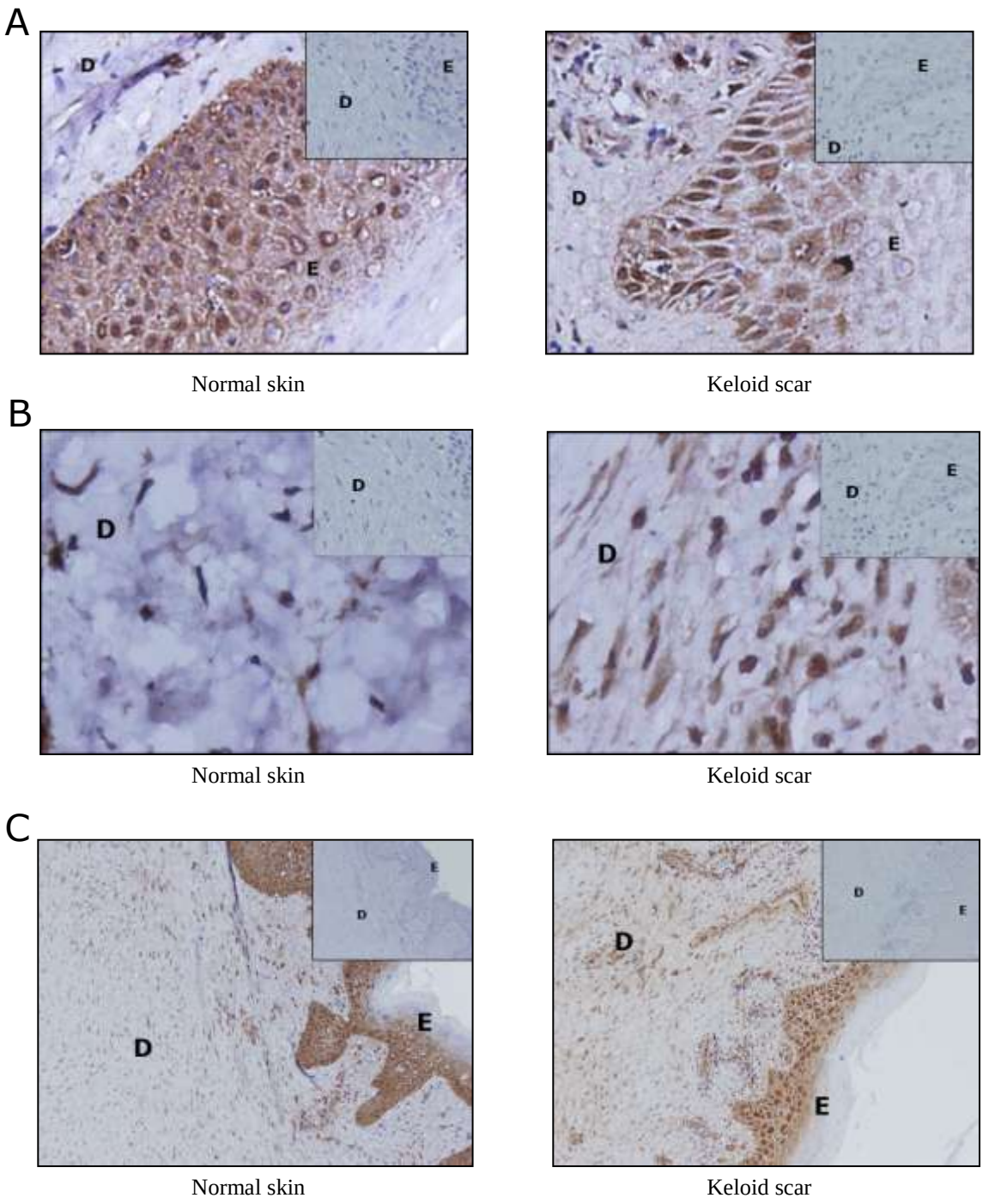


Figure 4.1: Immunohistochemical staining of keloid and normal tissue for HDGF. Paraffin sections of normal and keloid tissue were prepared and stained with antibodies against HDGF. Pictures were taken with magnification at 40X (**A**, **B**) and 10X (**C**). The dermis and the epidermis are represented by (**D**) and (**E**), respectively. In each panel, the inset shows the same tissue labelled with a non-immune mouse antibody of the appropriate immunoglobulin isotype as a negative control. HDGF was detected in both the epidermis (**A**) and the dermis (**B**) of normal and keloid tissue. Increased expression was observed in the dermis of keloid tissue compared with the dermis of normal tissue (**B**, **C**).

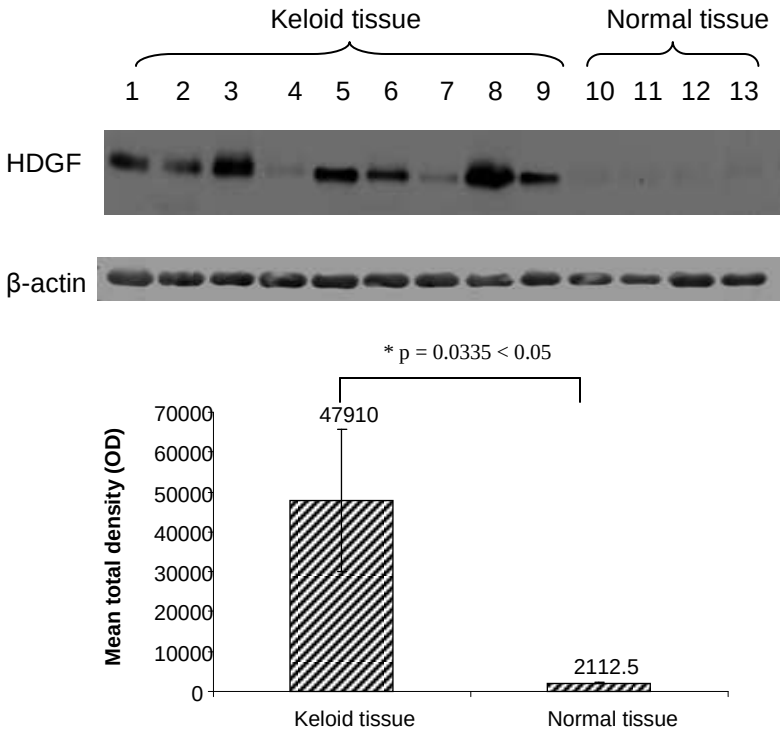


Figure 4.2: Western blot of keloid and normal whole tissue extract. In total, 50 µg of tissue extracts from nine keloid tissue specimens and four normal skin specimens was subjected to Western blot analysis with antibodies against HDGF. The whole-tissue extracts include both the epidermis and the dermis. The blots were probed with anti-β-actin antibody to confirm equal loading. The bar graph represents the mean ± SEM of HDGF levels in the normal and keloid samples, as quantified by gel densitometry. *indicates statistical significance as assessed by Welch's t-test.

4.2.2 Serum stimulation and epithelial-mesenchymal interactions had no effect on intracellular HDGF expression

Western blot results indicated that treatment with serum had no significant effect on intracellular HDGF expression in both normal fibroblasts (NF) and keloid fibroblasts (KF) (Fig. 4.3). In addition, NFs co-cultured with normal keratinocytes (NK) and KFs co-cultured with keloid keratinocytes (KK) did not show any significant difference in HDGF expression levels when compared with the monocultured controls or when compared with each other.

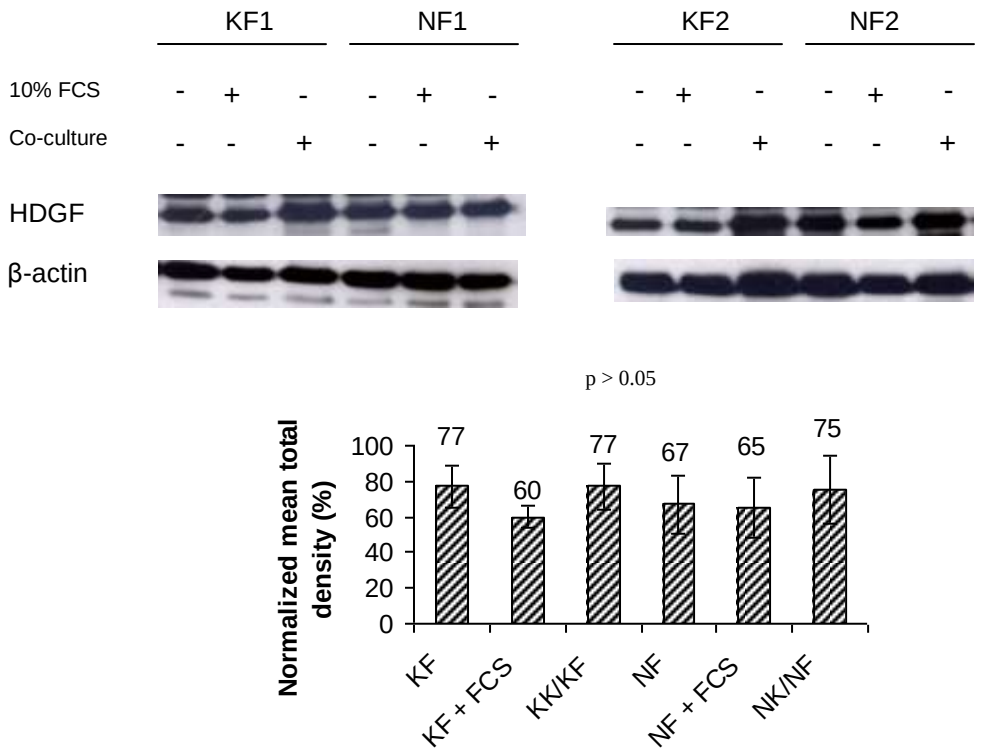


Figure 4.3: Effect of serum and epithelial–mesenchymal interactions on intracellular HDGF expression. Six different strains of keloid/normal fibroblasts were cultured with DMEM, 10% FCS or co-cultured with keloid/normal keratinocytes for 5 days. In total, 50 µg of total protein extracts was subjected to Western blot analysis with HDGF antibodies. Two representative strains are shown. The bar graphs represent the normalized mean ± S.E.M. of HDGF levels in the different conditions. All blots were probed and normalized with β-actin.

4.2.3 Epithelial-mesenchymal interactions in keloid co-culture increased secretion of HDGF

Conditioned media obtained when KFs were co-cultured with KKs showed a significant increase in HDGF compared with conditioned media obtained from monocultured KFs from day 1 to 5 ($P < 0.05$ for asterisks; Fig. 4.4A). In contrast, HDGF was undetected from day 1 to 3, and a weak increase was only detected when NFs were cocultured with NKs on day 5 ($P < 0.05$; Fig. 4.4B). Monocultured KFs showed a higher secretion compared with monocultured NFs and keloid co-cultures showed a higher secretion

compared with normal skin cell co-cultures. Monocultured keratinocytes show a moderately high secretion of HDGF but no significant difference was seen between NKs and KKs harvested at day 5 (Fig. 4.4C).

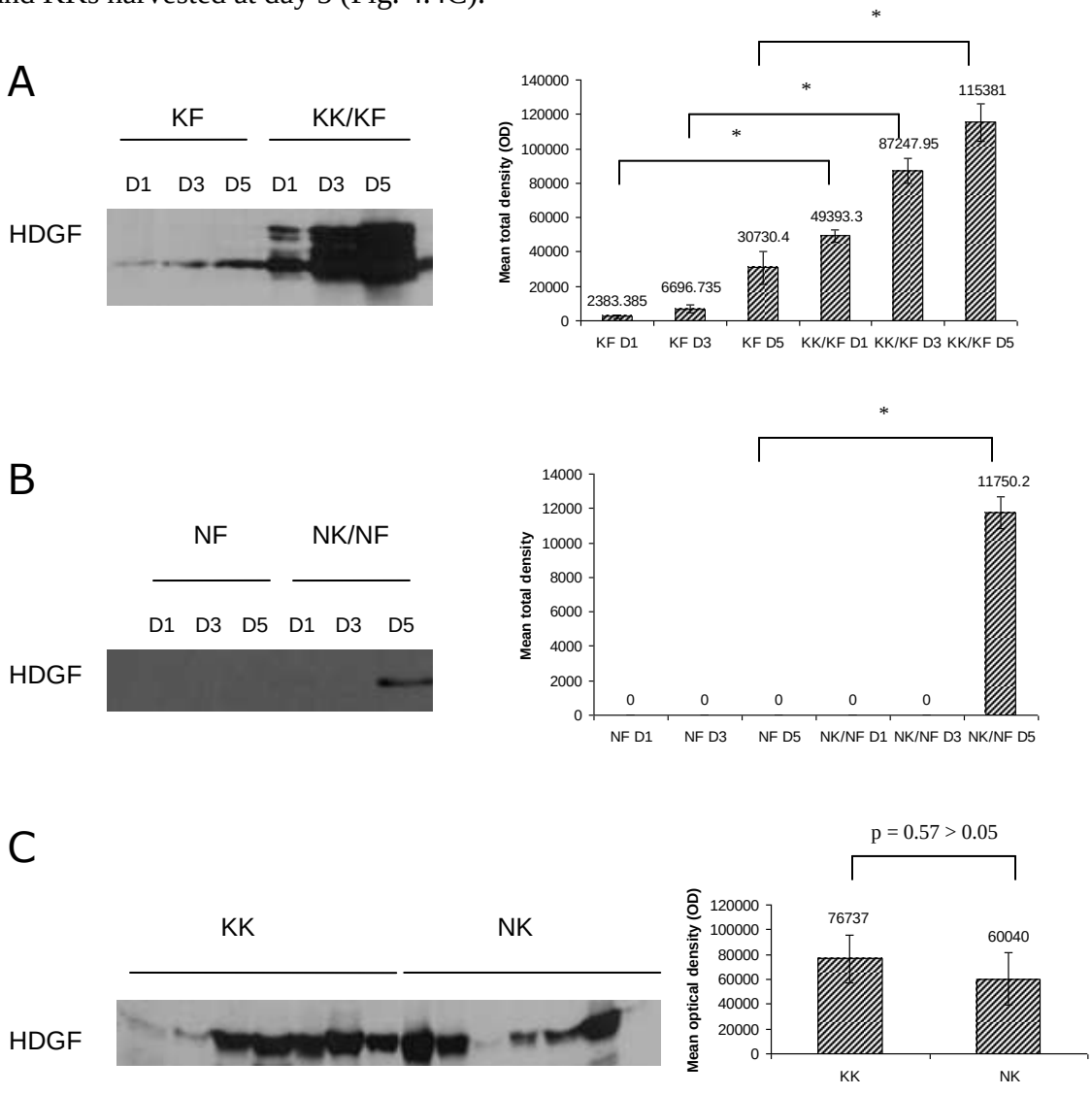


Figure 4.4: Expression of HDGF in conditioned media of monocultured and co-cultured cells. (**A**) Conditioned media of keloid fibroblasts monoculture (KF) and keloid fibroblasts cocultured with keloid keratinocytes (KK/KF) were collected at days 1, 3 and 5. (**B**) Conditioned media of normal fibroblast monoculture (NF) and normal fibroblast co-cultured with normal keratinocytes (NK/NF) were collected at days 1, 3 and 5. Experiments were performed in duplicates. (**C**) Conditioned media of seven samples of singly cultured keloid keratinocytes (KK) and normal keratinocytes (NK) were collected at day 5. Four millilitres of the conditioned media from (**A**), (**B**) and (**C**) was then concentrated and subjected to Western blot analysis with anti-HDGF antibody. Representative figures are shown. The bar graphs represent the mean ± S.E.M. of HDGF levels. * indicates statistical significance as determined by the paired t-test.

4.2.4 Increased keloid fibroblast proliferation upon stimulation with HDGF

There was no significant difference in NF proliferation when treated for 72 hrs with various doses of HDGF (Fig. 4.5A). However, there was a significant dose-dependent increase of up to ~13% in the proliferation of KFs ($P < 0.05$ for asterisks; Fig. 4.5B). In addition, Western blot analysis after 48 hrs showed a significant increase of proliferating cell nuclear antigen (PCNA) expression in treated KFs compared with untreated keloid controls, but this effect was not seen in treated NFs compared with untreated normal controls ($P < 0.05$; Fig. 4.6A). In total, 250 ng/ml of recombinant HDGF was used for treatment of both KFs and NFs.

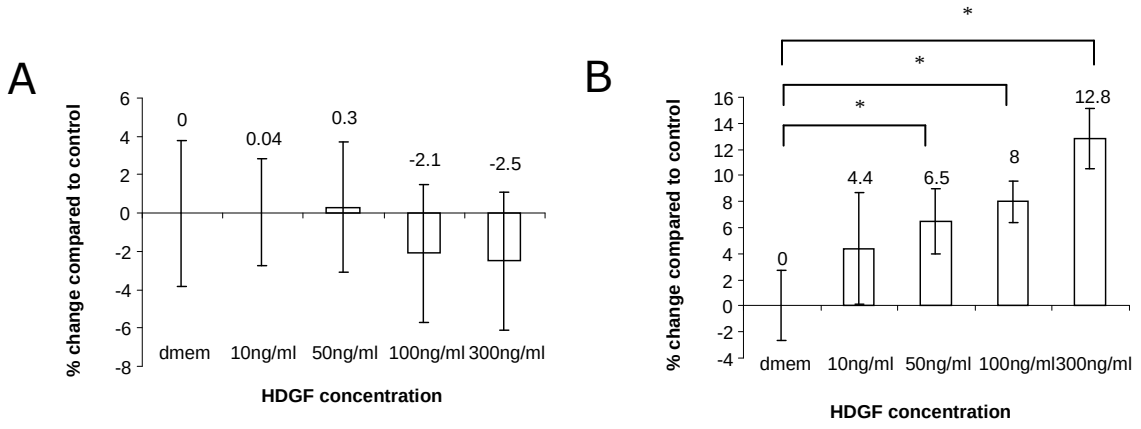
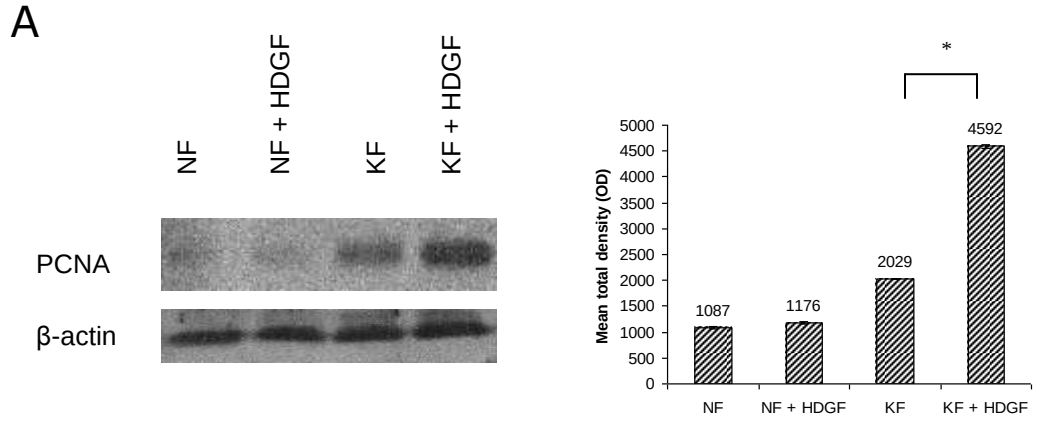
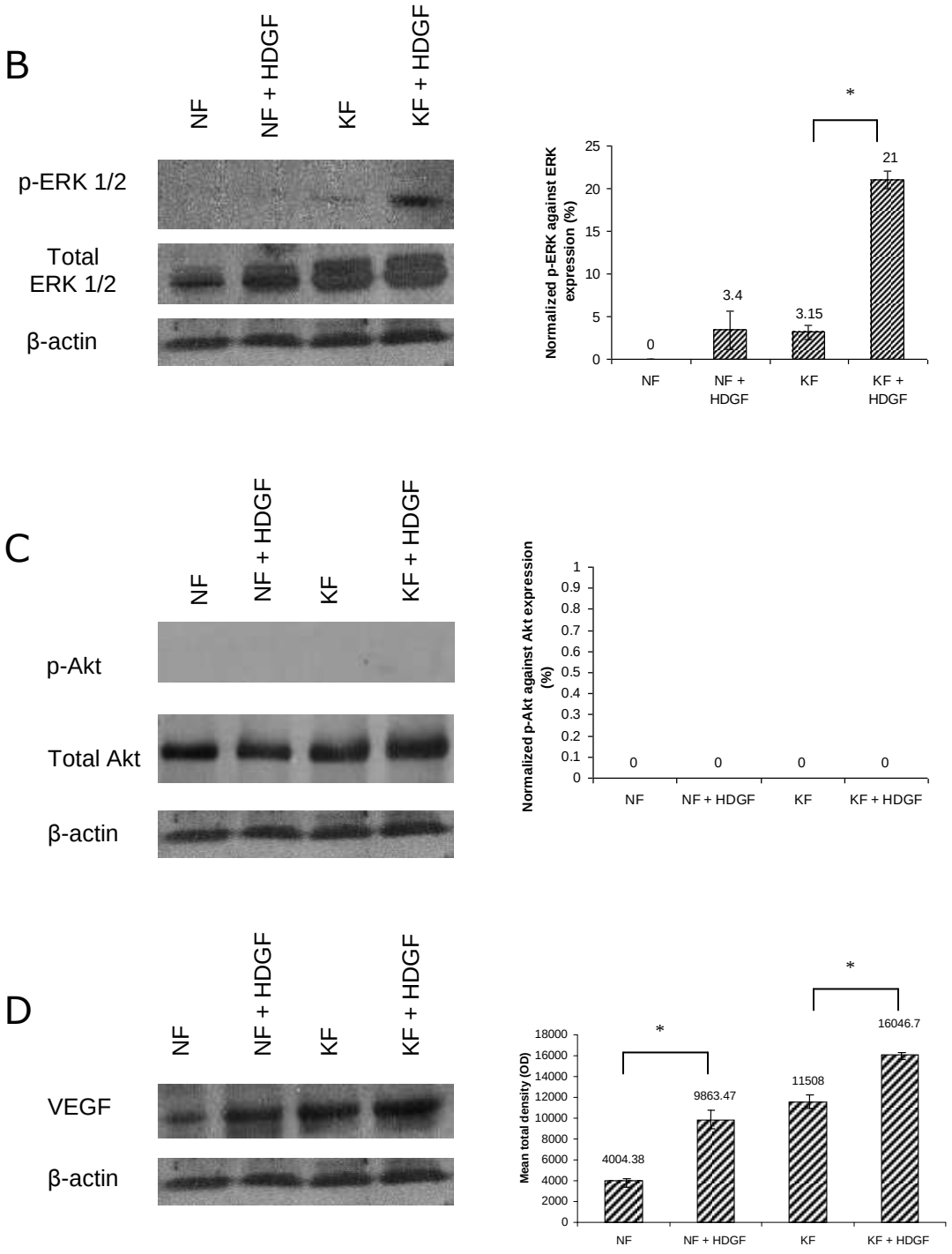


Figure 4.5: Increased proliferation of keloid fibroblasts treated with recombinant HDGF. Cultures of keloid or normal fibroblasts were grown until 50% confluence and then serum starved for 48 hrs. The fibroblasts were then treated with HDGF (10, 50, 100 and 300 ng/ml) for 72 hrs and then subjected to the MTT proliferation assay. Untreated samples were used as control. The bar graph in (**A**) represents the mean proliferative response of treated normal fibroblasts as a percentage of the control. The bar graph in (**B**) represents the mean proliferative response of treated keloid fibroblasts as a percentage of the control. * indicates statistical significance compared with DMEM control as assessed by Student's t-test.

4.2.5 Treatment of fibroblasts with HDGF activated the ERK pathway, increased the secretion of VEGF, and decreased the secretion of collagen I

KFs treated with HDGF after 48 hrs showed a significant increase in the expression of intracellular phospho-extracellular signal regulated kinase (ERK) 1/2 compared with untreated keloid controls, but this increase was not seen in treated NFs compared with untreated normal controls ($P < 0.05$; Fig. 4.6B). Intracellular expression ($P < 0.05$, Fig. 4.6D) and extracellular secretion ($P < 0.05$, Fig. 4.7A) of VEGF from both keloid and normal fibroblasts were significantly increased upon treatment with HDGF. Secretion of collagen I was downregulated in the conditioned media of treated keloid and normal fibroblasts compared to untreated controls ($p<0.05$, Fig. 4.7B). Treatment with HDGF did not produce any significant difference in the expression of intracellular α-SMA (Fig. 4.6E), extracellular fibronectin (Fig. 4.7C) and extracellular CTGF (Fig. 4.7D). Furthermore, expression of p-Akt was undetectable in all samples (Fig. 4.6C). At earlier time points, no significant increase in phospho-ERK was detected (Fig. 4.6F). In total, 250 ng/ml of recombinant HDGF was used for treatment of both KFs and NFs.




B
NF
NF + HDGF
KF
KF + HDGF
p-ERK 1/2
Total ERK 1/2
β-actin
Normalized p-ERK against ERK expression (%)
0
3.4
3.15
21
*
C
p-Akt
Total Akt
β-actin
Normalized p-Akt against Akt expression (%)
D
VEGF
β-actin
Mean total density (OD)
4004.38
9863.47
11508
16046.7

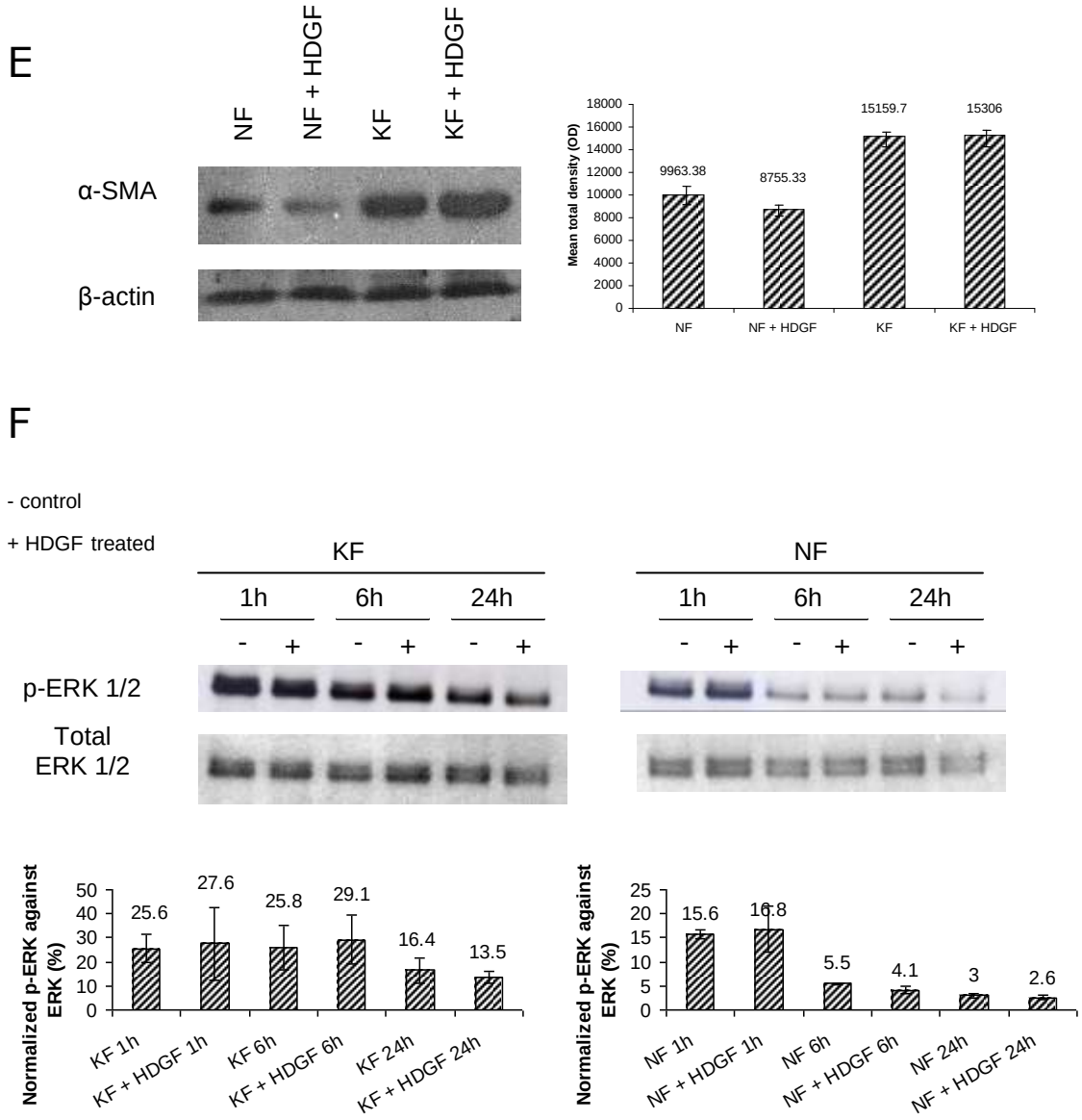


Figure 4.6: Effect of HDGF on the expression of downstream intracellular targets. Normal fibroblasts and keloid fibroblasts were treated with either DMEM or 250 ng/ml of recombinant HDGF, harvested after 48 hrs and lysed for Western blot analysis, as described under experimental procedures. Blots were incubated with anti-PCNA (**A**), anti-phospho-ERK 1/2 and total ERK 1/2 (**B**), anti-phospho-Akt and total Akt (**C**), anti-VEGF **(D)** and anti-α-SMA **(E)** antibodies The blots were also incubated with anti-β-actin antibody to confirm equal loading. In another set of experiments, normal fibroblasts and keloid fibroblasts were treated with either DMEM or HDGF and harvested after 1 hr, 6 hrs and 24 hrs for Western blot analysis. Blots were incubated with anti-phospho-ERK 1/2 and total ERK 1/2 (**F**) antibodies. All experiments were performed in duplicates. Representative figures are shown. The bar graphs represent the mean ± S.E.M. of protein levels. Phospho-ERK 1/2 was normalized against total ERK 1/2 expression and phosphor-Akt was normalized against total Akt expression. * indicates statistical significance as determined by the paired t-test.

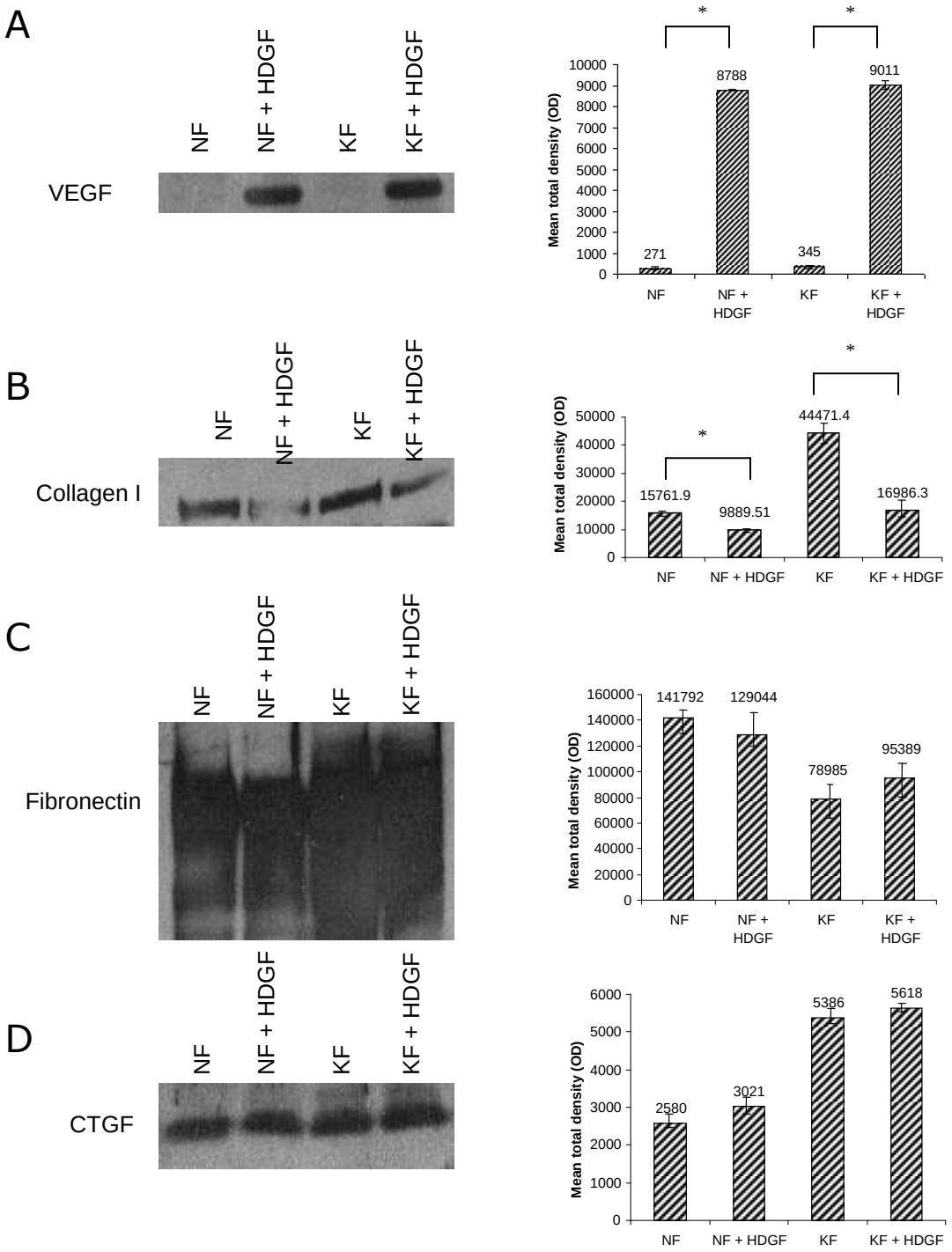


Figure 4.7: Effect of HDGF on the expression of downstream extracellular targets. Normal fibroblasts and keloid fibroblasts were treated with either DMEM or 250 ng/ml of recombinant HDGF. After 48 h, four millilitres of conditioned media was concentrated and subjected to Western blot analysis. Blots were then incubated with anti-VEGF (**A**), anti-collagen I (**B**), anti-fibronectin (**C**) and anti-CTGF (**D**) antibodies. All experiments were performed in duplicates. Representative figures are shown. The bar graphs represent the mean ± S.E.M. of protein levels. * indicates statistical significance as determined by the paired t-test.

4.2.6 Treatment with mTOR and Sp1 inhibitors did not significantly affect the production of HDGF

Treatment with the mTOR inhibitor, Rapamycin and Sp1 inhibitors, Wp631 and Mitoxantrone, for 72 hours did significantly affect the production of intracellular (Fig. 4.8A) as well as extracellular (Fig 4.8B) HDGF. Extracellular levels of HDGF appear to decrease slightly upon treatment with both the mTOR and Sp1 inhibitors but the decrease was not significant. Intracellular levels of HDGF appear to decrease slightly upon treatment with Sp1 inhibitors, but again this decrease was not significant.

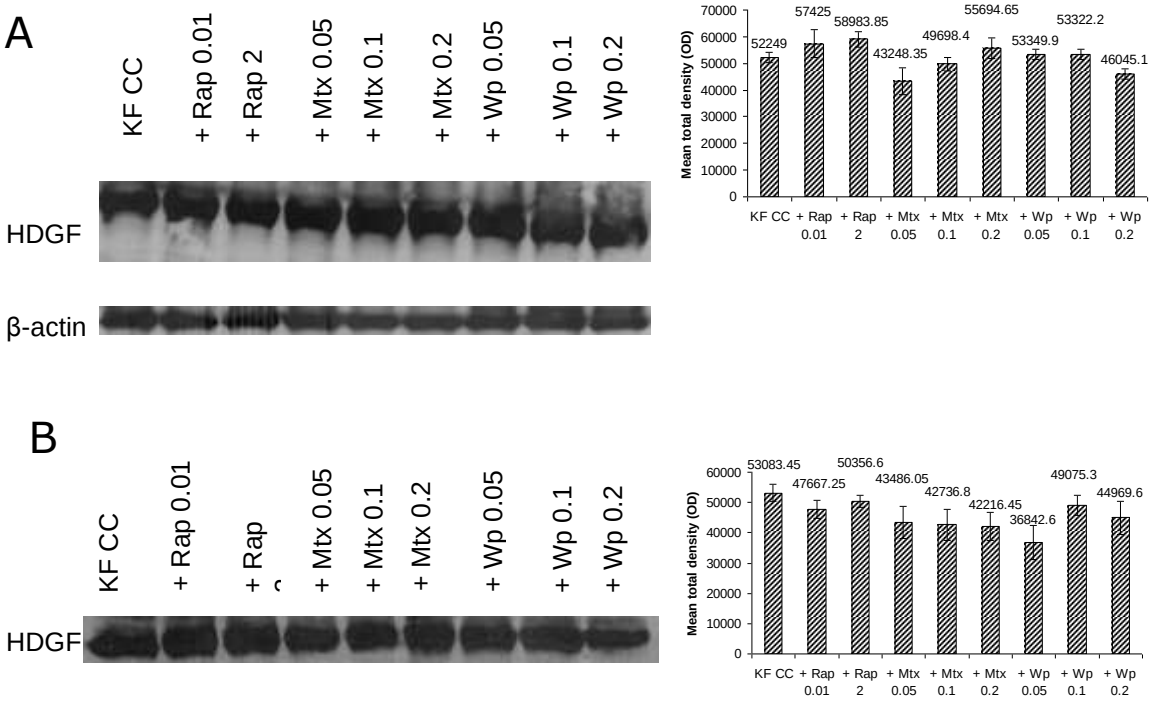


Figure 4.8: Effect of mTOR and Sp1 inhibitors on the expression of HDGF. mTOR inhibitor (rapamycin) and Sp1 inhibitors (WP631 and mitoxanthrone) were added to the keloid coculture for 72 h with indicated concentrations as described in Material and Methods. Untreated co-cultures served as controls. (**A**) For analysis of intracellular HDGF production, KFs were harvested and lysed for Western blot analysis, as described under experimental procedures. The blots were also incubated with anti-β-actin antibody to confirm equal loading. (**B**) For analysis of secreted HDGF levels, 4 ml of conditioned media was concentrated and subjected to Western blot analysis with anti-HDGF antibody. All experiments were performed in duplicates. Representative figures are shown. The bar graphs represent the mean ± S.E.M. of HDGF levels.

4.2.7 Knockout of Smad 2/3 signaling increases intracellular HDGF expression while knockout of Smad 1 signaling increases extracellular HDGF expression

Smad $2^{-/-}$ ($P<0.05$, Fig. 4.9C) and Smad $3^{-/-}$ ($P<0.05$, Fig. 4.9D) mouse embryo fibroblast cells had a significantly higher expression of intracellular HDGF compared to their respective wild type cells. However, intracellular HDGF expression in Smad $1^{-/-}$ (Fig. 4.9A) and Smad $1^{+/+}$ (Fig. 4.9B) cells was not significantly different from the wild type. Conditioned media from monocultured Smad $1^{-/-}$ cells had a significantly higher expression of HDGF compared to the wild type cells and conversely, conditioned media from monocultured Smad $1^{+/+}$ cells had a significantly lower expression of HDGF compared to the wild type control ($P<0.05$ for asterisks, Fig. 4.10A). Secretion from monocultured Smad $2^{-/-}$ and Smad $3^{-/-}$ cells does not appear to be significantly different from wild type controls. Upon co-culturing with keloid keratinocytes, this effect was abrogated and there was no significant difference in secretion from the Smad co-cultures compared to the keloid co-culture control (Fig. 4.10B).

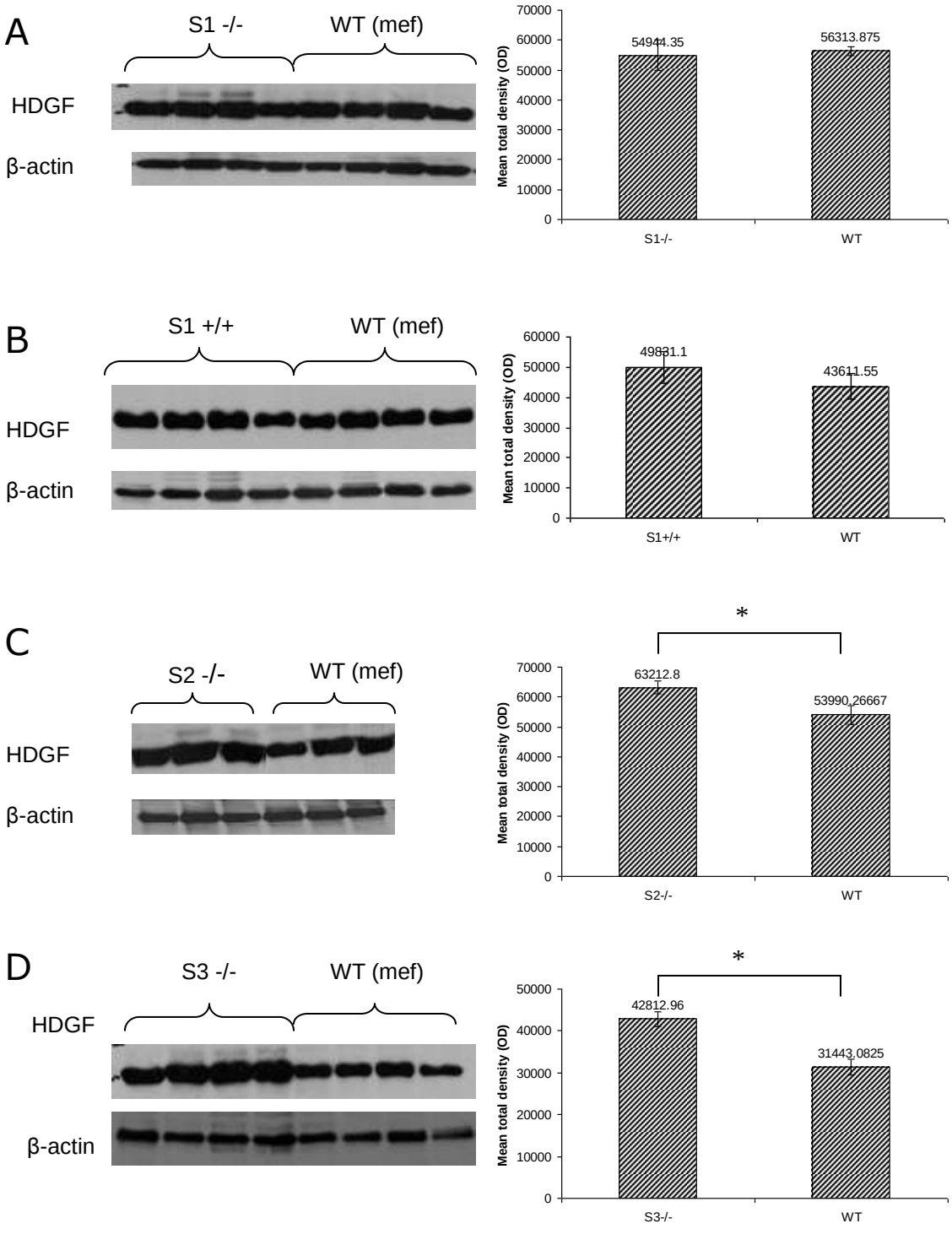


Figure 4.9: Effect of Smad signaling on intracellular HDGF expression. (**A**) $S1^{-/-}$, (**B**) $S1^{+/+}$, (**C**) $S2^{-/-}$ and (**D**) $S3^{-/-}$ cells were seeded in six-well plates for 24 h in 10% FCS. Medium was replaced by serum-free DMEM for another 48 h before co-culturing with keloid keratinocytes. Cells were harvested for Western blot analysis of HDGF. The bar graphs represent the mean ± S.E.M of HDGF levels. All blots were probed with β-actin antibody to confirm equal loading. * indicates statistical significance as assessed by the paired t-test.

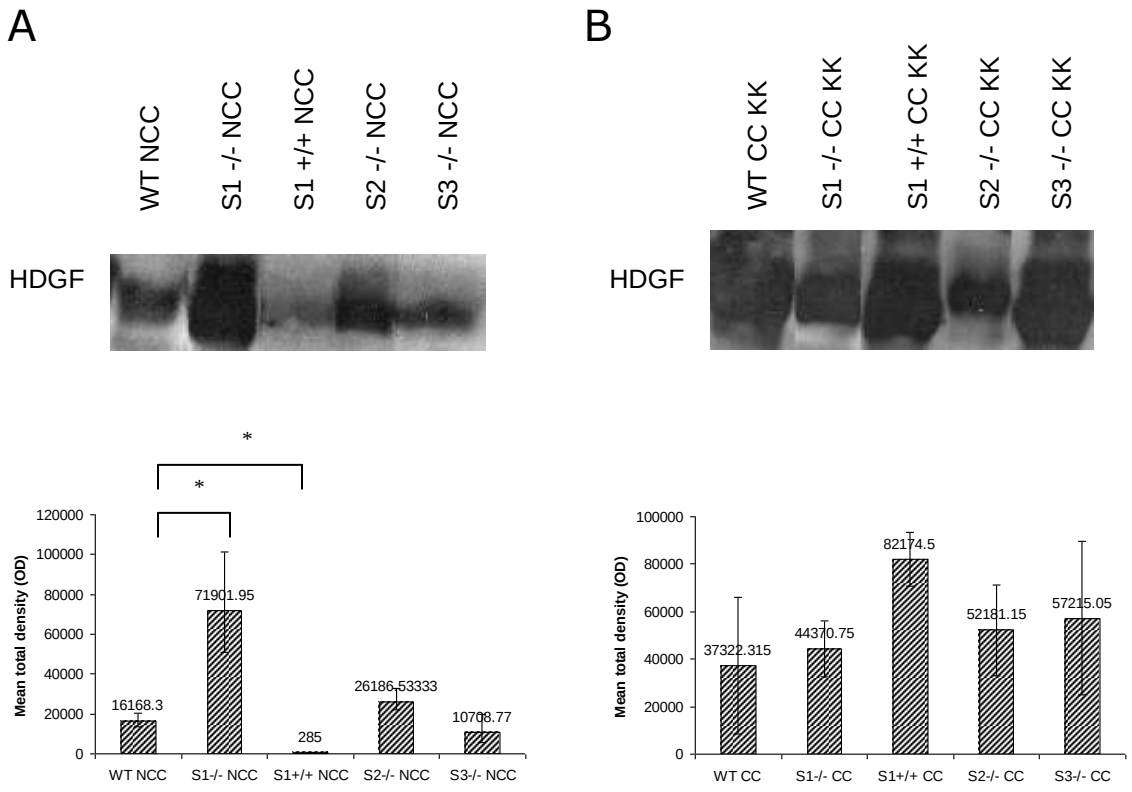


Figure 4.10: Effect of Smad signaling on extracellular HDGF expression. 4 ml of conditioned media from $S1^{-/-}$, $S1^{+/+}$, $S2^{-/-}$ and $S3^{-/-}$ mouse embryo fibroblasts monoculture (**A**) and co-cultured with keloid keratinocytes (**B**) was collected after 48 h. The samples were then concentrated and subjected to Western blot analysis with anti-HDGF antibody. Experiments were performed in duplicates. Representative figures are shown. The bar graphs represent the mean ± S.E.M. of HDGF levels. * indicates statistical significance as determined by the paired t-test.

4.3 Discussion

HDGF has not been as well studied as other growth factors and cytokines, but there is a growing wealth of information underlining its importance in cancer. In this study, we examined the hypothesis that HDGF could play a role in the dysregulation of wound

healing leading to keloid formation. Our data suggest that this growth factor is likely to be only one of a myriad number of players involved in keloid pathogenesis

Immunohistochemical results from this study indicate that HDGF is quite highly expressed in the epidermis of both keloid and normal skin tissue. However, in the dermal layer, there is a higher expression of HDGF in keloid tissue compared with normal skin. This result led us to focus our efforts on the fibroblasts, which are the major cell type in the dermis. Preliminary *in vitro* results demonstrated a significant expression of both intracellular and extracellular HDGF in monocultured epidermal keratinocytes. We were unable to detect any significant difference in HDGF expression between KK and NK samples grown in vitro. However, the very presence of HDGF in these cells suggested that this growth factor played some as yet unknown role in epidermal biology that is worth investigating.

The process of cutaneous wound healing can be arbitrarily divided into four phases - hemostasis, inflammation, proliferation and remodeling. The serum stimulation model is an in vitro model that can be used to determine the involvement of growth factors or cytokines in the early stages of the wound healing phase. Fibroblasts interpret the presence of serum as a physiological wounding signal and would respond to it as if it had occurred in vivo (Iyer et al. 1999). Our data show no difference in HDGF expression levels between serum-stimulated and non-serum-stimulated fibroblasts, suggesting the lack of involvement of HDGF in the early phases of wound healing.

Recent studies from our group and others have shown the importance of paracrine signalling via epithelial–mesenchymal interactions in keloid pathogenesis (Lim et al. 2001; Funayama et al. 2003). The co-culture experiments were used to assess the effect

of this paracrine signaling on HDGF expression. We found no significant difference in intracellular HDGF expression in whole-cell extracts of co-cultured fibroblasts compared with the singly cultured fibroblasts. However, conditioned media collected from the keloid co-culture configuration had significantly higher levels of HDGF compared with those collected from the singly cultured KFs. Furthermore, secretion of HDGF from the keloid co-culture configuration was also higher than from normal skin cell co-culture configuration. When singly cultured, the keratinocytes secreted fairly high amounts of HDGF but there was no significant difference between KKs and NKs. This suggests that the secretion of HDGF is somehow modulated by epithelial–mesenchymal interactions.

Taken together, our in vitro results suggest that intracellular HDGF may have some role to play in the normal process of wound healing, as evident by its high basal expression in both normal and keloid fibroblasts. However, intracellular HDGF levels alone cannot account for the higher expression of HDGF that was observed in *in vivo* samples of keloid tissue. It appears that the contribution of HDGF to keloid pathogenesis lies mainly in its *secreted* form and this secretion is modulated by epithelial-mesenchymal interactions. To date, two different pathways have been proposed to account for the activity of HDGF, either by binding to a cell surface receptor and acting as a secreted growth factor (Abouzied et al. 2005), or by translocating to the nucleus where it acts as a nuclear transcription factor (Everett, Stoops & McNamara 2001; Kishima et al. 2002). Our data suggests that in the formation of keloids, the secretory pathway of HDGF is of greater importance. However, we are not able to completely rule out nuclear translocation as it is possible that the increase in intracellular HDGF is too slight to be detected by Western analysis.

HDGF has been reported to act as a potent exogenous mitogen for a wide variety of cells, including 3T3 fibroblasts (Nakamura et al. 1989; Abouzied et al. 2005) and mice dermal fibroblasts (Gallitzendoerfer et al. 2008). The susceptibility of KFs to other mitogenic stimuli has also been previously established (Lim et al. 2001; Phan et al. 2003; Xia et al. 2004). Therefore, it was not surprising that both our MTT and PCNA results showed KFs having a better proliferative response to HDGF stimuli compared with their normal skin counterparts.

Downstream molecular targets of HDGF that were found to be up-regulated after 48 hrs include ERK and VEGF, while those found to be unaffected include alpha-smooth muscle actin (α-SMA), fibronectin and connective tissue growth factor (CTGF). We were however unable to detect any increase in ERK phosphorylation at the earlier time points of 1 hr, 6 hrs and 24 hrs. HDGF has previously been found to induce the phosphorylation of ERK in human pulmonary endothelial cells (Everett et al. 2004) and gastric epithelial cells (Mao et al. 2008). However, in 3T3 fibroblasts, it has been reported that extracellular HDGF does not enter the cell but instead binds to the cell membrane. Furthermore, it stimulates proliferation but does not activate the ERK signalling pathway (Abouzied et al. 2005). Our findings that there is no increase in intracellular HDGF and that extracellular HDGF has mitogenic activity without activating ERK at early time points suggest that this may also be the case in KFs. However, while the ERK pathway may not be directly activated by HDGF, it does seem to be somehow involved downstream of HDGF stimulation. In addition to the ERK pathway, we also tested the Akt pathway but were unable to detect any phosphorylation of Akt. This is consistent with the finding in human pulmonary endothelial cells, where no phosphorylation of Akt

was detected as well (Everett et al. 2004). Thus, the pathway by which HDGF exerts its proliferative activity remains as elusive as ever.

The induction of VEGF by HDGF has also been shown previously in the process of tumourigenesis (Okuda et al. 2003). In the study conducted by Okuda et al., NIH3T3 fibroblasts overexpressing HDGF were found to induce sarcomatous tumours after injection into nude mice, and the tumour formation was induced mainly by angiogenesis due to induction of VEGF. We report a similar induction of VEGF by HDGF in both keloid and normal primary skin fibroblasts, suggesting that HDGF could play some role in the normal angiogenic process during wound healing. However, in our in vitro co-culture models, we observed very little secretion of HDGF in the normal co-culture experiments. Therefore, while HDGF might induce production of VEGF in both NFs and KFs, the absence of HDGF in the normal condition limits the production of VEGF through this mechanism. This result tallies with previous findings from our group. In the study conducted by Ong et al., keloid tissue was shown to have a higher expression of VEGF compared with normal tissue, and there was also a significant increase in VEGF from keloid co-culture compared with normal co-culture conditions (Ong et al. 2007). These results suggest that the production of VEGF could be tied to the presence or absence of HDGF.

Interestingly, secretion of collagen I was observed to be downregulated when HDGF was exogenously applied to the fibroblast cells. Furthermore, there was no significant difference in α-SMA, fibronectin or CTGF expression between the HDGF induced cells and the controls. Our results suggest that HDGF induces only a mitogenic response in the fibroblasts and does not participate in their differentiation into

myofibroblasts (of which α-SMA is a marker) or in ECM (collagen I and fibronectin) production. Inhibition of CTGF has been found to cause a significant reduction in the number of myofibroblasts in scars and also decreased transcription of types I and III collagen (Sisco et al. 2008). Therefore, the lack of effect of HDGF on CTGF is consonant with this hypothesis. Furthermore, Grotendorst et al. have previously reported that fibroblast cells treated with TGF-β that are proliferating do not express α-SMA or elevated levels of collagen synthesis (Grotendorst, Rahmanie & Duncan 2004). Conversely, cells expressing α-SMA do not exhibit DNA synthesis but coexpress higher levels of types I and III collagen mRNA. The authors concluded that these responses to TGF-β are mutually exclusive and are controlled by combinatorial signaling pathways involving not only components of the TGF-β pathway, but also signaling events induced by other growth factors. HDGF appears to be one of the growth factors involved in eliciting a mitogenic response from the fibroblasts, but on its own, it may also cause the downregulation of certain ECM components. It should be noted that these features of HDGF show some similarity to the insulin-like growth factor I, a potent mitogen that requires the synergistic effect of TGF-β for ECM protein production (Daian et al. 2003; Phan et al. 2003).

The Sp1 transcription factor is known to regulate several ECM promoters, and our group has found that the Sp1 inhibitors Wp631 and mitoxanthrone were able to reduce the expression of ECM components in KF as well as to inhibit its proliferation (Mukhopadhyay et al. 2007). mTOR, on the other hand, is a serine/theronine kinase which has been shown to regulate collagen type I expression via a phosphatidylinositol 3-kinase (PI3-K)-independent pathway in human dermal fibroblasts (Shegogue &

Trojanowska 2004). The mTOR pathway inhibitor rapamycin is a naturally occurring antibiotic that has been shown to downregulate the expression of cytoplasmic PCNA, fibronectin, collagen and a-SMA (Ong et al. 2007). However, we found no significant effect of both the Sp1 inhibitors as well as the mTOR inhibitor on the production of HDGF.

Smad proteins are downstream signaling targets of the TGF-β family of growth factors (Massagué 1998; Derynck & Zhang 2003). Once activated by TGF-β receptors, Smad 2/3 oligomerizes with Smad 4, and the hetero-oligomeric Smad 2/3-Smad4 complex subsequently translocates from the cytoplasm into the nucleus where it activates collagen gene transcription (Ghosh et al. 2000). Increased Smad 3 signaling has been observed in different fibrotic disorders, including keloids (Chin et al. 2001; Phan et al. 2005). Our results show that the suppression of pro-fibrotic TGF-β mediated Smad 2/3 signaling results in an increased basal expression of cytoplasmic HDGF. This is consistent with our finding that HDGF reduces ECM protein production. We also found a significant increase in secreted HDGF from monocultured Smad 1-null cells, and a corresponding decrease in secreted HDGF from Smad 1-overexpressing cells. Very little is known about the role of Smad 1 in the biology of dermal fibroblasts. However, in lung fibroblasts, BMP4 mediated Smad 1 signaling has been found to inhibit proliferation and promote differentiation of the lung fibroblasts (Jeffery et al. 2005). This anti-proliferative effect of Smad 1 signaling could result in the downregulation of HDGF secretion. Further investigation into the role of Smad 1 signaling in dermal fibroblasts has to be done to validate this hypothesis.

Figure 4.11 summarizes the main findings of our study. First, we have shown that in the keloid condition, HDGF acts mainly as a secreted growth factor that is modulated by epithelial–mesenchymal interactions. Second, exogenous HDGF exerts a proliferative effect on KFs and is likely to be indirectly involved in ERK signalling. Finally, the presence of HDGF increases the production of VEGF and indirectly contributes to the process of angiogenesis. However, it has to be noted that on its own, HDGF appears to downregulate collagen production and has no effect on fibronectin, α-SMA and CTGF. Therefore, it is the combination of and interplay between various molecular factors that ultimately decides the fate of the wound healing process.

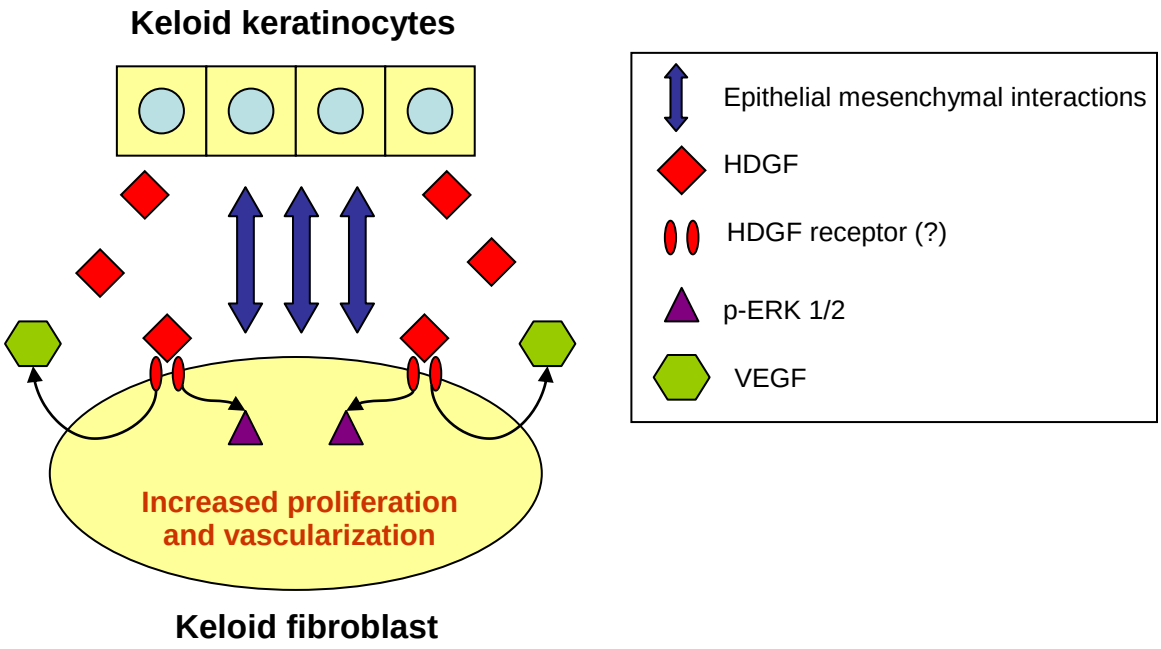


Figure 4.11: Schematic representation of the role of HDGF in keloid pathogenesis. Epithelial–mesenchymal interactions result in an increased secretion of HDGF in keloids. Overproduction of extracellular HDGF leads to the phosphorylation of ERK 1/2 and increased proliferation of keloid fibroblasts, most likely through a receptor-mediated pathway. HDGF also stimulates the fibroblasts to produce VEGF.

CHAPTER FIVE

GENOME WIDE TRANSCRIPTIONAL PROFILING OF SERUM STARVED KELOID AND NORMAL FIBROBLASTS

5.1 Introduction

Since their first reported use in the mid-1990s, microarray technology has been adopted rapidly within the research community and it is now a standard technique in a molecular biologist's toolbox. The appeal of this technology can be easily understood; with microarray technology, thousands of genes can be measured simultaneously, giving researchers a peek into the transcriptional profile of a cell. It has to be noted however, that unlike in the previous section where protein expression was measured, microarrays typically measure gene expression levels, specifically messenger RNA (mRNA) levels. mRNA has to be translated into protein before they become functional, hence requiring a further processing step that cannot be elucidated using this techonology.

The principle of a microarray experiment is that mRNA from a given cell line or tissue is used to generate a labelled sample, sometimes termed as the 'target', which is hybridized in parallel to a large number of DNA sequences immobilized on a solid surface in an ordered array (Schena et al. 1995). Although many different microarray systems have been developed by academic groups and commercial suppliers, the most commonly used systems today can be divided into two groups, according to the arrayed material: complementary DNA (cDNA) and oligonucleotide microarrays. The arrayed material has generally been termed the probe since it is equivalent to the probe used in a

northern blot analysis. Probes for cDNA arrays are usually products of the polymerase chain reaction (PCR) generated from cDNA libraries or clone collections, using either vector-specific or gene-specific primers, and are printed onto glass slides or nylon membranes as spots at defined locations. For oligonucleotide arrays, short 20–25mers are synthesized *in situ*, either by photolithography onto silicon wafers (high-density-oligonucleotide arrays from Affymetrix) or by ink-jet technology (licensed to Agilent Technologies) (Schulze & Downward 2001).

Methods based on synthetic oligonucleotides offer an advantage because sequence information alone is sufficient to generate the DNA to be arrayed, no time-consuming handling of cDNA resources is required. Another important difference between high-density oligonucleotide arrays and spotted arrays lies in the fact that the high reproducibility of *in situ* synthesis of oligonucleotide chips allows accurate comparison of signals generated by samples hybridized to separate arrays. In the case of spotted arrays, the process of gridding is not accurate enough to allow comparison between different arrays (Schulze & Downward 2001). However, oligonucleotide chips are also more expensive compared to their cDNA counterparts.

Comparing between the different oligonucelotide chips, Agilent arrays typically have a single spot per gene (single probe measurement), whereas Affymetrix arrays provide multiple measurements: a series of independent or semi-independent oligonucleotides (the probe set) query each RNA in solution. Affymetrix probe sets are constructed from a series of perfect-match and paired-mismatch oligonucleotides, allowing some assessment of non-specific binding and performance of the probes. Overall, the Affymetrix probe sets provide a variety of measurements that allow robust

measures of gene expression. The use of multiple perfect-match and mismatch probes for each gene enables the development of different methods of interpreting the hybridization patterns across the probe set and calculating a single 'expression level' or 'signal' that reflect the gene's relative expression level (Tumor Analysis Best Practices Working Group 2004). Due to these reasons, we have decided to use the Affymetrix platform for our experiments.

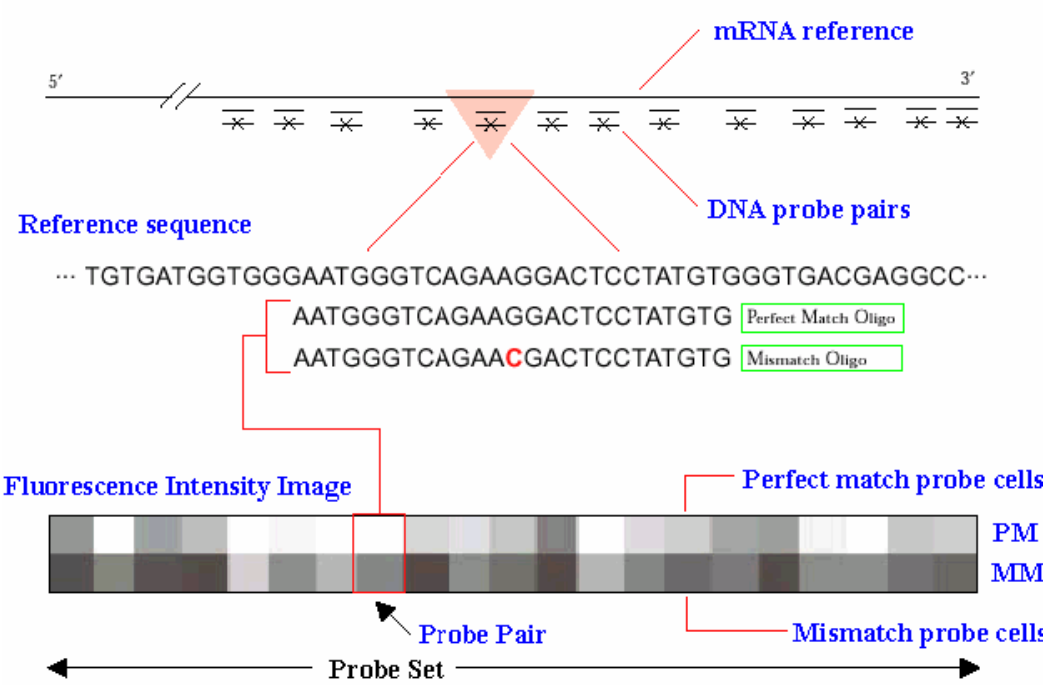


Figure 5.1: Affymetrix GeneChip Expression Array design. Each probe pair consists of one perfect match probe and one mismatch probe.

Other groups have conducted microarray experiments on keloid fibroblasts in an attempt to identify some of the transcriptional level differences underlying this condition. The most recent study examined the differential gene expression between keloid and normal fibroblasts grown in the absence and presence of the steroid hydrocortisone (Smith et al. 2008). When the fibroblasts were grown only in serum supplemented medium in the absence of hydrocortisone, 511 genes were found to be expressed at

significantly different levels in keloid and normal cells. In the presence of hydrocortisone, 515 genes were found to be differentially expressed. The study showed increased expression of several IGF-binding and IGF-binding-related proteins and decreased expression of a subset of Wnt-pathway inhibitors and multiple IL-1-inducible genes. Most genes are up- or down-regulated by hydrocortisone to a similar extent in normal and keloid cells. However, increased expression of CTGF and insulin-like growth factor binding protein (IGFBP)-3 was observed in keloid fibroblasts only in the presence of hydrocortisone, suggesting a role for glucocorticoid resistance in the pathogenesis of keloids. These findings support a role for multiple fibrosis-related pathways in the pathogenesis of keloids.

Seifert et al. performed a study comparing the gene expression profiles between different lesional sites of keloids (Seifert et al. 2008). The Affymetrix microarray chip used in this study covered 38,500 genes. Gene expression patterns in the central part of keloids involve up-regulation of apoptosis inducing genes such as a disintegrin and metalloprotease 12 (ADAM12) and ECM degrading genes as matrix metalloproteinase (MMP) - 19. Overexpression of apoptosis inhibitors such as apoptosis caspase activation inhibitor (AVEN) and down-regulation of angiogenesis inhibiting genes as pentraxin-related gene (PTX3) at the active margin of keloids may be responsible for the invasive character of the keloid margin. The results of the study support the important role of the biopsy site for research in keloids as these results show that different genes are regulated in different sites of keloids.

Another study analyzed 22,000 genes in keloid fibroblasts compared with normal skin fibroblasts using a different Affymetrix chip and revealed 43 up- and 6 down-

regulated genes (Satish et al. 2006). The authors described up-regulation of annexin A2, transgelin, and RPS18 in keloids and they reported for the first time that a few tumor-related genes were overexpressed in keloid fibroblasts. In this study, the age of the participating patients were different. Of the three patients, the first was an 8-year-old male, the second was a 57-year-old and the third was of unknown age. Furthermore, the site of the biopsy within the keloid was not presented and the race of the individuals and the reason for keloid development was not recorded.

Chen et al. performed microarray analysis of three keloids after burn injury and three normal skin samples in Chinese patients using cDNA microarray technology (Chen et al. 2003). In this study 250 genes were up and 152 genes were down-regulated. The authors describe differential expression of collagen, fibronectin, proteoglycan, growth factors, and apoptosis-related genes consistent with the published biochemical and clinical observations of keloids and found higher expression of TGF-β1 and nerve growth factor (NGF) in keloids versus normal skin.

A comparison of results from all four of these independent microarray studies was done by Seifert et al. and interestingly, no overlapping gene expression pattern was found (Seifert & Mrowietz 2009). Table 5.1 gives an overview of some of the regulated genes.

Table 5.1: Comparison of different microarray studies (Seifert & Mrowietz 2009)

	Satish [134]	Chen [135]	Smith [133]	Seifert [136]
Apoptosis				
p53 Binding protein		+		+
Brain cellular apoptosis susceptibility protein		+		+
Annexin A2	+			+
Matrix				
Proteoglycan 1, secretory granule		+		+
Chondroitin sulfate proteoglycan		+		+
Collagen type I, alpha 1 (COL1A1)	+			+
Collagen, type XI, alpha 1 (COL11A1)		+		+
Collagen, type IV, alpha 1 (COL4A1)		+		+
Collagen, type V, alpha 1 (COL5A1)		+		+
Fibronectin	+	+		+
Collagen, type VIII, alpha 1 (COL8A1)			−	+
Growth factors				
β-Actin	+			+
Insulin-like growth factor binding protein (IGFBP)	+	+	+	+
Heparin-binding EGF-like growth factor		+		+
Transforming growth factor, beta receptor II (TGF-βRII)		+		+
Transforming growth factor, beta 1 (TGF-β1)		+		+
Epidermal-growth factor receptor kinase substrate (EPS)		+		+
Human nerve growth factor (HBNF-1)		+	+	+
Cytokines				
CXCL12			−	−
Other				
Tissue factor pathway inhibitor 2 (TFPI2)			−	+
Tenascin 1 (TNC)			−	−
Syndecan 1 (SDC1)			−	−
MMP3			−	−
Tropomyosin 1	+			+
Mesoderm specific transcript (MEST)			+	+
Signal transducer and activator of transcription (STAT1)			+	+
Disc large homolog 7 (DLG7)			−	+
Coagulation factor II receptor (F2R)			+	+

+ up-regulated genes; − down-regulated genes

For our study, we wanted to examine the transcriptional differences in keloid and normal fibroblasts in the absence of external signals from serum. It is a well known fact that keloid fibroblasts have a reduced dependence on serum growth factors as compared to normal fibroblasts (Russell et al. 1988). One of the aims of this study is to elucidate some possible reasons for this phenomenon. In addition, we also wanted to examine if there were any systematic transcriptional level differences between fibroblasts that were harvested at different time points.

5.2 Results

5.2.1 The time factor did not result in any systematic differences in the transcriptional profile of the fibroblast cells

The Affymetrix GeneChip U133A array is capable of measuring expression levels of 22283 probe sets which represent approximately 18400 gene transcripts and variants. Our Two-Way ANOVA results indicate that only the type of cell (keloid or normal) resulted in significant systematic differences in gene expression levels. Both the time factor and the interactive effect between time and type of cell did not result in any significant differences in the transcriptional profile of the cells, both when MAS 5.0 was used (Table 5.2) and when RMA was used (Table 5.3)

Table 5.2: Two-way ANOVA results for determining the contribution of the time and type of cell on gene expression with probe summarization by MAS 5.0

Result Summary

	P all	P < 0.05	P < 0.02	P < 0.01	P < 0.0050	P < 0.0010
Corrected p-value(Type)	22283	395	199	132	77	46
Corrected p-value(Type-Time point)	22283	0	0	0	0	0
Corrected p-value(Time point)	22283	0	0	0	0	0
Expected by chance		19	3	1	0	0

Table 5.3: Two-way ANOVA results for determining the contribution of the time and type of cell on gene expression with probe summarization by RMA

Result Summary

	P all	P < 0.05	P < 0.02	P < 0.01	P < 0.0050	P < 0.0010
Corrected p-value(Type)	22283	330	180	141	94	45
Corrected p-value(Type-Time point)	22283	0	0	0	0	0
Corrected p-value(Time point)	22283	0	0	0	0	0
Expected by chance		16	3	1	0	0

5.2.2 Genes significantly upregulated in keloid compared to normal fibroblasts

ECM and glycoproteins such as collagen type I alpha 1 (COL1A1), collagen type XV alpha I (COL15A1), extracellular matrix protein 1 (ECM1), thrombospondin-1 (THBS1) and laminin alpha 2 (LAMA2), signaling molecules such as insulin-like growth factor binding protein 3 (IGFBP3), platelet-derived growth factor receptor beta (PDGFRB), wingless-type MMTV integration site family member 5A (WNT5A) and ras-related C3 botulinum toxin substrate 2 (RAC2), as well as transcriptional regulators such as homeobox D10 (HOXD10) and A11 (HOXA11) were found to be significantly upregulated in keloid fibroblasts compared to normal fibroblasts in at least one of the top 25 lists (Welch's t-test $P<0.05$, Tables 5.4 and 5.5). Genes found to be upregulated in both of the top 25 lists include the osteoblast specific factor periostin (POSTN), cytoskeletal protein keratin 19 (KRT19), cell adhesion molecule 1 (CADM1), neurodegenerative disorder protein ataxin-1 (ATXN1) and the axonal protein semophorin-5A (SEM5A) (Welch's t-test $P<0.05$, Tables 5.4 and 5.5). Other notable genes that were not in the top 25 list include collagen type V alpha 1 (COL5A1), collagen type V alpha 3 (COL5A3), collagen type XVII alpha 1 (COL17A1), myosins 1D (MYO1D) and 19 (MYO19), mediator of cell motility 1 (MEMO1), G-protein-coupled receptors 137B (GPR137B) and 153 (GPR153), G-protein signaling modulator 2 (GPSM2), son of sevenless homolog 2 (SOS2), growth factor receptor-bound protein 10 (GRB10) and Ephrin type-B receptor 4 (EPHB4) (Welch's t-test $P < 0.05$). The full list of genes that were upregulated with p-value < 0.05 can be found in Appendix A.1 and A.3.

Table 5.4: Top 25 upregulated genes in keloid compared to normal fibroblasts using the MAS 5.0 summarization algorithm ranked by fold change

Fold change	Gene Symbol	Gene Title	Corrected p-value
26.2568	POSTN	periostin, osteoblast specific factor	2.51E-04
19.88896	ZIC1	Zic family member 1 (odd-paired homolog, Drosophila)	0.046789
14.57531	HOXD10	homeobox D10	0.004014
10.38634	COL15A1	collagen, type XV, alpha 1	4.25E-04
8.887936	EGR2	early growth response 2 (Krox-20 homolog, Drosophila)	0.006852
7.998279	HOXA11	homeobox A11	0.033229
7.9743	CCDC102B	coiled-coil domain containing 102B	0.021884
6.770347	KCNJ6	potassium inwardly-rectifying channel, subfamily J, member 6	0.014057
6.420784	JUP /// KRT19	junction plakoglobin /// keratin 19	0.014453
5.80834	MAP7	microtubule-associated protein 7	0.00785
5.553577	IGFBP3	insulin-like growth factor binding protein 3	0.021725
5.515989	ADRA2A	adrenergic, alpha-2A-, receptor	0.011306
5.489978	RAC2	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	0.049297
5.301598	CDYL	chromodomain protein, Y-like	0.039061
5.168533	NPTX1	neuronal pentraxin I	0.015499
4.974008	PMEPA1	prostate transmembrane protein, androgen induced 1	0.037054
4.940097	EFNB2	ephrin-B2	0.022704
4.889467	ATXN1	ataxin 1	0.003232
4.783432	CADM1	cell adhesion molecule 1	0.033839
4.513727	SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	0.008288
4.476735	WNT5A	wingless-type MMTV integration site family, member 5A	0.0251
4.305683	AK5	adenylate kinase 5	0.032527
4.167339	EVI2A /// EVI2B	ecotropic viral integration site 2A /// ecotropic viral integration site 2B	0.003941
4.105316	THBS1	thrombospondin 1	0.003338
4.064728	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)	0.038054

Table 5.5: Top 25 upregulated genes in keloid compared to normal fibroblasts using the RMA summarization algorithm ranked by fold change

Fold change	Gene Symbol	Gene Title	Corrected p-value
18.0305	POSTN	periostin, osteoblast specific factor	0.006498
5.40679	IGFBP3	insulin-like growth factor binding protein 3	0.024325
3.959037	COL15A1	collagen, type XV, alpha 1	0.028774
3.21548	SEMA5A	sema domain, seven thrombospondin repeats (type 1	0.013197

		and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	
3.042947	SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	0.035137
2.726196	CADM1	cell adhesion molecule 1	0.001646
2.631936	ATXN1	ataxin 1	0.020822
2.531559	FARP1	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)	0.045664
2.375326	MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2	0.037055
2.176466	ECM1	extracellular matrix protein 1	0.047073
2.129065	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	0.019936
2.053115	KCNJ6	potassium inwardly-rectifying channel, subfamily J, member 6	0.021022
1.990826	TBC1D2	TBC1 domain family, member 2	0.048158
1.985587	CADM1	cell adhesion molecule 1	0.046985
1.959726	NXN	nucleoredoxin	0.002448
1.906954	MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2	0.017212
1.897066	COL1A1	collagen, type I, alpha 1	0.008827
1.878355	PDGFRB	platelet-derived growth factor receptor, beta polypeptide	0.002791
1.869649	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	0.046273
1.865537	GPSM2	G-protein signaling modulator 2 (AGS3-like, C. elegans)	0.04761
1.852414	LOC644191 /// LOC728937 /// RPS26	similar to hCG15685 /// similar to 40S ribosomal protein S26 /// ribosomal protein S26	0.038141
1.84265	CTSB	cathepsin B	0.018933
1.836179	ODZ3	odz, odd Oz/ten-m homolog 3 (Drosophila)	6.59E-04
1.835598	JUP /// KRT19	junction plakoglobin /// keratin 19	0.02721
1.807992	LAMA2	laminin, alpha 2	0.006566
1.796024	FHOD1	formin homology 2 domain containing 1	0.014803
1.764263	CTDSPL	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like	0.018992
1.749725	SHMT2	serine hydroxymethyltransferase 2 (mitochondrial)	0.038598
1.747925	HDLBP	high density lipoprotein binding protein	0.037741

5.2.3 Genes significantly downregulated in keloid compared to normal fibroblasts

A host of chemokine factors including chemokine ligands 6 (CXCL6), 1 (CXCL1), and 2 (CXCL2) as well as interleukin 8 (IL8) were among the genes found to be significantly

downregulated in keloid compared to normal fibroblasts in both the top 25 lists (Welch's t-test P<0.05, Tables 5.6 and 5.7). In addition, cytokines such as interleukins 6 (IL6) and 32 (IL32) as well as tumor necrosis factor alpha-induced protein 6 (TNFAIP6), 3 (TNFAIP3) and tumor necrosis factor superfamily member 10 (TNFSF10) were also found to be downregulated in at least one of the top 25 lists (Welch's t-test P<0.05, Tables 5.6 and 5.7). Other interesting genes that were downregulated include matrix metalloproteinase 2 (MMP2), hydroxysteroid (11-beta) dehydrogenase 1 (HSD11B1), complement factor B (CFB), complement component 3 (C3), radical S-adenosyl methionine domain containing 2 (RSAD2), 2',5'-oligoadenylate synthetase 1 (OAS1), solute carrier family 39 member 8 (SLC39A8), G0/G1switch 2 (G0S2), interferon-induced protein with tetratricopeptide repeats 1 (IFIT1) and 3 (IFIT3), prostaglandin E synthase (PTGES) as well as secreted frizzled-related protein 1 (SFRP1) (Welch's t-test P < 0.05) . The full list of genes that were downregulated with p-value < 0.05 can be found in Appendix A.2 and A.4.

Table 5.6: Top 25 downregulated genes in keloid compared to normal fibroblasts using the MAS 5.0 summarization algorithm ranked by fold change

Fold change	Gene Symbol	Gene Title	Corrected p-value
77.61498	CXCL6	chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)	1.48E-05
73.48209	CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	7.93E-05
67.89986	IL8	interleukin 8	0.006303
64.39615	CXCL11	chemokine (C-X-C motif) ligand 11	0.006303
49.49322	HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1	1.98E-06
41.41703	CCL5	chemokine (C-C motif) ligand 5	0.005142
39.93947	CXCL2	chemokine (C-X-C motif) ligand 2	0.002401
32.48277	RARRES1	retinoic acid receptor responder (tazarotene induced) 1	0.004841
29.77878	RSAD2	radical S-adenosyl methionine domain containing 2	0.019981
27.68656	PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)	0.001149
27.15919	C2 /// CFB	complement component 2 /// complement factor B	2.15E-05

27.04305	CXCL5	chemokine (C-X-C motif) ligand 5	2.44E-04
26.4216	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	0.008288
26.13762	CXCL3	chemokine (C-X-C motif) ligand 3	0.005384
23.80846	IL32	interleukin 32	0.001636
23.43326	CP	ceruloplasmin (ferroxidase)	0.003941
23.40176	CXCL10	chemokine (C-X-C motif) ligand 10	0.032337
21.83553	CHI3L2	chitinase 3-like 2	5.70E-04
21.53614	IDO1	indoleamine 2,3-dioxygenase 1	0.004053
19.30517	NTRK2	neurotrophic tyrosine kinase, receptor, type 2	0.011306
15.52373	C3	complement component 3	2.51E-04
15.49821	SLC39A8	solute carrier family 39 (zinc transporter), member 8	3.85E-07
14.28212	G0S2	G0/G1switch 2	0.003903
14.04543	OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa	0.039061
13.35598	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	0.016414

Table 5.7: Top 25 downregulated genes in keloid compared to normal fibroblasts using the RMA summarization algorithm ranked by fold change

Fold change	**Gene Symbol**	**Gene Title**	**Corrected p-value**
44.34992	CXCL6	chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)	9.29E-10
41.64695	CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	1.89E-05
37.03845	C2 /// CFB	complement component 2 /// complement factor B	1.75E-08
33.61232	HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1	1.50E-06
29.71772	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	0.008827
21.50683	CXCL2	chemokine (C-X-C motif) ligand 2	5.89E-05
20.34249	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	0.011388
19.02187	IL8	interleukin 8	0.003057
18.80414	SLC39A8	solute carrier family 39 (zinc transporter), member 8	1.22E-05
16.76682	SLC39A8	solute carrier family 39 (zinc transporter), member 8	7.69E-04
14.08095	C3	complement component 3	2.94E-04
13.78165	RSAD2	radical S-adenosyl methionine domain containing 2	0.013835
12.01816	SOD2	superoxide dismutase 2, mitochondrial	0.002706
11.75815	IL8	interleukin 8	0.001761
11.65778	CCL2	chemokine (C-C motif) ligand 2	9.46E-07
10.88791	SFRP1	secreted frizzled-related protein 1	0.026217
9.990352	G0S2	G0/G1switch 2	0.001019
9.679501	IFI44L	interferon-induced protein 44-like	0.04903
8.604651	CHI3L2	chitinase 3-like 2	0.019936
8.378408	IL6	interleukin 6 (interferon, beta 2)	4.59E-04
8.213031	CA12	carbonic anhydrase XII	2.40E-04
8.204023	OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa	0.015085
8.195308	GCH1	GTP cyclohydrolase 1	2.29E-04
8.068323	CA12	carbonic anhydrase XII	0.00631
7.890376	SOD2	superoxide dismutase 2, mitochondrial	0.001861

7.244113	IFIT1	interferon-induced protein with tetratricopeptide repeats 1	0.035976
7.188362	TNFAIP3	tumor necrosis factor, alpha-induced protein 3	8.07E-05
6.98734	IFIT3	interferon-induced protein with tetratricopeptide repeats 3	0.014427
6.850272	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	0.03935

5.2.4 Hierarchical clustering and principal components analysis revealed that genes chosen were capable of distinguishing between keloid and normal samples

When the 18 samples of keloid and normal fibroblasts were grouped by principal components analysis using all ~23 000 probe sets in the GeneChip U133A arrays, keloid samples were fairly well separated from normal samples (Fig. 5.2A), although there was a slight overlap when genes were summarized using the RMA algorithm (Fig. 5.3A). When samples were grouped using only probe sets that were found to be significantly different, there was a clear separation between keloid and normal samples (Figs. 5.2B and 5.3B). The same outcome was also observed when hierarchical clustering was used. When using the full list of genes, keloid samples were generally clustered together and normal samples were also generally clustered together, but again there were some samples that were not clustered accordingly (Figs. 5.2C and 5.3C). When samples were clustered using only probe sets that were found to be significantly different, two major clusters were formed with keloid samples in one cluster and normal samples in the other (Figs. 5.2D and 5.3D).

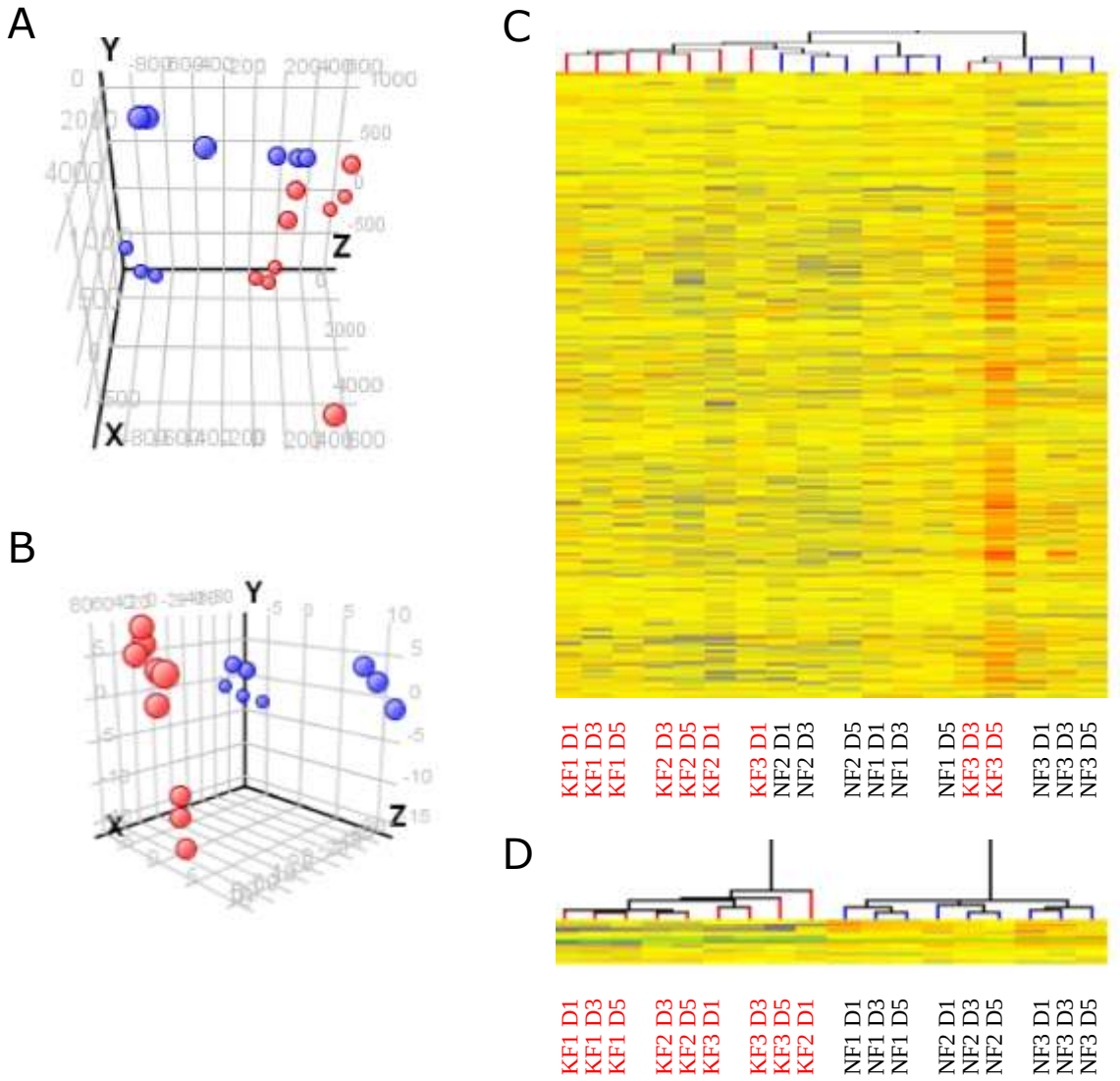


Figure 5.2: Principal components analysis and hierarchical clustering using the MAS 5.0 algorithm. (**A**) PCA of all samples using the full list of genes (**B**) PCA of all samples using differentially expressed genes (P<0.05) (**C**) Hierarchical clustering of all samples using the full list of genes (**D**) Hierarchical clustering of all samples using differentially expressed genes (P<0.05). Red balls and lines denote keloid samples while blue balls and lines denote normal samples.

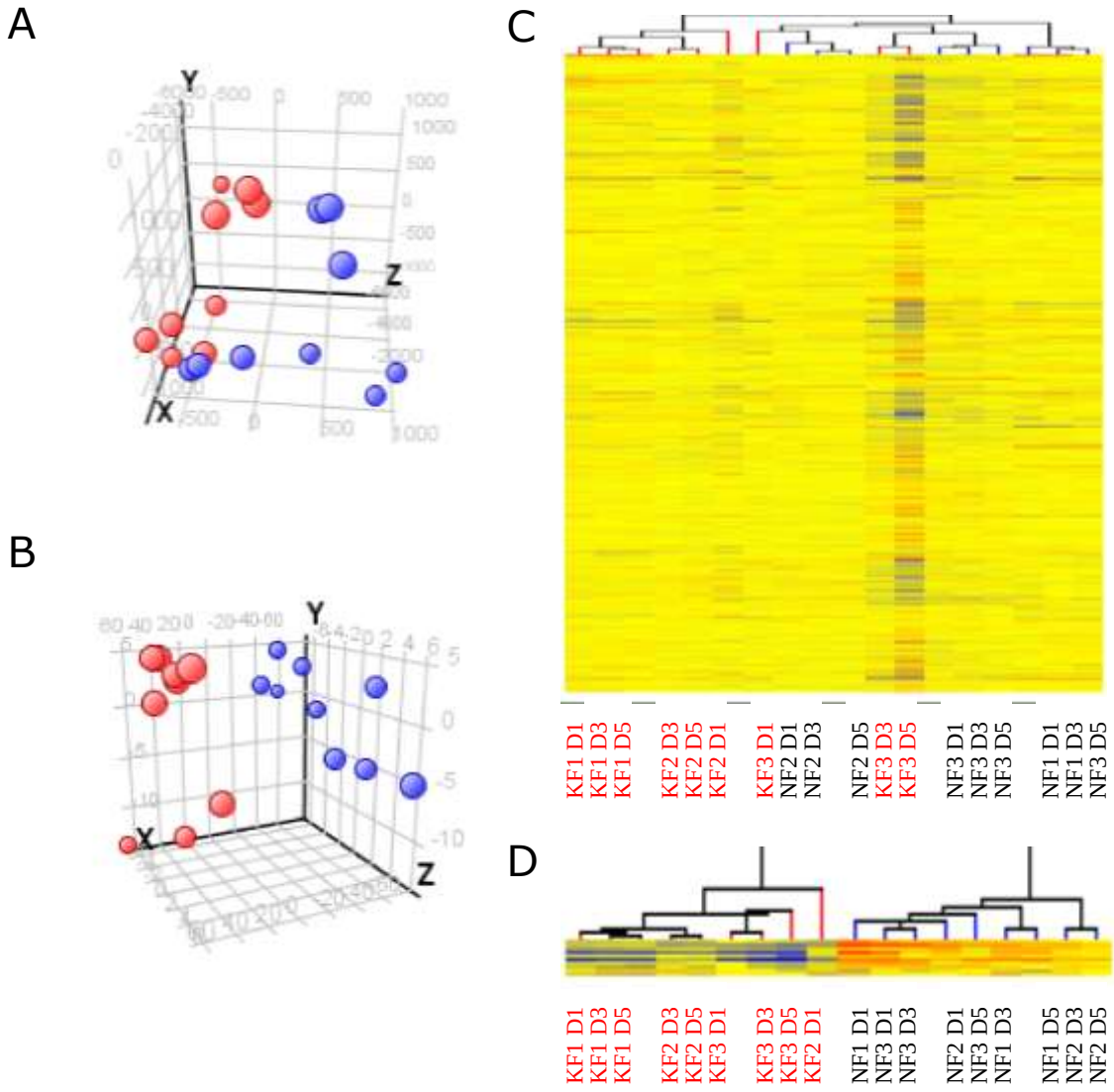


Figure 5.3: Principal components analysis and hierarchical clustering using the RMA algorithm. (**A**) PCA of all samples using the full list of genes (**B**) PCA of all samples using differentially expressed genes (P<0.05) (**C**) Hierarchical clustering of all samples using the full list of genes (**D**) Hierarchical clustering of all samples using differentially expressed genes (P<0.05). Red balls and lines denote keloid samples while blue balls and lines denote normal samples.

5.2.5 DAVID analysis suggests a role for immunological factors and ribosomal proteins in keloid pathogenesis

A total of 18 Gene Ontology (GO) terms were found to be statistically enriched (P < 0.05) using the DAVID Gene Functional Classification Tool with the list of significantly

upregulated genes in keloid as inputs (Table 5.8). When the list of significantly downregulated genes in keloid was used, a total of 35 GO terms were found to be statistically enriched ($P<0.05$) (Table 5.9). These terms include biological processes such as immune response, response to wounding and locomotory behaviour, molecular functions such as chemokine and cytokine activity and cellular components such as extracellular matrix and cytosolic ribosome. Two Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were also found to be significantly enriched ($P<0.05$) when the full set of differentially expressed genes were used. These were the antigen processing and presentation pathway (Fig. 5.4, Table 5.10) and the ribosome pathway (Fig. 5.5, Table 5.11). Most of the significantly different genes that were involved in the antigen processing and presentation pathway were downregulated in keloid while all the significantly different genes that were involved in the ribosome pathway were upregulated in keloids.

Table 5.8: List of Gene Ontology terms that were found to be statistically enriched using the DAVID Gene Functional Classification Tool with the list of significantly upregulated genes in keloid as input

Category	Term	Corrected p-value
GOTERM_CC_ALL	GO:0005830~cytosolic ribosome (sensu Eukaryota)	5.10E-07
GOTERM_CC_ALL	GO:0005840~ribosome	2.10E-06
GOTERM_CC_ALL	GO:0044445~cytosolic part	7.54E-06
GOTERM_CC_ALL	GO:0033279~ribosomal subunit	8.89E-06
GOTERM_MF_ALL	GO:0003735~structural constituent of ribosome	1.52E-05
GOTERM_MF_ALL	GO:0005198~structural molecule activity	3.88E-05
GOTERM_CC_ALL	GO:0030529~ribonucleoprotein complex	8.31E-05
GOTERM_BP_ALL	GO:0006412~translation	0.00109
GOTERM_CC_ALL	GO:0015935~small ribosomal subunit	0.001301
GOTERM_CC_ALL	GO:0005843~cytosolic small ribosomal subunit (sensu Eukaryota)	0.00132
GOTERM_CC_ALL	GO:0043232~intracellular non-membrane-bound organelle	0.001642
GOTERM_CC_ALL	GO:0043228~non-membrane-bound organelle	0.001642
GOTERM_MF_ALL	GO:0003723~RNA binding	0.029371

GOTERM_CC_ALL	GO:0031012~extracellular matrix	0.030531
GOTERM_CC_ALL	GO:0005737~cytoplasm	0.0307
GOTERM_CC_ALL	GO:0005578~proteinaceous extracellular matrix	0.031002
GOTERM_BP_ALL	GO:0009059~macromolecule biosynthetic process	0.042621
GOTERM_BP_ALL	GO:0044249~cellular biosynthetic process	0.04839

Table 5.9: List of Gene Ontology terms that were found to be statistically enriched using the DAVID Gene Functional Classification Tool with the list of significantly downregulated genes in keloid as input

Category	**Term**	**Corrected p-value**
GOTERM_BP_ALL	GO:0006955~immune response	2.22E-21
GOTERM_BP_ALL	GO:0002376~immune system process	3.44E-19
GOTERM_BP_ALL	GO:0050896~response to stimulus	9.19E-14
GOTERM_BP_ALL	GO:0019882~antigen processing and presentation	2.09E-09
GOTERM_BP_ALL	GO:0006952~defense response	2.10E-09
GOTERM_MF_ALL	GO:0005125~cytokine activity	2.35E-08
GOTERM_MF_ALL	GO:0042379~chemokine receptor binding	2.42E-08
GOTERM_MF_ALL	GO:0008009~chemokine activity	3.61E-08
GOTERM_BP_ALL	GO:0006954~inflammatory response	5.17E-08
GOTERM_MF_ALL	GO:0001664~G-protein-coupled receptor binding	1.23E-06
GOTERM_BP_ALL	GO:0048002~antigen processing and presentation of peptide antigen	2.09E-06
GOTERM_BP_ALL	GO:0009611~response to wounding	2.93E-06
GOTERM_CC_ALL	GO:0005615~extracellular space	5.60E-06
GOTERM_BP_ALL	GO:0002474~antigen processing and presentation of peptide antigen via MHC class I	7.17E-06
GOTERM_BP_ALL	GO:0009607~response to biotic stimulus	2.44E-05
GOTERM_BP_ALL	GO:0042221~response to chemical stimulus	5.25E-05
GOTERM_CC_ALL	GO:0042611~MHC protein complex	6.93E-05
GOTERM_BP_ALL	GO:0007626~locomotory behavior	7.96E-05
GOTERM_CC_ALL	GO:0005576~extracellular region	1.16E-04
GOTERM_CC_ALL	GO:0044421~extracellular region part	2.63E-04
GOTERM_BP_ALL	GO:0006950~response to stress	3.39E-04
GOTERM_BP_ALL	GO:0009605~response to external stimulus	3.81E-04
GOTERM_BP_ALL	GO:0042330~taxis	3.93E-04
GOTERM_BP_ALL	GO:0006935~chemotaxis	3.93E-04
GOTERM_BP_ALL	GO:0051707~response to other organism	3.97E-04
GOTERM_BP_ALL	GO:0051704~multi-organism process	5.05E-04
GOTERM_CC_ALL	GO:0042612~MHC class I protein complex	6.51E-04
GOTERM_BP_ALL	GO:0009615~response to virus	0.001549
GOTERM_BP_ALL	GO:0000041~transition metal ion transport	0.00606
GOTERM_MF_ALL	GO:0005507~copper ion binding	0.008833
GOTERM_MF_ALL	GO:0005102~receptor binding	0.010917
GOTERM_BP_ALL	GO:0007610~behavior	0.011518
GOTERM_MF_ALL	GO:0046870~cadmium ion binding	0.023054

GOTERM_BP_ALL	GO:0042127~regulation of cell proliferation	0.037833

Table 5.10: List of downregulated genes in keloid compared to normal fibroblasts involved in GO term Antigen Processing and Presentation

Fold change	Gene Symbol	Gene Title
2.737904	HLA-F	major histocompatibility complex, class I, F
2.702095	HLA-C	major histocompatibility complex, class I, C
2.648943	HLA-DMA	major histocompatibility complex, class II, DM alpha
2.638515	TAPBPL	TAP binding protein-like
2.566287	TAPBPL	TAP binding protein-like
2.421571	HLA-F	major histocompatibility complex, class I, F
2.231546	HLA-G	major histocompatibility complex, class I, G
2.21116	HLA-G	major histocompatibility complex, class I, G
2.182933	HLA-B	major histocompatibility complex, class I, B
2.131292	HLA-G	major histocompatibility complex, class I, G
2.117058	TAPBP	TAP binding protein (tapasin)
2.042615	HLA-B	major histocompatibility complex, class I, B
1.798085	HLA-C	major histocompatibility complex, class I, C
1.795796	HLA-A /// HLA-A29.1 /// HLA-B /// HLA-G /// HLA-H /// HLA-J	major histocompatibility complex, class I, A /// major histocompatibility complex class I HLA-A29.1 /// major histocompatibility complex, class I, B /// major histocompatibility complex, class I, G /// major histocompatibility complex, class I, H (pseudogene) /// major histocompatibility complex, class I, J (pseudogene)
1.773213	HLA-C	major histocompatibility complex, class I, C
1.691625	HLA-C	major histocompatibility complex, class I, C
1.678448	HLA-B /// MICA	major histocompatibility complex, class I, B /// MHC class I polypeptide-related sequence A
1.520112	HLA-A	major histocompatibility complex, class I, A
1.229484	B2M	beta-2-microglobulin

Table 5.11: List of upregulated genes in keloid compared to normal fibroblasts involved in GO term Ribosome

Fold change	Gene Symbol	Gene Title
2.662594	MGC87895 /// RPS14	similar to ribosomal protein S14 /// ribosomal protein S14
1.877402	LOC285053 /// LOC390354 /// RPL18A	similar to ribosomal protein L18a /// ribosomal protein L18a pseudogene /// ribosomal protein L18a
1.865347	LOC644191 /// LOC728937 /// RPS26	similar to hCG15685 /// similar to 40S ribosomal protein S26 /// ribosomal protein S26
1.780408	RPL10	ribosomal protein L10
1.764005	RPS16	ribosomal protein S16
1.696983	EIF3A	eukaryotic translation initiation factor 3, subunit A
1.645023	RPS9	ribosomal protein S9

1.636263	RPL13	ribosomal protein L13
1.602853	RPS2	ribosomal protein S2
1.597647	RPL4	ribosomal protein L4
1.59623	RPS8	ribosomal protein S8
1.531203	SERP1	stress-associated endoplasmic reticulum protein 1
1.523152	RPL13	ribosomal protein L13
1.474366	RPS6	ribosomal protein S6
1.415472	RPL8	ribosomal protein L8
1.371258	SERP1	stress-associated endoplasmic reticulum protein 1
1.360637	RPL13	ribosomal protein L13

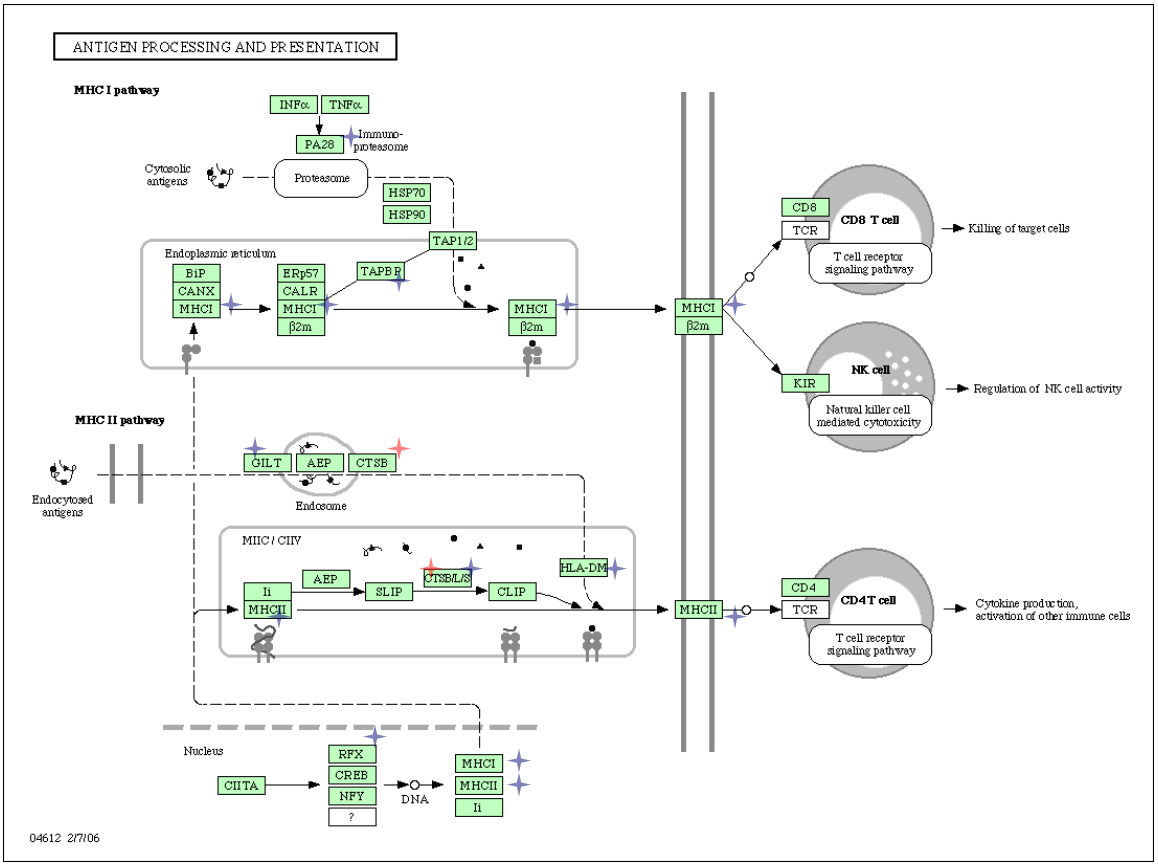


Figure 5.4: Antigen processing and presentation pathway from the KEGG database. List of differentially expressed genes were submitted to the DAVID Gene Functional Classification Tool using the full list of Affymetrix U133A genes as background for statistical analysis. The antigen processing and presentation pathway had a Benjamini corrected P-value of 0.0029. Genes that were significantly upregulated in the pathway are denoted with a red cross while genes that were significantly downregulated in the pathway are denoted with a blue cross.

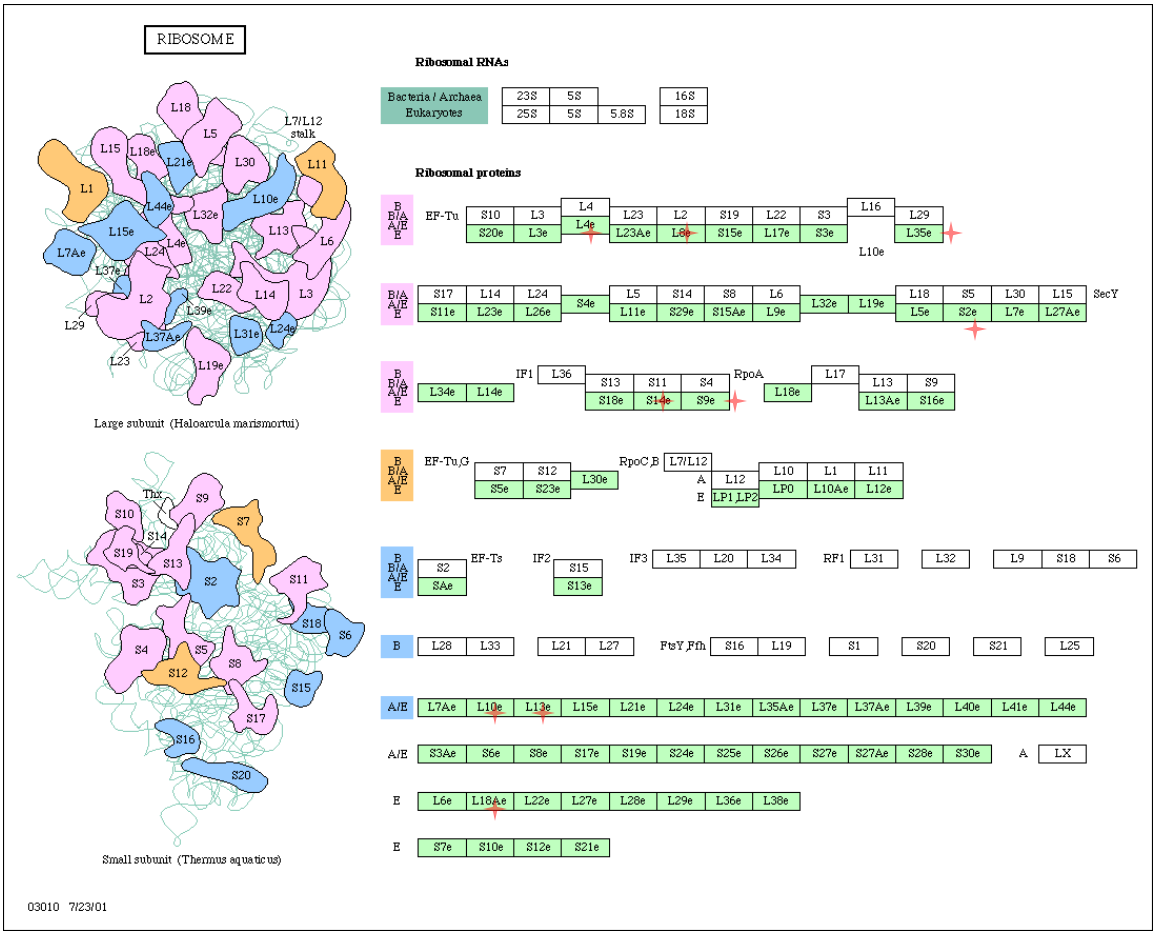


Figure 5.5: Ribosome pathway from the KEGG database. List of differentially expressed genes were submitted to the DAVID Gene Functional Classification Tool using the full list of Affymetrix U133A genes as background for statistical analysis. The ribosome pathway had a Benjamini corrected P-value of 0.029. Genes that were significantly upregulated in the pathway are denoted with a red cross while genes that were significantly downregulated in the pathway are denoted with a blue cross.

5.3 Discussion

A number of different groups have performed microarray studies on keloid fibroblasts previously (Chen et al. 2003; Satish et al. 2006; Seifert et al. 2008; Smith et al. 2008) but none have examined the transcriptional level and time dependent effects of serum starvation on these cells. In this study, we report that the time factor did not have any

significant and systematic effect on gene expression levels of keloid and normal fibroblasts. This result is consistent with the common assumption that cells left in minimal media show no major transcriptional level differences. However, it is also possible that our time points and number of samples are too small to detect any systematic differences due to the time factor. Under serum starvation conditions, it was found that there was a greater number of genes downregulated and fewer genes upregulated in the keloid compared to the normal fibroblasts. Interestingly, this same observation has also been made when the fibroblasts were left in serum supplemented media (175 genes upregulated, 559 genes downregulated) or when hydrocortisone supplemented media was used (221 genes upregulated, 547 genes downregulated) (Smith et al. 2008). Research in keloid has focused mainly on the upregulation of candidate genes such as TGF-β and PDGF, but these results suggest that gene downregulation could also be a very important aspect of keloid formation.

The microarray platform allows for the measurement of thousands of gene transcripts simultaneously, but its strength could also be its weakness. It is not feasible, for example, to independently validate all the hundreds of abnormally expressed genes using other molecular techniques. Furthermore, the vast amounts of data combined with the limited number of samples make for a profoundly under-determined problem; there is not enough data to distinguish between many of the different hypothesis that could be consistent with the data set. Finally, the different methods that can be used for background correction, normalization and summarization of the arrays can lead to different results. Some of the more widely used algorithms include the default Affymetrix MAS 5.0 algorithm (default Affymetrix approach; Affymetrix Users Guide,

www.affymetrix.com), the RMA algorithm (Bolstad et al. 2003) and the dChip algorithm (Li & Wong 2001). At present, there is no consensus as to which method is best, with different studies giving conflicting results (Bolstad et al. 2003; Shedden et al. 2005; Harr & Schlötterer 2006; Lim et al. 2007).

In our experiments, we have used the MAS 5.0 algorithm and the RMA algorithm for normalization and summarization. We found that using the MAS 5.0 algorithm gives a larger number of differentially expressed genes (471 genes, $P < 0.05$) compared to the RMA algorithm (344 genes, $P < 0.05$). Furthermore, the list of genes that were most highly upregulated or downregulated were also different when a different algorithm was used. In total, the intersection between the RMA summarized data and the MAS 5.0 summarized data showed 217 genes to be differentially expressed (Appendix A.5). It remains unclear which method produces the more accurate and reliable results, but we have chosen to use the MAS 5.0 summarization algorithm for DAVID analysis by virtue it producing a larger number of differentially expressed genes. For enrichment analysis, a larger gene list has higher statistical power resulting in a higher sensitivity to slightly enriched terms, as well as to more specific terms (Huang, Sherman & Lempicki 2009).

Many of the genes that were found to be significantly different in this study were similar to those that have been found in other studies previously. ECM proteins such as COL1A1 and COL5A1 were found to be significantly upregulated in keloids. Upregulation of collagen is a well known characteristic of keloid fibroblasts and upregulation of these two collagen types in particular has also been reported in previous microarray studies (Table 5.1). PDGFRB (Messadi et al. 1998), THBS1 (Chipev et al. 2000) and RAC2 (Witt et al. 2008) were also found to be significantly upregulated in

keloids consistent with previous findings. In a microarray array study done by Seifert et al., it was reported that keratin 18 (KRT18) was upregulated in keloid compared to normal samples, and that there was upregulation of POSTN and ECM1 in fibroblasts derived from the deeper part of the keloid compared to the superficial part of the keloid (Seifert et al. 2008). While we did not see any difference in KRT18 expression levels, we found a higher expression of KRT19, POSTN and ECM1 in keloid compared to normal fibroblasts.

IGFBP3 is another important upregulated factor that was found in our study. Higher expression of IGFBP3 in dermal fibroblasts from patients with systemic sclerosis (Feghali & Wright 1999), idiopathic pulmonary fibrosis (Yasuoka et al. 2006), and leiomyomas (Tsibris et al. 2002) indicate that this protein has some role to play in fibrosis. Furthermore, IGFBP3 is capable of inducing production of ECM components such as collagen type I and fibronectin in normal primary adult lung fibroblasts (Pilewski et al. 2005). A known function of IGFBP3 is to bind to IGF-1 (Collett-Solberg & Cohen 1996) but it has also been found to bind to extracellular matrix components, have nuclear localization signals, and bind to putative receptors on the cell surface (Mohan & Baylink 2002). It is possible that these other functions of IGFBP3 contribute to its role in fibrosis and could be an interesting future area of study. Increased expression of IGFBP3 has also been found in other studies of keloids (Satish et al. 2006; Smith et al. 2008).

Our results also show a reduction in the Wnt signaling antagonist SFRP1 as well as an increase in the WNT5A gene expression levels in keloid compared to normal fibroblasts. Low levels of SFRP1 in keloid fibroblasts have been reported previously (Smith et al. 2008) indicating a role for the Wnt pathway in keloid pathogenesis. SFRP1

suppresion has also been found in numerous different carcinomas including breast (Shulewitz et al. 2006), bladder (Marsit et al. 2005), ovarian (Takada et al. 2004), colon (Caldwell et al. 2004) and prostrate (Lodygin et al. 2005) cancers. Furthermore, it has also been shown to have an antiproliferative effect on vascular cells (Ezan et al. 2004), making it a very attractive candidate for further study. SFRP1 suppresses both the canonical Wnt/β-catenin pathway, as well as the non canonical β-catenin independent pathway (Yang et al. 2009). WNT5A, on the other hand, is only associated with the non-canonical Wnt signaling pathway (Slusarski, Corces & Moon 1997; Kilian et al. 2003). The role of WNT5A in cancer remains unclear with some cancers showing an upregulation of this factor while others show a decrease in expression (McDonald & Silver 2009). However, it has been found to promote angiogenesis by inducing endothelial cell proliferation and enhanced cell survival under serum-deprived conditions (Masckauchán et al. 2006). While the canonical Wnt signaling has been shown to be upregulated in keloids (Sato 2006), not much is known of the non-canonical pathway.

Expression of some IL-1 responsive genes such as CXCL-1, CXCL-6, MMP-2 and TNFAIP6 were found to be downregulated in keloids and again, this result is consonant with that obtained in previous studies (Smith et al. 2008; Yeh, Shen & Tai 2009). Lower levels of MMP-2 could be one of the reasons for the accumulation of ECM components as the MMPs are responsible for ECM degradation. Other inflammatory cytokines such as IL8 and IL32 were also found to be downregulated, suggesting a role for inflammation in keloid pathogenesis.

Other interesting genes that were upregulated include those involved in G protein coupled receptor signaling (GPR137B, GPR153, SOS2), cell motility proteins (MYO1D,

MYO19, MEMO1) and tyrosine kinase signaling molecules (EPHB4, SOS2) while those that were downregulated were involved in apoptosis (PTGES, TNFSF10), production of cortisol (HSD11B1), activation of the complement pathway of the immune system (CFB, C3), antiviral defense (RSAD2, OAS1) and interferon induced proteins (IFIT1 and IFIT3). These genes, as well as a host of other genes in the top 25 list of differentially expressed genes could be important targets for further study and may be important in shedding light on the keloid condition. Of particular interest are the proapoptotic genes TNFSF10 and PTGES which were found to be decreased in keloids. This finding has never been reported before and could explain the resilience of keloids even under serum starvation conditions.

However, there were some discrepancies in our results compared to previous studies as well. Smith et al. reported a reduction in multiple homeotic (HOX) genes such as HOXA11, among others (Smith et al. 2008). We found an increased expression of HOXA11 and HOXD10 in keloid compared to normal fibroblasts. HOX genes are highly conserved master control genes that play major roles in anterior-posterior development in the embryo but they have also been found to regulate gene expression in adult differentiated cells, including human dermal fibroblasts (Chang et al. 2002). The differential expression of HOX genes may account for the tumorigenic phenotype of keloids, but as suggested by Smith et al., it is also possible that these differences may just be due to the different anatomic sites from which cultures were isolated (Smith et al. 2008). This could possibly account for the discrepancy in our results. In addition, these differences could also be due to the different culture conditions (serum/hydrocortisone instead of serum starvation), different microarray chips (U133A 2.0 instead of U133A),

different normalization methods (RMA instead of MAS 5.0), sample to sample variation or false positive results. Our results also showed a downregulation of IL6, contrary to previous reports where higher expression of IL6 was found in keloids (Tosa et al. 2005; Ghazizadeh et al. 2007). Overexpression of IL6 related genes in keloid fibroblasts was also not seen in the microarray study done by Smith et al. and they attributed this to the differences in origin of cultures, culture conditions, or the microarray platform (Smith et al. 2008). The very same factors could also be the reasons for our discrepancy.

In general, there are two ways in which microarrays can be used to investigate problems in cell biology (Schulze & Downward 2001). The first method can be thought of as a local approach to understanding gene expression changes, where the investigator is interested only in finding the single change in gene expression that might be the key to a given alteration in phenotype. This is the method that we have utilized so far, and in doing so we have come up with a list of potential genes that could be important in understanding keloid pathophysiology. However, we are never really certain about the importance of this list; microarray experiments are highly capable of generating long lists of genes with altered expression, but they provide few clues as to which of these changes are important in establishing a given phenotype. This deduction is left to the ingenuity of the experimenter, and the temptation is to stick to familiar genes, or genes that conform to existing ideas about how the system works, thus resulting in a certain level of biasedness. In our case, we have tried to eliminate this bias by giving the top 25 lists of genes that were differentially expressed with p-values less than 0.05, but even so, the list differs significantly depending on the normalization method used.

Another way of working with microarray data is to look at the global picture of gene expression patterns. Unsupervised methods such as *k*-means clustering, principal-component analysis and self-organizing maps can be used to group closely related genes or samples together. Using hierarchical clustering and principal-component analysis on the full set of genes, we found that in general, keloid samples were more similar to each other compared to normal samples, regardless of the sample type or the time point when they were harvested. This was despite the fact that the samples were processed in batches, that is, RNA extraction and hybridization of KF1 and NF1 were done together, KF2 was paired with NF2 and KF3 was paired with NF3. Batch effects are a very common problem faced by researchers in the area of microarray studies, particularly when combining multiple batches of data from different experiments or if an experiment cannot be conducted all at once (Johnson, Li & Rabinovic 2007). Our results indicate that for our case, there are minimal systematic batch effects. There was a small degree of overlap between the keloid and normal samples when the full set of genes was used, but this could also be due to the fact that KF3 and NF3 were from the same patient. This overlap was eliminated when the subset containing only genes found to be differentially expressed was used for the unsupervised methods. This result reconfirms the ability of the list of genes found to discriminate between keloid and normal fibroblasts.

We were also interested in looking at clusters of related genes that were differentially expressed to give some biological meaning to the data. Hierarchical clustering can be used to cluster similarly expressed genes together, but here we have opted for an enrichment analysis based method instead. The strategy employed here is to systematically map a large number of interesting genes in a list to the associated

biological annotation (e.g., gene ontology terms), and then statistically highlight the most overrepresented (enriched) biological annotation out of thousands of linked terms and contents. The DAVID Gene Functional Classification Tool is one of a number of tools that is capable of performing enrichment analysis, and compared with similar services, it has some unique features and capabilities, such as an integrated and expanded back-end annotation database, advanced modular enrichment algorithms and powerful exploratory ability in an integrated data-mining environment (Huang, Sherman & Lempicki 2009).

For DAVID analysis, we have decided to use the MAS 5.0 algorithm for summarization and normalization as this produced a larger set of differentially expressed genes as compared to the RMA algorithm. In general, when it comes to enrichment analysis, a larger gene list can have higher statistical power resulting in a higher sensitivity to slightly enriched terms, as well as to more specific terms. Otherwise, the sensitivity is decreased toward largely enriched terms and broader/general terms (Huang, Sherman & Lempicki 2009). Huang et al. also gives a checklist of characteristics of a 'good' gene list for analysis, and this includes the presence of important marker genes, a reasonable number of genes ranging from hundreds to thousands and the passing of important statistical thresholds such as t-tests and fold changes. However, he also states that important statistical thresholds do not have to be sacrificed (e.g., fold changes>1.1 and P-value < 0.2) to reach a comfortable gene size. For our gene list, we have chosen a stringent P-value of 0.05 but have not placed any restriction on the fold change as we believe the stringent P-value would be sufficient in identifying the differentially expressed genes.

One deficiency of the DAVID system is that it does not say in what way the enriched terms differ when comparing between keloid and normal fibroblast cells, that is to say, whether for example, immune response is heightened or suppressed in keloid compared to normal cells. All the system does is to take in a list of genes and determine which terms are enriched from the list that was received. One way of dealing with this limitation is to input genes that are upregulated and genes that are downregulated separately. Since this approach reduces the size of the gene list, this may result in decreased sensitivity and specificity in the enrichment analysis. However, we have found that in our case, breaking up the list into upregulated and downregulated genes results in about the same number and type of terms enriched compared to when the full set of differentially expressed genes was used as input. Furthermore, this approach has the added advantage of giving us a rough gauge as to how the enriched terms differ when comparing keloid and normal cells. However, bearing in mind the multiple and complex roles that a gene could play in a biological system, and also the antagonistic roles of some genes, this approach would at best give us a very rough idea about the roles that the enriched terms play in the keloid condition. For a more accurate treatment, detailed literature review and further experiments have to be done on each of the enriched terms found.

When comparing keloid to normal fibroblasts, our results indicate an increased expression of genes involved in ribosomal activity and extracellular matrix and a decreased expression of genes involved in immune response, locomotory behaviour and chemokine and cytokine activity, among others. This is an interesting finding that could be explored in detail in future studies. We also found two KEGG pathways that were

significantly enriched when the full list of differentially expressed genes was used. These pathways are the antigen processing and presentation pathway and the ribosome pathway.

Although fibroblasts are frequently considered just as structural elements of a tissue, they represent in fact a dynamic cell population that is actively involved not only in tissue remodeling, but also in autoimmune and inflammatory processes (Smith 2005). Human leukocyte antigen (HLA) class I molecules are expressed on the surface of all nucleated human cells including fibroblasts and their role is to present peptides derived from endogenous proteins to cytotoxic $CD8^+$ T cells(York & Rock 1996; Grommé & Neefjes 2002; Cresswell et al. 2005). The reduced levels of these class I molecules implies that keloid fibroblasts are unable to deal with infections, and this result appears to be consistent with other findings from this study about the lower levels of immunological factors in keloids. Interferon gamma has been used in the treatment of keloids, and its efficacy has been attributed to its ability to downregulate collagen synthesis (Larrabee et al. 1990). Interferon gamma has been known to also increase HLA class I expression in dermal fibroblasts (Hengel et al. 1995; Zimmer et al. 2006) and this could possibly be another reason for its success in treating keloids.

It has been previously reported that ribosomal protein (RPS) L23A, RPS10 and RPS 18 were upregulated in keloids (Satish et al. 2006). Here we find further support for the upregulation of ribosomal proteins in keloids. Overexpression of several ribosomal proteins has been reported in carcinomas of the colon (Pogue-Geile et al. 1991) and breast (Henry, Coggin & King 1993). There is also evidence suggesting that ribosomal proteins, in addition to participating in protein synthesis, are likely to be involved in other extraribosomal functions such as DNA replication, transcription and repair, RNA splicing

and modification, cell growth and proliferation, regulation of apoptosis and development, and cellular transformation (Wool 1996; Lai & Xu 2007). All these make the ribosomal proteins interesting candidates for study as they may shed further light on the keloid condition.

In conclusion, this study has shown that many of the genes found to be differentially expressed in keloid fibroblasts when left in serum starvation media are similar to what has been reported in previous studies, despite the different conditions used in the other studies. Furthermore, we found no systematic difference when the cells were harvested at different time points. This study has also revealed a list of novel differentially expressed genes that could be utilized for further research, out of which the antigen presentation pathway and ribosomal proteins appear to be interesting candidates.

CHAPTER SIX

REVERSE ENGINEERING GENE NETWORKS IN KELOID AND NORMAL FIBROBLASTS

6.1 Introduction

The genome plays a central role in the control of cellular processes, such as in the response of cells to environmental signals, the differentiation of cells during development, and the replication of the genome preceding cell division. A protein synthesized from the information contained in a coding region of DNA may function as a transcription factor binding to regulatory sites elsewhere on the DNA, as an enzyme catalyzing a metabolic reaction, or as a component of a signal transduction pathway. With few exceptions, all cells in an organism contain the same genetic material. This implies that in order to understand how genes are implicated in the control of intracellular and intercellular processes, the scope should be broadened from sequences of nucleotides encoding proteins to regulatory systems determining which genes are expressed, when and where in the organism, and to what extent. The goal of reverse engineering methods is to infer gene networks from observational data, thus providing insight into the inner workings of a cell (Hartwell et al. 1999; Schadt, Sachs & Friend 2005).

Reverse engineering strategies generally fall within two broad categories - “physical” approaches and “influence” approaches (Gardner & Faith 2010). The modeling of the physical interactions between transcription factors and their promoters is what is known as the physical approach to reverse engineering. Gene expression is

predominantly controlled by auxiliary proteins called transcription factors (TF). A TF binds directly to a specific upstream region of the target gene known as the promoter region, which triggers the enzyme, RNA polymerase, to transcribe DNA to RNA. TFs can therefore be viewed as a class of specialized proteins that govern the on-off switch of gene expression through either repressing (down-regulation) or inducing (up-regulation) its output. An advantage of this strategy is that it enables the use of genome sequence data, in combination with RNA expression data, to enhance the sensitivity and specificity of predicted interactions but its limitation is that it cannot describe regulatory control by mechanisms other than transcription factors.

On the other hand, the influence approach abstracts out this mechanistic process and instead can be viewed just as an input-output device. In other words, it looks for transcripts that act as "inputs" whose concentration changes can explain the changes in "output" transcripts. Such a model does not generally describe physical interactions since transcription is rarely controlled directly by RNA. Nevertheless, in some cases, the input transcripts may encode the transcription factors that directly regulate transcription. In such cases, the influence model may accurately reflect a physical interaction. An advantage of the influence strategy is that the model can implicitly capture regulatory mechanisms at the protein and metabolite level that are not physically measured but the limitation of this approach is that the model can be difficult to interpret in terms of the physical structure of the cell, and therefore difficult to integrate or extend with further research. Moreover, the implicit description of hidden regulatory factors may lead to prediction errors.

In addition to these two modeling approaches, reverse engineering methods also differ in terms of the mathematical formalisms used and can be static or dynamic, continuous or discrete, linear or nonlinear and deterministic or stochastic (Hache, Lehrach & Herwig 2009). For the purposes of this study, we have chosen to use both the physical as well as the influence approach for reconstructing the networks. For the physical approach, we will use the regression method fREDUCE (fast-Regulatory Element Detection Using Correlation with Expression) (Wu et al. 2007) with the objective of identifying important cis-binding motifs and their targets in keloid fibroblasts. For the influence approach, we will compare the performance of the information theoretic method ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) (Basso et al. 2005) and the Bayesian package BANJO (Bayesian Network Inference with Java Objects) (Yu et al. 2004) in uncovering regulatory interactions in keloid and normal fibroblasts. The effect of different normalization/summarization methods and lowly expressed probes on gene network inference is also not clear and will be examined in this system.

Generally, the process of modeling gene regulatory networks consists of a few main steps: designing experiments that produce maximally informative observations, developing methodologies for choosing a candidate model that 'best' fits the observations, analyzing and validating the model, and using the model to formulate and test new hypotheses (Goutsias & Lee 2007). Microarray data from the previous study will be used to learn the networks. However, learning the structure of a gene network using the influence approach is difficult as the number of possibilities scale exponentially with the number of variables. Therefore, modeling and testing such large structures would

require large amounts of data for accuracy. Due to our limited data, we have decided to focus on small networks of genes that have been found to be differentially expressed from the second part of this dissertation. Furthermore, to increase the number of samples, we will use data from other independent microarray experiments performed in our lab and also data from Smith et al (2008), which is the only keloid fibroblast data publicly available at the Gene Expression Omnibus (GEO) database. For the physical approach, since the binding motif repeats are regressed against the expression levels of each gene, it is the number of genes that constitute the sample size. Therefore, the full range of genes is used for this approach instead of the smaller transcriptional networks that have found to be differentially expressed.

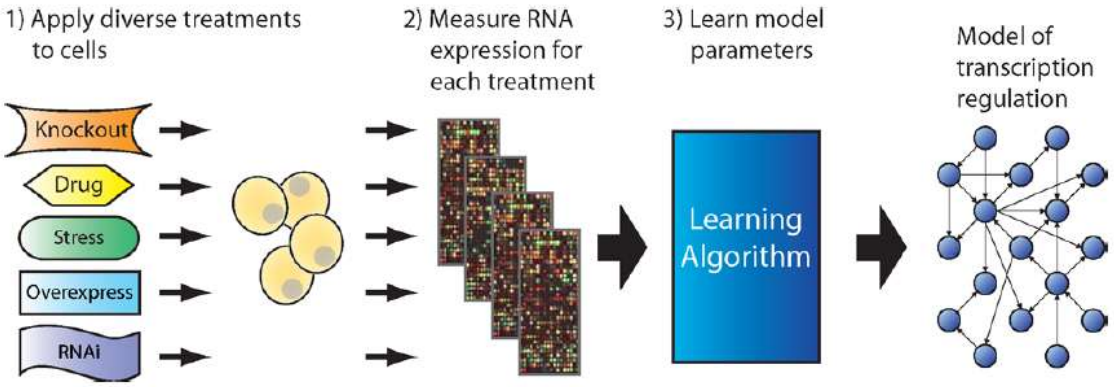


Figure 6.1: The general strategy for reverse-engineering transcription control systems (Gardner & Faith 2010). (1) The experimenter perturbs cells with various treatments to elicit distinct responses. (2) After each perturbation, the experimenter measures the expression (concentration) of many or all RNA transcripts in the cells. (3) A learning algorithm calculates the parameters of a model that describes the transcription control system underlying the observed responses. The resulting model may then be used in the analysis and prediction of the control system function.

In total, we have four different treatment conditions (serum-treated, serum-free, hydrocortisone-treated and HDGF-treated) and two different cell derivations (keloid and normal) from multiple patients. Although our datasets consist of some time-series, the gap between each time point is very large (in the order of days) and may lead to

inaccurate results if used to infer time-series regulatory networks. Therefore, we have limited our study to steady state conditions with the assumption that each time point is statistically independent from others. This is a possibly valid assumption as the sampling time is very long. Furthermore, the genes were not directly perturbed by knockdown or overexpression in our experiments and it is very likely that the different conditions used will result in multiple unknown perturbations. As such, inference algorithms such as dynamic Bayesian networks (which require numerous closely spaced time points) and differential equation approaches (which require either time series data or knowledge of perturbations) cannot be applied in our case.

6.2 Algorithms

6.2.1 fREDUCE

This method is an extension to the REDUCE (Regulatory Element Detection Using Correlation with Expression) algorithm (Bussemaker, Li & Siggia 2001). REDUCE is a deterministic method that first enumerates oligonucleotides and then identifies words whose occurrence in promoter sequences correlate most strongly with expression data. This procedure is applied iteratively to produce a set of oligonucleotides that produce the best simultaneous fit to the data.

One weakness of REDUCE is that it can miss weak but biologically significant variants of the regulator site. Highly degenerate motifs whose individual variants fall below the detection threshold will be missed altogether. This is particularly the case for

regulators in higher mammalian genomes, which can exhibit strong site to site variation in specificity. However, a straightforward extension of REDUCE using exhaustive enumeration of degenerate motifs becomes impractical when the motif length or number of degenerate positions increases as the computation of the Pearson correlation coefficient which is required for identifying the motifs is computationally laborious.

fREDUCE uses the following strategy to efficiently compute the Pearson coefficients of the most significant degenerate motifs: 1) A list of degenerate motifs that can be derived from the sequence data is generated. 2) For each degenerate motif, a "pseudo-Pearson" coefficient, an estimate of the actual Pearson coefficient can be calculated. The pseudo-Pearson coefficient is guaranteed to be an upper-bound on the actual Pearson coefficient and is used as a filter to eliminate most (typically >99.9%) of the motif list. 3) Actual Pearson coefficients are computed and the top motif is found and 4) The contribution from the top motif is subtracted from the expression data to form a residual, which is used for subsequent rounds of motif searching. This algorithm has been shown to outperform many of the other motif finding algorithms, including its predecessor REDUCE (Wu et al. 2007).

6.2.2 ARACNE

Information-theoretic approaches use a pseudo-distance between probability distributions called Mutual Information (MI), to compare expression profiles from a set of microarrays. For each pair of genes (i, j), their MI_{ij} is computed and the edge $a_{ij} = a_{ji}$ is set to 0 or 1 depending on a significance threshold to which MI_{ij} is compared. MI can be

used to measure the degree of independence between two genes. Mutual information MI_{ij} between gene i and gene j is computed as:

$$M_{ij} = H_i + H_j - H_{ij}$$

where *H*, the entropy, is defined as:

$$H_k = -\sum_{k=1}^{n} p(x_k)\log(p(x_k))$$

The entropy H_k has many interesting properties, specifically it reaches a maximum for uniformly distributed variables, i.e. the higher the entropy, the more randomly distributed are gene expression levels across the experiments. From the definition, it follows that MI becomes zero if the two variable x_i and x_j are statistically independent $(p(x_ix_j) = p(x_i)p(x_j)$, since their joint entropy $H_{ij} = H_i + H_j$. A higher MI indicates that the two genes are non-randomly associated to each other. It can be easily shown that MI is symmetric, $M_{ij} = M_{ji}$, therefore the network is described by an undirected graph G. The definition of MI requires each data point, i.e. each experiment, to be statistically independent from the others, thus information-theoretic approaches, as described here, can deal with steady-state gene expression data set, or with time-series data as long as the sampling time is long enough to assume that each point is independent from the previous ones. Edges in networks derived by information-theoretic approaches represent statistical dependences among gene expression profiles.

ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) belongs to the family of information-theoretic approaches to gene network inference by implementing the relevance network algorithm (Basso et al. 2005). ARACNE computes M_{ij} for all pairs of genes *i* and *j* in the data set. M*ij* is estimated using the method of Gaussian kernel density (Steuer et al. 2002). Once M_{ij} for all gene pairs has been

computed, ARACNE excludes all the pairs for which the null hypothesis of mutually independent genes cannot be ruled out ($H_0 : MI_{ij} = 0$). A p-value for the null hypothesis, computed using Monte Carlo simulations, is associated to each value of the mutual information. The final step of this algorithm is a pruning step based on the Data Processing Inequality (DPI) principle that tries to reduce the number of false positives. This principle asserts that if both (*i*,*j*) and (*j*,*k*) are directly interacting, and (*i*,*k*) is indirectly interacting through j, then $M_{i,k} \leq \min(M_{ij}, M_{jk})$. This condition is necessary but not sufficient, i.e. the inequality can be satisfied even if (*i*,*k*) are directly interacting, therefore the authors acknowledge that by applying this pruning step using DPI they may be discarding some direct interactions as well.

6.2.3 BANJO

A Bayesian network is a graphical model for probabilistic relationships among a set of random variables X_i, with $i = 1 \ldots n$. These relationships are encoded in the structure of a directed acyclic graph G whose vertices (or nodes) are the random variables X_i. The relationships between variables are described by a joint probability distribution $P(X_1,\ldots,X_n)$ that is consistent with the independence assertions embedded in the graph G and has the form:

$$P(X_1,...,X_n) = \prod_{i=1}^{n} P(X_i = x_i \parallel X_j = x_j,...,X_{j+p} = x_{j+p})$$

where the p+1 genes on which the probability is conditioned are called the parents of gene *i* and represent its regulators, and the joint probability density is expressed as a product of conditional probabilities by applying the chain rule of probabilities and independence. This rule is based on Bayes theorem: P(A,B) = P(B||A) * P(A) = P(A||B) *

P(B). We observe that the joint probabilty distribution can be decomposed as the product of conditional probabilities only if the Markov assumption holds, i.e. each variable X_i is independent of its non-descendants, given its parent in the directed acyclic graph G. In order to reverse-engineer a Bayesian network model of a gene network we must find the directed acyclic graph G (i.e. the regulators of each transcript) that best describes the gene expression data D. This is done by choosing a scoring function that evaluates each graph G (i.e. a possible network topology) with respect to the gene expression data D, and then searching for the graph G that maximizes the score.

BANJO is a gene network inference software that has been developed by the group of Hartemink (Yu et al. 2004). In BANJO, heuristic approaches are used to search the network space to find the network graph G (Proposer/Searcher module in BANJO). For each network structure explored, the parameters of the conditional probability density distribution are inferred and an overall network's score is computed using the Bayesian metric with Dirichlet priors and equivalence (BDe) metric in BANJO's Evaluator module. The output network will be the one with the best score (BANJO's Decider module). BANJO outputs a signed directed graph indicating regulation among genes. BANJO can analyse both steady-state and time-series data. In the case of steady-state data, BANJO, as well as the other Bayesian networks algorithms, is not able to infer networks involving cycles (e.g. feedback or feed-forward loops).

6.3 Results

6.3.1 Binding motifs found from fREDUCE for keloid versus normal fibroblasts under serum starvation condition

Binding motifs found using the gene expression values from set A (keloid versus normal fibroblasts under serum starvation conditions) are shown in Table 6.1. Highlighted motifs indicate motifs found in at least two variations of the conditions/parameters. Both MAS5 and RMA normalization as well as filtered and unfiltered gene lists provided hits for the binding motifs. Of particular note are the binding motifs CGCCGA (found in 5 of the conditions), GCCGAC (found in 3 of the conditions), CTTCTT (found in 3 of the conditions) and CACATAT (found in 3 of the conditions). A search through the TRANSFAC database did not produce any results for the binding motif CACATAT, but found possible gene targets for CGCCGA (MYB), GCCGAC (ATF2) and CTTCTT (ADH1) (Table 6.2).

Table 6.1: Binding motifs found from fREDUCE for keloid versus normal fibroblasts under serum starvation condition

Normalization	Parameters	Binding Motif	p-value
MAS 5 (unfiltered)	Length 7 (0 IUPAC)	CCGGCC	5.31
		GCCGAC	1.99
		CGTAGC	1.22
	Length 7 (1 IUPAC)	CGCBGA	5.30
		MCGGAA	1.42
RMA (unfiltered)	Length 7 (0 IUPAC)	GCCGAC	3.35
		CACATAT	2.56
		CCGGCC	1.12
	Length 7 (1 IUPAC)	GBCGAC	3.56
		CACATAT	2.02
MAS 5 (filtered)	Length 7 (0 IUPAC)	CGCCGA	2.86
		CTTCTT	1.25
		CGTAGC	1.11

	Length 7 (1 IUPAC)	**CGCCBA** (3.65)	3.65
		CCTTCYT (0.27)	0.27
RMA (filtered)	Length 7 (0 IUPAC)	**CGCCGA**	2.58
		TATACAC	1.95
	Length 7 (1 IUPAC)	**CACAKAT**	2.33
		CGCCGA	2.03
		CTTCTTV	0.18

Note: P-values are shown as $-\log_{10}$ values.
IUPAC characters M = C/A; Y = C/T; K = T/G; B = C/T/G, V = A/C/G

Table 6.2: Possible gene targets and TFs found from the TRANSFAC database for top binding motifs from Table 6.1

Binding Motif	Possible gene targets	Possible TFs
CCGGCC	MC2R (melanocortin 2 receptor)	SF-1
	MT1G (metallothionein 1G)	
	EPO (erythropoietin)	Tf-LF1 and Tf-LF2
	SURF1 and SURF2 (surfeit 1 and 2)	YY1
GCCGAC	ATF2 (activating transcription factor 2)	SP1
CGTAGC	-	-
CGCCGA	c-myb	MZF-1
CTTCTT	ADH1 (alcohol dehydrogenase)	-
CACATAT	-	-

6.3.2 Binding motifs found from fREDUCE for keloid versus normal fibroblasts under serum induced condition

No binding motifs were found for unfiltered RMA normalized set B (keloid versus normal fibroblasts under serum conditions), but binding motifs were found for the other conditions (Table 6.3). Of particular note is the binding motif GGGGCTC which was found to be consistent in 4 of the conditions, although all these 4 conditions were using the MAS 5 normalization. A search through the TRANSFAC database found ADA as a possible gene with this binding motif (Table 6.4).

Table 6.3: Binding motifs found from fREDUCE for keloid versus normal fibroblasts under serum induced condition

Normalization	Parameters	Binding Motif	p-value
MAS 5 (unfiltered)	Length 7 (0 IUPAC)	**CCACACA**	2.44
		GGGGCTC	2.19
	Length 7 (1 IUPAC)	**CCACACA**	2.14
		GGVCTC	1.91
		AGGCAH	1.30
MAS 5 (filtered)	Length 7 (0 IUPAC)	**GGGGCTC**	2.28
	Length 7 (1 IUPAC)	**GGGGHTC**	2.56
		CGAGRA	0.11
RMA (filtered)	Length 7 (0 IUPAC)	GCGCCA	2.52
		GTCCCG	1.46
	Length 7 (1 IUPAC)	**GTCVCG**	4.29
		CAACGW	0.95

Note: P-values are shown as $-\log_{10}$ values.
IUPAC characters R = A/G; W = T/A; H = A/T/C, V = A/C/G

Table 6.4: Possible gene targets and TFs found from the TRANSFAC database for top binding motifs from Table 6.3

Binding Motif	Possible gene targets	Possible TFs
CCACACA	-	-
GGGGCTC	ADA (adenosine deaminase)	SP1
GTCCCG	EGFR (EGF receptor)	-
	ATF2 (activating transcription factor 2)	SP1
	CCNE1 (cyclin E1)	E2F-1
	MET (hepatocyte growth factor receptor)	PAX-3

6.3.3 Binding motifs found from fREDUCE for sets C and D suggest consistent effects from steroid induction for both keloid and normal fibroblasts

Binding motifs were found for set C (keloid treated with steroid versus serum induced keloid fibroblasts) and D (normal treated with steroid versus serum induced normal fibroblasts) when fREDUCE was run using parameters length 6 with 0 IUPAC

substitutions. Other parameters did not produce any results. Furthermore, results were only obtained when MAS 5 normalization was used. The effect of hydrocortisone appears to be realized through the binding motifs GGAGGG and GCCCCC and this was consistent for both keloid (Table 6.5) and normal (Table 6.6) fibroblasts. A search through the TRANSFAC database using these binding motifs found a large list of genes containing these binding motifs, including COL1A2, FN, TGFB1, PDGF1 and IGF2 (Table 6.7). Of particular note is the fact that most of the genes found in this list have SP1 as its transcription factor (Table 6.7).

Table 6.5: Binding motifs found from fREDUCE for steroid treated versus control keloid fibroblasts

Normalization	Parameters	Binding Motif	p-value
MAS 5 (filtered)	Length 6 (0 IUPAC)	GGAGGG	24.62
		GCCCCC	11
		CCTGGG	7.33
		TGTGTG	3.93
		GGCTGG	3.45
		CTGTGC	1.73
		AAACAC	1.32
	Length 6 (1 IUPAC)	GGWGGG	30.68
		CCDGGG	12.92
		CTCCCH	6.23
		TGTGDG	4.52
		HACGAA	3.63
		ACCGCD	2.03
		CVGTAA	0.91

Note: P-values are shown as $-\log_{10}$ values.
IUPAC characters W = T/A; H = A/T/C; V = A/C/G; D = A/T/G

Table 6.6: Binding motifs found from fREDUCE for steroid treated versus control normal fibroblasts

Normalization	Parameters	Binding Motif	p-value
MAS 5 (filtered)	Length 6 (0 IUPAC)	GCCCCC	30.08
		GGAGGG	19.87
		CTGGGG	10.31

		TGGGCC	5
		CCCAGA	2.67
		AGAACG	2.44
		TGGGTG	2.25
		GCGAAA	1.53
		CCTGAG	1.19

Note: P-values are shown as -$\log_{10}$ values.

Table 6.7: Possible gene targets and TFs found from the TRANSFAC database for top binding motifs from Tables 6.5 and 6.6

Binding Motif	Possible gene targets	Possible TFs
GGAGGG	EPO (erythropoietin)	Tf-LF1 and Tf-LF2
	ATF2 (activating transcription factor 2)	SP1
	RARG (retinoic acid receptor, gamma)	SP1
	ACTC1 (actin, alpha, cardiac muscle 1)	SP1
	FN (fibronectin)	-
	c-myc	Pur factor
	CEACAM5 (carcinoembryonic antigen-related cell adhesion molecule 5	SP1
	CYP17 (cytochrome P450, subfamily XVII)	PBX1B
	SFTPB (surfactant protein B)	NKX2-1
	PDGFA (platelet derived growth factor A chain)	SP1, WT1
	ADA (adenosine deaminase)	SP1
	SA-ACT (skeletal alpha actin)	COUP-TF2
	MIP (major intrinsic protein of lens fiber)	SP1
	c-myb	MZF-1
	COL1A2 (collagen I alpha 2)	SP1
	ALDC (aldolase C)	-
GCCCCC	TGFB1 (transforming growth factor beta 1)	SP1 and AP1
	apoE (apolipoprotein E)	-
	c-jun	SP1
	ACTC1 (actin, alpha, cardiac muscle 1)	SP1
	ATF2 (activating transcription factor 2)	SP1
	apoB (apolipoprotein B)	-
	GFAP (glial fibrillary acidic protein)	NF1, SP1
	Cyclin D1	c-Ets-2
	Insulin	-
	ALDC (aldolase C)	-
	HRAS (transforming protein p21)	SP1
	PFKM (muscle phosphofructokinase)	SP1
	DRD1 (dopamine receptor D1)	-
	IGF2 (insulin-like growth factor 2)	WT1

6.3.4 Not many binding motifs found from fREDUCE for sets E and F

fREDUCE found few binding motifs for set E (keloid versus normal fibroblasts both treated with steroid) and no binding motifs for set F (keloid treated with HDGF versus untreated keloid fibroblasts). Binding motifs for set E were found only when the MAS 5 unfiltered condition and the RMA filtered condition were used (Table 6.8). Furthermore, binding motifs found in these conditions were not very consistent. A search through the TRANSFAC database using the top binding motifs from Table 6.8 found EGFR, ADM and CGA as possible gene targets (Table 6.9).

Table 6.8: Binding motifs found from fREDUCE for keloid versus normal fibroblasts under steroid treated condition

Normalization	Parameters	Binding Motif	p-value
MAS 5 (unfiltered)	Length 6 (0 IUPAC)	**CGCCGC**	1.53
	Length 6 (1 IUPAC)	**CGCCGC**	1.23
		GCGYTT	1.42
RMA (filtered)	Length 7 (0 IUPAC)	**GGGTTG**	2.75
		CGTTTT	1.80
		AGCGAC	1.73

Note: P-values are shown as $-\log_{10}$ values
IUPAC character Y = C/T

Table 6.9: Possible gene targets and TFs found from the TRANSFAC database for top binding motifs from Table 6.8

Binding Motif	Possible gene targets	Possible TFs
CGCCGC	EGFR (EGF receptor)	SP1
	ADM (adrenomedullin)	TFAP2A
GGGTTG	CGA (glycoprotein hormone alpha subunit)	-

6.3.5 Mean sensitivity performance of BANJO in recovering influence networks was significantly better than that of ARACNE

On average, BANJO was significantly more sensitive compared to ARACNE in recovering influence networks (Fig. 6.2C). However, there was no significant difference in average accuracy (PPV) between BANJO and ARACNE (Fig. 6.2A). Furthermore, there was no significant difference between RMA and MAS 5 normalization both in terms of mean accuracy (PPV) (Fig. 6.2B) as well as mean sensitivity (Fig. 6.2D) although p-values were fairly close to 0.05, with RMA being the better choice for both measures.

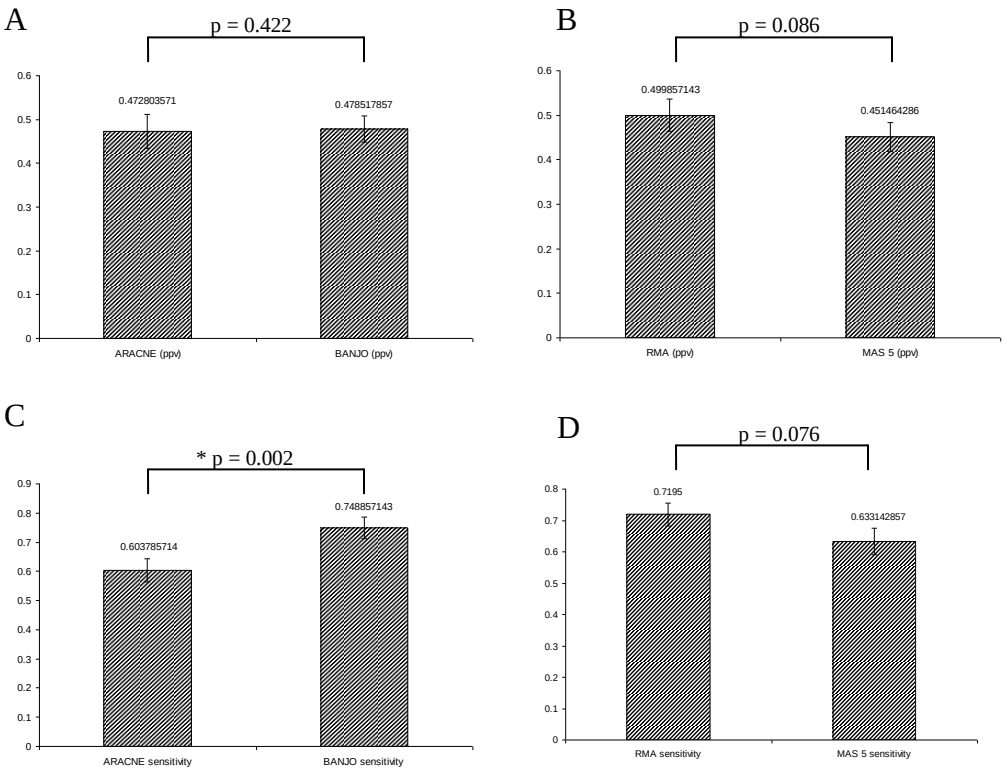


Figure 6.2: Comparison between ARACNE, BANJO, RMA and MAS 5 based on PPV and sensitivity values. (**A**) ARACNE (ppv) compared with BANJO (ppv). (**B**) RMA (ppv) compared with MAS 5 (ppv) (**C**) ARACNE (sensitivity) compared with BANJO (sensitivity) (**D**) RMA (sensitivity) compared with MAS 5 (sensitivity). Bar graphs represent mean ± S.E.M values. * indicates statistical significance as assessed by the paired t-test.

6.3.6 Transcriptional networks were better suited for network inference compared to cytokine receptor interactions and intracellular signaling networks

Transcriptional networks (Fig. 3.2C, E and G) were better suited for network inference compared to cytokine receptor interactions (Fig. 3.2A and B) and intracellular signaling networks (Fig. 3.2D and F). The full list of results for all data sets is given in Table 6.10, with bold typeface indicating performance better than random. In particular, RMA normalization for transcriptional networks (Networks C, E and G) had consistently better accuracy (PPV) compared to random, and also had sensitivity values higher than 0.5, regardless of the algorithm used (Table 6.10). BANJO appears to perform particularly well for intracellular signaling network F, but did not do so well for network D. For the NFKB transcriptional network (network E), performance using keloid sets were consistently lower than performance using normal sets, but there was very little difference in performance between the keloid and normal datasets for the other networks (Table 6.10).

Table 6.10: PPV and sensitivity results for all data sets run using BANJO and ARACNE

Data sets	A				B			
	ARACNE		BANJO		ARACNE		BANJO	
	PPV	Se	PPV	Se	PPV	Se	PPV	Se
Keloid RMA	0.238	0.556	0.3	0.667	**0.6**	1	0.4	0.667
Normal RMA	0.1875	0.333	0.3125	0.556	0.5	0.667	0.4	0.667
Keloid MAS5	0.214	0.333	0.286	0.667	0.5	0.667	**0.75**	1
Normal MAS5	0.25	0.667	0.304	0.778	**0.75**	1	0.6	1
Random	0.389				0.5			

Data sets	C				D			
	ARACNE		BANJO		ARACNE		BANJO	
	PPV	Se	PPV	Se	PPV	Se	PPV	Se
Keloid RMA	**0.8**	1	**0.75**	0.75	0.5	0.6	0.5	0.8
Normal RMA	**1**	0.75	**0.8**	1	0.5	0.6	0.2	0.2

Keloid MAS5	0.5	0.25	0.6	0.75	0.5	0.6	**0.625**	1
Normal MAS5	0.667	0.5	0.5	0.5	0.333	0.4	0.5	0.8
Random	0.667				0.5			

Data sets	E				F			
	ARACNE		BANJO		ARACNE		BANJO	
	PPV	Se	PPV	Se	PPV	Se	PPV	Se
Keloid RMA	**0.375**	0.6	**0.364**	0.8	0.375	0.75	**0.5**	1
Normal RMA	**0.444**	0.8	**0.444**	0.8	**0.5**	0.75	**0.5**	1
Keloid MAS5	0.286	0.4	0.273	0.6	0.286	0.5	**0.429**	0.75
Normal MAS5	0.333	0.6	**0.364**	0.8	0.2	0.25	**0.429**	0.75
Random	0.333				0.4			

Data sets	G			
	ARACNE		BANJO	
	PPV	Se	PPV	Se
Keloid RMA	0.5	0.5	**0.714**	0.833
Normal RMA	**0.667**	0.667	**0.625**	0.833
Keloid MAS5	0.4	0.333	0.429	0.5
Normal MAS5	**0.833**	0.833	0.5	0.5
Random	0.6			

6.4 Discussion

Reverse engineering gene networks from expression data is a considerably difficult problem, with challenges arising from the nature of the data which is typically noisy, high dimensional, and significantly undersampled. Most evaluations of reverse engineering techniques are done on simulated data (Camacho et al. 2007; Hache, Lehrach & Herwig 2009) although some have extended this to small sets of experimental data (Bansal et al. 2007; Cantone et al. 2009). While simulated data can model the high dimensionality as well as the indeterminacy of the problem accurately, the nature of noise as well as the underlying function governing the regulatory interactions has to be assumed *a priori*. A major problem of working with experimental data, however, is that not enough is known

about the real networks and this could lead to difficulties in validating the inferred networks.

Our results from the physical approach show that MAS 5 normalization was better suited for the recovery of significant binding motifs as more binding motifs were obtained when the fREDUCE method was used. However, the results from influence methods show that RMA is better for the inference of gene networks especially for the case of transcriptional networks. The performance of different normalization approaches have been assessed in a previous study by Lim et al. In the study, the Spearman rank correlation was used to compare between gene expression profile pairs from replicate samples as well as from samples with randomly permuted probe values (Lim et al. 2007). The authors found that the GCRMA procedure produced significant correlation artefacts (false positives), and that the MAS 5 procedure was best suited for the reverse engineering process. However, for most of their tests, RMA performs similarly, albeit at a lower level, compared to MAS 5. Here, we report that the performance of RMA or MAS 5 normalization appears to be dependent on the type of inference done.

The physical approach using the fREDUCE algorithm found binding motifs that were active in keloid fibroblasts compared to normal fibroblasts under various conditions. fREDUCE also found binding motifs that were responsive to steroid treatment. One limitation of the fREDUCE algorithm is that it cannot determine which TFs bind to the discovered motifs. By manually searching through the TRANSFAC database, we are able to get some idea about target genes containing these motifs, as well as possible transcription factors that bind to these motifs. The TRANSFAC database is not complete

however, as it is based on published data, therefore undiscovered interactions will not be reflected in the database.

Our results suggest that steroid treatment affected both keloid and normal fibroblasts in a similar fashion as the top binding motifs found when these two cell types were treated with hydrocortisone were the same. Many of the possible gene targets containing these binding motifs are involved in wound healing, for example fibronectin, erythropoietin, PDGF, COL1A1 and TGFB. This is consistent with the fact that steroids are known to have a depressive effect on wound healing (Wicke et al. 2000). Furthermore, SP1 was the most common TF found for these gene targets. This result suggests that hydrocortisone exerts its depressive effect on fibroblasts by affecting the activity of SP1, and could be a future area of research. There were fewer gene targets found for binding motifs that were active when comparing keloid to normal fibroblasts. Furthermore, the TFs found for these conditions were also less consistent. This could be due to the fact that the keloid condition is a result of the effect of multiple TFs, and unlike the effect of hydrocortisone, no single TF is most responsible.

The success of fREDUCE depends on a number of assumptions regarding the dynamics of transcription. Most notably, it relates the influence of combinations of TFs as a log-linear function of RNA levels. Such a highly constrained model may lead to errors in predictions. Furthermore, it assumes that the 1000 base pairs upstream of the transcription start site play some role in the regulation of the gene. Despite these limitations, fREDUCE has been used successfully to discover binding motifs in human liver tissue (Wu et al. 2007).

We used fairly naïve methods in the preprocessing of data for our influence based inference methods. To enable comparison between multiple datasets, we normalized the expression values with the average of GAPDH and B-actin expression values for each individual chip whereas for discretization, we used the 7 bin quantile discretization that is available in BANJO. More sophisticated discretization techniques (Friedman et al. 2000; Pe'er et al. 2001; Becquet et al. 2002) might potentially produce better results.

Due to our limited data, it would be unwise to run the influence algorithms on the full list of genes. The subsets of genes that we selected were based on KEGG pathways that were found to be significantly enriched by the differentially expressed genes found in the previous section. We further subdivided the pathways found into three major groups – cytokine receptor interactions, transcriptional networks and intracellular signaling. These lists are by no means complete, and there is bound to be many hidden factors and feedback interactions that were not explicitly modeled or taken into account. However, in the absence of further biological knowledge to guide us in our selection, this seemed to be the most logical step to take. Futhermore, it is hoped that the influence methods, being of the 'black box' variety, would be able to cope with these deficiencies.

Our results show that both ARACNE and BANJO seem to perform better for transcriptional networks compared to cytokine receptor interations or intracellular signaling networks. This makes some intuitive sense as there is a causal link between transcription factors and their target genes, whereas in cytokine receptor interactions and in intracellular signaling, there is at most only a correlation (in most cases, there may not even be a correlation). However, a cellular signaling network has been successfully reconstructed previously using a Bayesian approach (Sachs et al. 2005). It should be

noted however that Sach's study measured phosphorylation levels in addition to expression levels and used the flow cytometry platform instead of microarray expression data. On a related note, it is worth pointing out that influence methods using microarray data do not take the actual binding affinities of transcription factors into consideration as only expression values are used. This could be a source of inaccuracies in the networks inferred.

Between the two influence methods tested, BANJO produced significantly better results compared to ARACNE. The superior performance of BANJO for small data sets with 'global' perturbations can also be seen in the results of the *in silico* study done by Bansal et al. (Bansal et al. 2007). The lower performance of ARACNE could be due to the small number of data sets used; ARACNE has been recommended to be used on data sets containing a minimum of 100 microarray expression profiles as this represents an empirical lower bound on the amount of data needed to estimate the MI reliably (Margolin et al. 2006). Having said that, none of the PPV values obtained either through BANJO or ARACNE was able to beat the random score significantly as assessed by the chi-squared test, although the absolute PPV values were higher in some of the cases. Achieving statistical significance with a small number of genes requires the difference in data distributions to be very large and may be too demanding for our small networks.

The fact that performance for the transcriptional network involving NFKB (nuclear factor kappa-light-chain-enhancer of activated B cells, Set E) using influence methods was better for normal fibroblasts compared to keloid fibroblasts suggests that the influence between NFKB and its targets was weaker in keloid fibroblasts, or that there were more links in the keloid network that were not captured by our simplified diagram.

This would imply that targeting of NFKB alone may not be sufficient in reducing the expression of its targets in keloids. However, more work needs to be done to verify this hypothesis.

The ability to infer molecular interactions in cellular systems is one of the most exciting promises of systems biology. As the most widely available high throughput technology, gene expression microarrays provide a good test set for the application of inference algorithms that infer dynamic models from static, genome-scale data. However, the critical assumption underlying this methodology is that mRNA measurements are predictive of molecular activity. This assumption has been thrown into question as new studies reveal the substantial role of alternate regulatory mechanisms, such as translation, post-translational modifications, genetic and epigenetic factors, as well as the increasingly appreciated regulatory role of non-coding RNAs. Furthermore, data from the microarray platform is typically noisy, and is also hidden in multiple probes that can be combined in multiple ways to produce different expression values. Yet in spite of all these difficulties, the topic of reverse engineering gene networks is surely worth pursuing, as it provides us with a means of understanding biology not only in terms of the genes themselves, but also through their interactions.

CHAPTER SEVEN

CONCLUSION

"Now they [genes] are trapped in huge colonies, locked inside highly intelligent beings, moulded by the outside world, communicating with it by complex processes, through which, blindly, as if by magic, function emerges."

– Denis Noble, Emeritus Professor, University of Oxford

The reductionist approach to biological problems has had tremendous success in the past, culminating in the discovery of complete genomes of several organisms, including humans. Thanks to molecular biology, we now understand how linear arrangement of nucleotides encodes linear arrangement of amino acids and how proteins interact to form functional groups such as signal transduction and metabolic pathways. In recent years however, there has been a greater realization about the limitations of reductionist approaches. Having reduced the biological universe to a myriad of minute parts, we are now unable to assemble them back together again in manner that increases our biological understanding. This has led to the development of the nascent field of systems biology, where methods adapted from math and engineering disciplines are employed to shed light on the complex cellular networks present in living organisms.

We have utilized both reductionist as well as top down approaches in this dissertation in an attempt to better understand keloid scarring. We first investigated the novel growth factor HDGF and its role in the keloid system using cell and molecular techniques. Our results suggest that this growth factor is upregulated in the keloid

condition and has angiogenic and proliferative potential. However, it does not seem to increase extracellular matrix production and thus HDGF appears to be just one of a myriad number of players underlying the keloid condition. Targeting HDGF alone would probably not be sufficient in ameliorating this condition.

In the second part of this dissertation, we decided to take a top down approach in an attempt to identify groups of genes that can be implicated in the formation of keloids. Surveying the global transcriptional landscape is now possible with the advent of microarray technology, although this platform is still far from perfect as it is inherently noisy. Furthermore, there is still no single standard for the processing of raw data for the Affymetrix Genechips, and different algorithms result in different expression values for the same raw data. Despite all these shortcomings, the microarray platform has been widely used to profile mRNA expression values, and our results indicate consistency with previous studies done on keloid fibroblasts. We have also uncovered differentially expressed genes that have not been reported previously, and enrichment analysis indicate that processes such as immune response, antigen processing and presentation, chemokine and cytokine activity, extracellular matrix and ribosomal proteins are among those that are affected in the keloid condition.

In the third part of this dissertation, we have attempted to reverse engineer gene networks from the collection of microarray expression data that we have collected. Using the physical approach of correlating expression values to binding motifs, we found some consensus sequences that were active in the keloid condition, as well as some sequences that were responsive to steroid treatment. Using influence approaches on experimental data, we found that the combination of the Bayesian algorithm, RMA normalization and

transcriptional networks gave the best results. Furthermore, our results show that the NFKB transcriptional network inferred from normal fibroblast data was more accurate than that inferred from keloid data, suggesting a more robust network in the keloid condition. This finding has therapeutic implications in that targeting NFKB alone may not be sufficient in reducing the transcriptional products of NFKB in keloid fibroblasts.

We are still far from a solution to the keloid condition, but it is hoped that the work done here is a small step towards finding this solution. In this thesis, we have mainly focused on fibroblasts, which is only one of the possible cell types involved in wound healing. Future work could involve microarray profiling of keratinocytes, as well as profiling of cells in the co-culture condition to examine the effect of epithelial mesenchymal interactions on transcriptional networks. Furthermore, novel genes obtained from our microarray experiments, or those containing the binding motifs that were active in the various conditions, could be studied in depth through reductionist approaches. Pathways such as the antigen processing and presentation pathway and the toll-like receptor signaling pathway keloid and normal fibroblasts could also be examined in further detail. Finally, improved methods that model alternate regulatory mechanisms such as post translational modifications and binding affinities could be developed to improve the accuracy of the reverse engineering process.

BIBLIOGRAPHY

Aarabi, S, Bhatt, KA, Shi, Y, Paterno, J, Chang, EI, Loh, SA, Holmes, JW, Longaker, MT, Yee, H & Gurtner, GC 2007, 'Mechanical load initiates hypertrophic scar formation through decreased cellular apoptosis', *FASEB J*, vol. 21, no. 12, pp. 3250-3261.

Abouzied, MM, Baader, SL, Dietz, F, Kappler, J, Gieselmann, V & Franken, S 2004, 'Expression patterns and different subcellular localization of the growth factors HDGF (hepatoma-derived growth factor) and HRP-3 (HDGF-related protein-3) suggest functions in addition to their mitogenic activity', *Biochem J*, vol. 378, no. Pt 1, pp. 169-176.

Abouzied, MM, El-Tahir, HM, Prenner, L, Häberlein, H, Gieselmann, V & Franken, S 2005, 'Hepatoma-derived growth factor. Significance of amino acid residues 81-100 in cell surface interaction and proliferative activity', *J Biol Chem*, vol. 280, no. 12, pp. 10945-10954.

Al-Attar, A, Mess, S, Thomassen, JM, Kauffman, CL & Davison, SP 2006, 'Keloid pathogenesis and treatment', *Plast Reconstr Surg*, vol. 117, no. 1, pp. 286-300.

Apfelberg, DB, Maser, MR, White, DN & Lash, H 1989, 'Failure of carbon dioxide laser excision of keloids', *Lasers Surg Med*, vol. 9, no. 4, pp. 382-388.

Appleton, I, Brown, NJ & Willoughby, DA 1996, 'Apoptosis, necrosis, and proliferation: possible implications in the etiology of keloids', *Am J Pathol*, vol. 149, no. 5, pp. 1441-1447.

Atiyeh, BS, Costagliola, M & Hayek, SN 2005, 'Keloid or hypertrophic scar: the controversy: review of the literature', *Ann Plast Surg*, vol. 54, no. 6, pp. 676-680.

Babu, M, Diegelmann, R & Oliver, N 1989, 'Fibronectin is overproduced by keloid fibroblasts during abnormal wound healing', *Mol Cell Biol*, vol. 9, no. 4, pp. 1642-1650.

Bansal, M, Belcastro, V, Ambesi-Impiombato, A & di Bernardo, D 2007, 'How to infer gene networks from expression profiles', *Mol Syst Biol*, vol. 3, pp. 78.

Basso, K, Margolin, AA, Stolovitzky, G, Klein, U, Dalla-Favera, R & Califano, A 2005, 'Reverse engineering of regulatory networks in human B cells', *Nat Genet*, vol. 37, no. 4, pp. 382-390.

Bayat, A, Arscott, G, Ollier, WER, Ferguson, MWJ & Mc Grouther, DA 2004, 'Description of site-specific morphology of keloid phenotypes in an Afrocaribbean population', *Br J Plast Surg*, vol. 57, no. 2, pp. 122-133.

Becquet, C, Blachon, S, Jeudy, B, Boulicaut, JF & Gandrillon, O 2002, 'Strong-association-rule mining for large-scale gene-expression data analysis: a case study on human SAGE data', *Genome Biol*, vol. 3, no. 12, pp. RESEARCH0067.

Bernard, K, Litman, E, Fitzpatrick, JL, Shellman, YG, Argast, G, Polvinen, K, Everett, AD, Fukasawa, K, Norris, DA, Ahn, NG & Resing, KA 2003, 'Functional proteomic analysis of melanoma progression', *Cancer Res*, vol. 63, no. 20, pp. 6716-6725.

Bloom, D 1956, 'Heredity of keloids; review of the literature and report of a family with multiple keloids in five generations', *N Y State J Med*, vol. 56, no. 4, pp. 511-519.

Bock, O, Schmid-Ott, G, Malewski, P & Mrowietz, U 2006, 'Quality of life of patients with keloid and hypertrophic scarring', *Arch Dermatol Res*, vol. 297, no. 10, pp. 433-438.

Bolstad, BM, Irizarry, RA, Astrand, M & Speed, TP 2003, 'A comparison of normalization methods for high density oligonucleotide array data based on variance and bias', *Bioinformatics*, vol. 19, no. 2, pp. 185-193.

Bussemaker, HJ, Li, H & Siggia, ED 2001, 'Regulatory element detection using correlation with expression', *Nat Genet*, vol. 27, no. 2, pp. 167-171.

Caldwell, GM, Jones, C, Gensberg, K, Jan, S, Hardy, RG, Byrd, P, Chughtai, S, Wallis, Y, Matthews, GM & Morton, DG 2004, 'The Wnt antagonist sFRP1 in colorectal tumorigenesis', *Cancer Res*, vol. 64, no. 3, pp. 883-888.

Camacho, D, Vera Licona, P, Mendes, P & Laubenbacher, R 2007, 'Comparison of reverse-engineering methods using an in silico network', *Ann N Y Acad Sci*, vol. 1115, pp. 73-89.

Cantone, I, Marucci, L, Iorio, F, Ricci, MA, Belcastro, V, Bansal, M, Santini, S, di Bernardo, M, di Bernardo, D & Cosma, MP 2009, 'A yeast synthetic network for in vivo assessment of reverse-engineering and modeling approaches', *Cell*, vol. 137, no. 1, pp. 172-181.

Chang, HY, Chi, JT, Dudoit, S, Bondre, C, van de Rijn, M, Botstein, D & Brown, PO 2002, 'Diversity, topographic differentiation, and positional memory in human fibroblasts', *Proc Natl Acad Sci U S A*, vol. 99, no. 20, pp. 12877-12882.

Chang, KC, Tai, MH, Lin, JW, Wang, CC, Huang, CC, Hung, CH, Chen, CH, Lu, SN, Lee, CM, Changchien, CS & Hu, TH 2007, 'Hepatoma-derived growth factor is a novel prognostic factor for gastrointestinal stromal tumors', *Int J Cancer*, vol. 121, no. 5, pp. 1059-1065.

Chen, W, Fu, X, Sun, X, Sun, T, Zhao, Z & Sheng, Z 2003, 'Analysis of differentially expressed genes in keloids and normal skin with cDNA microarray', *J Surg Res*, vol. 113, no. 2, pp. 208-216.

Cheng, JC, Evans, JH, Leung, KS, Clark, JA, Choy, TT & Leung, PC 1984, 'Pressure therapy in the treatment of post-burn hypertrophic scar--a critical look into its usefulness and fallacies by pressure monitoring', *Burns Incl Therm Inj*, vol. 10, no. 3, pp. 154-163.

Chin, GS, Liu, W, Peled, Z, Lee, TY, Steinbrech, DS, Hsu, M & Longaker, MT 2001, 'Differential expression of transforming growth factor-beta receptors I and II and activation of Smad 3 in keloid fibroblasts', *Plast Reconstr Surg*, vol. 108, no. 2, pp. 423-429.

Chipev, CC, Simman, R, Hatch, G, Katz, AE, Siegel, DM & Simon, M 2000, 'Myofibroblast phenotype and apoptosis in keloid and palmar fibroblasts in vitro', *Cell Death Differ*, vol. 7, no. 2, pp. 166-176.

Chodon, T, Sugihara, T, Igawa, HH, Funayama, E & Furukawa, H 2000, 'Keloid-derived fibroblasts are refractory to Fas-mediated apoptosis and neutralization of autocrine transforming growth factor-beta1 can abrogate this resistance', *Am J Pathol*, vol. 157, no. 5, pp. 1661-1669.

Clark, RA 2001, 'Fibrin and wound healing', *Ann N Y Acad Sci*, vol. 936, pp. 355-367.

Collett-Solberg, PF & Cohen, P 1996, 'The role of the insulin-like growth factor binding proteins and the IGFBP proteases in modulating IGF action', *Endocrinol Metab Clin North Am*, vol. 25, no. 3, pp. 591-614.

Cresswell, P, Ackerman, AL, Giodini, A, Peaper, DR & Wearsch, PA 2005, 'Mechanisms of MHC class I-restricted antigen processing and cross-presentation', *Immunol Rev*, vol. 207, pp. 145-157.

Curwen, V, Eyras, E, Andrews, TD, Clarke, L, Mongin, E, Searle, SMJ & Clamp, M 2004, 'The Ensembl automatic gene annotation system', *Genome Res*, vol. 14, no. 5, pp. 942-950.

Dai, M, Wang, P, Boyd, AD, Kostov, G, Athey, B, Jones, EG, Bunney, WE, Myers, RM, Speed, TP, Akil, H, Watson, SJ & Meng, F 2005, 'Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data', *Nucleic Acids Res*, vol. 33, no. 20, pp. e175.

Daian, T, Ohtsuru, A, Rogounovitch, T, Ishihara, H, Hirano, A, Akiyama-Uchida, Y, Saenko, V, Fujii, T & Yamashita, S 2003, 'Insulin-like growth factor-I enhances transforming growth factor-beta-induced extracellular matrix protein production through the P38/activating transcription factor-2 signaling pathway in keloid fibroblasts', *J Invest Dermatol*, vol. 120, no. 6, pp. 956-962.

Datubo-Brown, DD 1990, 'Keloids: a review of the literature', *Br J Plast Surg*, vol. 43, no. 1, pp. 70-77.

Dennis, G, Sherman, BT, Hosack, DA, Yang, J, Gao, W, Lane, HC & Lempicki, RA 2003, 'DAVID: Database for Annotation, Visualization, and Integrated Discovery', *Genome Biol*, vol. 4, no. 5, pp. P3.

Derynck, R & Zhang, YE 2003, 'Smad-dependent and Smad-independent pathways in TGF-beta family signalling', *Nature*, vol. 425, no. 6958, pp. 577-584.

Diegelmann, RF, Cohen, IK & Kaplan, AM 1981, 'The role of macrophages in wound repair: a review', *Plast Reconstr Surg*, vol. 68, no. 1, pp. 107-113.

English, RS & Shenefelt, PD 1999, 'Keloids and hypertrophic scars', *Dermatol Surg*, vol. 25, no. 8, pp. 631-638.

Enomoto, H, Yoshida, K, Kishima, Y, Kinoshita, T, Yamamoto, M, Everett, AD, Miyajima, A & Nakamura, H 2002, 'Hepatoma-derived growth factor is highly expressed in developing liver and promotes fetal hepatocyte proliferation', *Hepatology*, vol. 36, no. 6, pp. 1519-1527.

Everett, AD, Lobe, DR, Matsumura, ME, Nakamura, H & McNamara, CA 2000, 'Hepatoma-derived growth factor stimulates smooth muscle cell growth and is expressed in vascular development', *J Clin Invest*, vol. 105, no. 5, pp. 567-575.

Everett, AD, Stoops, T & McNamara, CA 2001, 'Nuclear targeting is required for hepatoma-derived growth factor-stimulated mitogenesis in vascular smooth muscle cells', *J Biol Chem*, vol. 276, no. 40, pp. 37564-37568.

Everett, AD, Narron, JV, Stoops, T, Nakamura, H & Tucker, A 2004, 'Hepatoma-derived growth factor is a pulmonary endothelial cell-expressed angiogenic factor', *Am J Physiol Lung Cell Mol Physiol*, vol. 286, no. 6, pp. L1194-L1201.

Ezan, J, Leroux, L, Barandon, L, Dufourcq, P, Jaspard, B, Moreau, C, Allières, C, Daret, D, Couffinhal, T & Duplàa, C 2004, 'FrzA/sFRP-1, a secreted antagonist of the Wnt-Frizzled pathway, controls vascular cell proliferation in vitro and in vivo', *Cardiovasc Res*, vol. 63, no. 4, pp. 731-738.

Feghali, CA & Wright, TM 1999, 'Identification of multiple, differentially expressed messenger RNAs in dermal fibroblasts from patients with systemic sclerosis', *Arthritis Rheum*, vol. 42, no. 7, pp. 1451-1457.

Ford, LC, King, DF, Lagasse, LD & Newcomer, V 1983, 'Increased androgen binding in keloids: a preliminary communication', *J Dermatol Surg Oncol*, vol. 9, no. 7, pp. 545-547.

Friedman, N, Linial, M, Nachman, I & Pe'er, D 2000, 'Using Bayesian networks to analyze expression data', *J Comput Biol*, vol. 7, no. 3-4, pp. 601-620.

Funayama, E, Chodon, T, Oyama, A & Sugihara, T 2003, 'Keratinocytes promote proliferation and inhibit apoptosis of the underlying fibroblasts: an important role in the pathogenesis of keloid', *J Invest Dermatol*, vol. 121, no. 6, pp. 1326-1331.

Gallitzendoerfer, R, Abouzied, MM, Hartmann, D, Dobrowolski, R, Gieselmann, V & Franken, S 2008, 'Hepatoma-derived growth factor (HDGF) is dispensable for normal mouse development', *Dev Dyn*, vol. 237, no. 7, pp. 1875-1885.

Gardner, TS & Faith, JJ 2010, 'Reverse-engineering transcription control networks', *Physics of life reviews*, vol. 2, no. 1, pp. 65-88.

Ghazizadeh, M 2007, 'Essential role of IL-6 signaling pathway in keloid pathogenesis', *J Nippon Med Sch*, vol. 74, no. 1, pp. 11-22.

Ghazizadeh, M, Tosa, M, Shimizu, H, Hyakusoku, H & Kawanami, O 2007, 'Functional implications of the IL-6 signaling pathway in keloid pathogenesis', *J Invest Dermatol*, vol. 127, no. 1, pp. 98-105.

Ghosh, AK, Yuan, W, Mori, Y & Varga, J 2000, 'Smad-dependent stimulation of type I collagen gene expression in human skin fibroblasts by TGF-beta involves functional cooperation with p300/CBP transcriptional coactivators', *Oncogene*, vol. 19, no. 31, pp. 3546-3555.

Goutsias, J & Lee, NH 2007, 'Computational and experimental approaches for modeling gene regulatory networks', *Curr Pharm Des*, vol. 13, no. 14, pp. 1415-1436.

Grommé, M & Neefjes, J 2002, 'Antigen degradation or presentation by MHC class I molecules via classical and non-classical pathways', *Mol Immunol*, vol. 39, no. 3-4, pp. 181-202.

Grotendorst, GR, Rahmanie, H & Duncan, MR 2004, 'Combinatorial signaling pathways determine fibroblast proliferation and myofibroblast differentiation', *FASEB J*, vol. 18, no. 3, pp. 469-479.

Hache, H, Lehrach, H & Herwig, R 2009, 'Reverse Engineering of Gene Regulatory Networks: A Comparative Study', *EURASIP journal on bioinformatics & systems biology*, vol. 2009, pp. 617281.

Hache, H, Lehrach, H & Herwig, R 2009, 'Reverse engineering of gene regulatory networks: a comparative study', *EURASIP journal on bioinformatics & systems biology*, pp. 617281.

Haisa, M, Okochi, H & Grotendorst, GR 1994, 'Elevated levels of PDGF alpha receptors in keloid fibroblasts contribute to an enhanced response to PDGF', *J Invest Dermatol*, vol. 103, no. 4, pp. 560-563.

Hansen, MB, Nielsen, SE & Berg, K 1989, 'Re-examination and further development of a precise and rapid dye method for measuring cell growth/cell kill', *J Immunol Methods*, vol. 119, no. 2, pp. 203-210.

Harr, B & Schlötterer, C 2006, 'Comparison of algorithms for the analysis of Affymetrix microarray data as evaluated by co-expression of genes in known operons', *Nucleic Acids Res*, vol. 34, no. 2, pp. e8.

Hartwell, LH, Hopfield, JJ, Leibler, S & Murray, AW 1999, 'From molecular to modular cell biology', *Nature*, vol. 402, no. 6761 Suppl, pp. C47-C52.

Hengel, H, Esslinger, C, Pool, J, Goulmy, E & Koszinowski, UH 1995, 'Cytokines restore MHC class I complex formation and control antigen presentation in human cytomegalovirus-infected cells', *J Gen Virol*, vol. 76 (Pt 12), pp. 2987-2997.

Henry, JL, Coggin, DL & King, CR 1993, 'High-level expression of the ribosomal protein L19 in human breast tumors that overexpress erbB-2', *Cancer Res*, vol. 53, no. 6, pp. 1403-1408.

Hinz, B 2006, 'Masters and servants of the force: the role of matrix adhesions in myofibroblast force perception and transmission', *Eur J Cell Biol*, vol. 85, no. 3-4, pp. 175-181.

Huang, DW, Sherman, BT & Lempicki, RA 2009, 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources', *Nat Protoc*, vol. 4, no. 1, pp. 44-57.

Ishihara, H, Yoshimoto, H, Fujioka, M, Murakami, R, Hirano, A, Fujii, T, Ohtsuru, A, Namba, H & Yamashita, S 2000, 'Keloid fibroblasts resist ceramide-induced apoptosis by overexpression of insulin-like growth factor I receptor', *J Invest Dermatol*, vol. 115, no. 6, pp. 1065-1071.

Iyer, VR, Eisen, MB, Ross, DT, Schuler, G, Moore, T, Lee, JC, Trent, JM, Staudt, LM, Hudson, J, Boguski, MS, Lashkari, D, Shalon, D, Botstein, D & Brown, PO 1999, 'The transcriptional program in the response of human fibroblasts to serum', *Science*, vol. 283, no. 5398, pp. 83-87.

Izumoto, Y, Kuroda, T, Harada, H, Kishimoto, T & Nakamura, H 1997, 'Hepatoma-derived growth factor belongs to a gene family in mice showing significant homology in the amino terminus', *Biochem Biophys Res Commun*, vol. 238, no. 1, pp. 26-32.

Jeffery, TK, Upton, PD, Trembath, RC & Morrell, NW 2005, 'BMP4 inhibits proliferation and promotes myocyte differentiation of lung fibroblasts via Smad1 and JNK pathways', *Am J Physiol Lung Cell Mol Physiol*, vol. 288, no. 2, pp. L370-L378.

Johnson, WE, Li, C & Rabinovic, A 2007, 'Adjusting batch effects in microarray expression data using empirical Bayes methods', *Biostatistics*, vol. 8, no. 1, pp. 118-127.

Kessler, D, Dethlefsen, S, Haase, I, Plomann, M, Hirche, F, Krieg, T & Eckes, B 2001, 'Fibroblasts in mechanically stressed collagen lattices assume a "synthetic" phenotype', *J Biol Chem*, vol. 276, no. 39, pp. 36575-36585.

Kilian, B, Mansukoski, H, Barbosa, FC, Ulrich, F, Tada, M & Heisenberg, CP 2003, 'The role of Ppt/Wnt5 in regulating cell shape and movement during zebrafish gastrulation', *Mech Dev*, vol. 120, no. 4, pp. 467-476.

Kischer, CW 1992, 'The microvessels in hypertrophic scars, keloids and related lesions: a review', *J Submicrosc Cytol Pathol*, vol. 24, no. 2, pp. 281-296.

Kischer, CW, Shetlar, MR, Shetlar, CL & Chvapil, M 1983, 'Immunoglobulins in hypertrophic scars and keloids', *Plast Reconstr Surg*, vol. 71, no. 6, pp. 821-825.

Kishima, Y, Yamamoto, H, Izumoto, Y, Yoshida, K, Enomoto, H, Yamamoto, M, Kuroda, T, Ito, H, Yoshizaki, K & Nakamura, H 2002, 'Hepatoma-derived growth factor stimulates cell growth after translocation to the nucleus by nuclear localization signals', *J Biol Chem*, vol. 277, no. 12, pp. 10315-10322.

Knighton, DR, Hunt, TK, Scheuenstuhl, H, Halliday, BJ, Werb, Z & Banda, MJ 1983, 'Oxygen tension regulates the expression of angiogenesis factor by macrophages', *Science*, vol. 221, no. 4617, pp. 1283-1285.

Lai, MD & Xu, J 2007, 'Ribosomal proteins and colorectal cancer', *Current genomics*, vol. 8, no. 1, pp. 43-49.

Larrabee, WF, East, CA, Jaffe, HS, Stephenson, C & Peterson, KE 1990, 'Intralesional interferon gamma treatment for keloids and hypertrophic scars', *Arch Otolaryngol Head Neck Surg*, vol. 116, no. 10, pp. 1159-1162.

Lee, SS, Yosipovitch, G, Chan, YH & Goh, CL 2004, 'Pruritus, pain, and small nerve fiber function in keloids: a controlled study', *J Am Acad Dermatol*, vol. 51, no. 6, pp. 1002-1006.

Lee, TY, Chin, GS, Kim, WJ, Chau, D, Gittes, GK & Longaker, MT 1999, 'Expression of transforming growth factor beta 1, 2, and 3 proteins in keloids', *Ann Plast Surg*, vol. 43, no. 2, pp. 179-184.

Li, C & Wong, WH 2001, 'Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection', *Proc Natl Acad Sci U S A*, vol. 98, no. 1, pp. 31-36.

Lim, IJ, Phan, TT, Song, C, Tan, WT & Longaker, MT 2001, 'Investigation of the

influence of keloid-derived keratinocytes on fibroblast growth and proliferation in vitro', *Plast Reconstr Surg*, vol. 107, no. 3, pp. 797-808.

Lim, IJ, Phan, TT, Bay, BH, Qi, R, Huynh, H, Tan, WTL, Lee, ST & Longaker, MT 2002, 'Fibroblasts cocultured with keloid keratinocytes: normal fibroblasts secrete collagen in a keloidlike manner', *Am J Physiol Cell Physiol*, vol. 283, no. 1, pp. C212-C222.

Lim, WK, Wang, K, Lefebvre, C & Califano, A 2007, 'Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks', *Bioinformatics*, vol. 23, no. 13, pp. i282-i288.

Lodygin, D, Epanchintsev, A, Menssen, A, Diebold, J & Hermeking, H 2005, 'Functional epigenomics identifies genes frequently silenced in prostate cancer', *Cancer Res*, vol. 65, no. 10, pp. 4218-4227.

Louw, L 2007, 'The keloid phenomenon: progress toward a solution', *Clin Anat*, vol. 20, no. 1, pp. 3-14.

Mao, J, Xu, Z, Fang, Y, Wang, H, Xu, J, Ye, J, Zheng, S & Zhu, Y 2008, 'Hepatoma-derived growth factor involved in the carcinogenesis of gastric epithelial cells through promotion of cell proliferation by Erk1/2 activation', *Cancer Sci*, vol. 99, no. 11, pp. 2120-2127.

Margolin, AA, Wang, K, Lim, WK, Kustagi, M, Nemenman, I & Califano, A 2006, 'Reverse engineering cellular networks', *Nat Protoc*, vol. 1, no. 2, pp. 662-671.

Marneros, AG & Krieg, T 2004, 'Keloids--clinical diagnosis, pathogenesis, and treatment options', *J Dtsch Dermatol Ges*, vol. 2, no. 11, pp. 905-913.

Marsit, CJ, Karagas, MR, Andrew, A, Liu, M, Danaee, H, Schned, AR, Nelson, HH & Kelsey, KT 2005, 'Epigenetic inactivation of SFRP genes and TP53 alteration act jointly as markers of invasive bladder cancer', *Cancer Res*, vol. 65, no. 16, pp. 7081-7085.

Masckauchán, TNH, Agalliu, D, Vorontchikhina, M, Ahn, A, Parmalee, NL, Li, CM, Khoo, A, Tycko, B, Brown, AMC & Kitajewski, J 2006, 'Wnt5a signaling induces proliferation and survival of endothelial cells in vitro and expression of MMP-1 and Tie-2', *Mol Biol Cell*, vol. 17, no. 12, pp. 5163-5172.

Massagué, J 1998, 'TGF-beta signal transduction', *Annu Rev Biochem*, vol. 67, pp. 753-791.

Mast, BA 1992, 'Wound Healing: Biochemical and Clinical Aspects' in *The Skin*, ed IK Cohen, B Diegelmann & WJ and Lindblad, W.B. Saunders Company, Philadelphia, pp. 344.

Matsuyama, A, Inoue, H, Shibuta, K, Tanaka, Y, Barnard, GF, Sugimachi, K & Mori, M 2001, 'Hepatoma-derived growth factor is associated with reduced sensitivity to irradiation in esophageal cancer', *Cancer Res*, vol. 61, no. 15, pp. 5714-5717.

Matys, V, Fricke, E, Geffers, R, Gössling, E, Haubrock, M, Hehl, R, Hornischer, K, Karas, D, Kel, AE, Kel-Margoulis, OV, Kloos, DU, Land, S, Lewicki-Potapov, B, Michael, H, Münch, R, Reuter, I, Rotert, S, Saxel, H, Scheer, M, Thiele, S & Wingender, E 2003, 'TRANSFAC: transcriptional regulation, from patterns to profiles', *Nucleic Acids Res*, vol. 31, no. 1, pp. 374-378.

McDonald, SL & Silver, A 2009, 'The opposing roles of Wnt-5a in cancer', *Br J Cancer*, vol. 101, no. 2, pp. 209-214.

Messadi, DV, Le, A, Berg, S, Huang, G, Zhuang, W & Bertolami, CN 1998, 'Effect of TGF-beta 1 on PDGF receptors expression in human scar fibroblasts', *Front Biosci*, vol. 3, pp. a16-a22.

Messadi, DV, Doung, HS, Zhang, Q, Kelly, AP, Tuan, TL, Reichenberger, E & Le, AD 2004, 'Activation of NFkappaB signal pathways in keloid fibroblasts', *Arch Dermatol Res*, vol. 296, no. 3, pp. 125-133.

Mohan, S & Baylink, DJ 2002, 'IGF-binding proteins are multifunctional and act via IGF-dependent and -independent mechanisms', *J Endocrinol*, vol. 175, no. 1, pp. 19-31.

Mori, M, Morishita, H, Nakamura, H, Matsuoka, H, Yoshida, K, Kishima, Y, Zhou, Z, Kida, H, Funakoshi, T, Goya, S, Yoshida, M, Kumagai, T, Tachibana, I, Yamamoto, Y, Kawase, I & Hayashi, S 2004, 'Hepatoma-derived growth factor is involved in lung remodeling by stimulating epithelial growth', *Am J Respir Cell Mol Biol*, vol. 30, no. 4, pp. 459-469.

Moustafa, MF, Abdel-Fattah, MA & Abdel-Fattah, DC 1975, 'Presumptive evidence of the effect of pregnancy estrogens on keloid growth. Case report', *Plast Reconstr Surg*, vol. 56, no. 4, pp. 450-453.

Mukhopadhyay, A, Khoo, A, Cheong, HH, Chan, SY, Aalami, O, Lim, IJ & Phan, TT 2007, 'Targeting of Sp1 transcription factor: a novel therapeutic approach for keloids, an in vitro analysis', *Exp Dermatol*, vol. 16, no. 12, pp. 1023-1031.

Nakamura, H, Izumoto, Y, Kambe, H, Kuroda, T, Mori, T, Kawamura, K, Yamamoto, H & Kishimoto, T 1994, 'Molecular cloning of complementary DNA for a novel human hepatoma-derived growth factor. Its homology with high mobility group-1 protein', *J Biol Chem*, vol. 269, no. 40, pp. 25143-25149.

Nakamura, H, Kambe, H, Egawa, T, Kimura, Y, Ito, H, Hayashi, E, Yamamoto, H, Sato, J & Kishimoto, S 1989, 'Partial purification and characterization of human hepatoma-derived growth factor', *Clin Chim Acta*, vol. 183, no. 3, pp. 273-284.

Nemeth, AJ 1993, 'Keloids and hypertrophic scars', *J Dermatol Surg Oncol*, vol. 19, no. 8, pp. 738-746.

Niessen, FB, Schalkwijk, J, Vos, H & Timens, W 2004, 'Hypertrophic scar formation is associated with an increased number of epidermal Langerhans cells', *J Pathol*, vol. 202, no. 1, pp. 121-129.

Okuda, Y, Nakamura, H, Yoshida, K, Enomoto, H, Uyama, H, Hirotani, T, Funamoto, M, Ito, H, Everett, AD, Hada, T & Kawase, I 2003, 'Hepatoma-derived growth factor induces tumorigenesis in vivo through both direct angiogenic activity and induction of vascular endothelial growth factor', *Cancer Sci*, vol. 94, no. 12, pp. 1034-1041.

Omo-Dare, P 1975, 'Genetic studies on keloid', *J Natl Med Assoc*, vol. 67, no. 6, pp. 428-432.

Ong, CT, Khoo, YT, Mukhopadhyay, A, Do, DV, Lim, IJ, Aalami, O & Phan, TT 2007, 'mTOR as a potential therapeutic target for treatment of keloids and excessive scars', *Exp Dermatol*, vol. 16, no. 5, pp. 394-404.

Ong, CT, Khoo, YT, Tan, EK, Mukhopadhyay, A, Do, DV, Han, HC, Lim, IJ & Phan, TT 2007, 'Epithelial-mesenchymal interactions in keloid pathogenesis modulate vascular endothelial growth factor expression and secretion', *J Pathol*, vol. 211, no. 1, pp. 95-108.

Parks, WC 1999, 'Matrix metalloproteinases in repair', *Wound Repair Regen*, vol. 7, no. 6, pp. 423-432.

Pe'er, D, Regev, A, Elidan, G & Friedman, N 2001, 'Inferring subnetworks from perturbed expression profiles', *Bioinformatics*, vol. 17 Suppl 1, pp. S215-S224.

Phan, TT, Lim, IJ, Aalami, O, Lorget, F, Khoo, A, Tan, EK, Mukhopadhyay, A & Longaker, MT 2005, 'Smad3 signalling plays an important role in keloid pathogenesis via epithelial-mesenchymal interactions', *J Pathol*, vol. 207, no. 2, pp. 232-242.

Phan, TT, Lim, IJ, Bay, BH, Qi, R, Longaker, MT, Lee, ST & Huynh, H 2003, 'Role of IGF system of mitogens in the induction of fibroblast proliferation by keloid-derived keratinocytes in vitro', *Am J Physiol Cell Physiol*, vol. 284, no. 4, pp. C860-C869.

Pilewski, JM, Liu, L, Henry, AC, Knauer, AV & Feghali-Bostwick, CA 2005, 'Insulin-like growth factor binding proteins 3 and 5 are overexpressed in idiopathic pulmonary fibrosis and contribute to extracellular matrix deposition', *Am J Pathol*, vol. 166, no. 2, pp. 399-407.

Pogue-Geile, K, Geiser, JR, Shu, M, Miller, C, Wool, IG, Meisler, AI & Pipas, JM 1991, 'Ribosomal protein genes are overexpressed in colorectal cancer: isolation of a cDNA clone encoding the human S3 ribosomal protein', *Mol Cell Biol*, vol. 11, no. 8, pp. 3842-

3849.

Prockop, DJ & Kivirikko, KI 1995, 'Collagens: molecular biology, diseases, and potentials for therapy', *Annu Rev Biochem*, vol. 64, pp. 403-434.

Raghow, R 1994, 'The role of extracellular matrix in postinflammatory wound healing and fibrosis', *FASEB J*, vol. 8, no. 11, pp. 823-831.

Ren, H, Tang, X, Lee, JJ, Feng, L, Everett, AD, Hong, WK, Khuri, FR & Mao, L 2004, 'Expression of hepatoma-derived growth factor is a strong prognostic predictor for patients with early-stage non-small-cell lung cancer', *J Clin Oncol*, vol. 22, no. 16, pp. 3230-3237.

Rockwell, WB, Cohen, IK & Ehrlich, HP 1989, 'Keloids and hypertrophic scars: a comprehensive review', *Plast Reconstr Surg*, vol. 84, no. 5, pp. 827-837.

Russell, SB, Trupin, KM, Rodríguez-Eaton, S, Russell, JD & Trupin, JS 1988, 'Reduced growth-factor requirement of keloid-derived fibroblasts may account for tumor growth', *Proc Natl Acad Sci U S A*, vol. 85, no. 2, pp. 587-591.

Sachs, K, Perez, O, Pe'er, D, Lauffenburger, DA & Nolan, GP 2005, 'Causal protein-signaling networks derived from multiparameter single-cell data', *Science*, vol. 308, no. 5721, pp. 523-529.

Santoro, MM & Gaudino, G 2005, 'Cellular and molecular facets of keratinocyte reepithelization during wound healing', *Exp Cell Res*, vol. 304, no. 1, pp. 274-286.

Santucci, M, Borgognoni, L, Reali, UM & Gabbiani, G 2001, 'Keloids and hypertrophic scars of Caucasians show distinctive morphologic and immunophenotypic profiles', *Virchows Arch*, vol. 438, no. 5, pp. 457-463.

Satish, L, Lyons-Weiler, J, Hebda, PA & Wells, A 2006, 'Gene expression patterns in isolated keloid fibroblasts', *Wound Repair Regen*, vol. 14, no. 4, pp. 463-470.

Sato, M 2006, 'Upregulation of the Wnt/beta-catenin pathway induced by transforming growth factor-beta in hypertrophic scars and keloids', *Acta Derm Venereol*, vol. 86, no. 4, pp. 300-307.

Sayah, DN, Soo, C, Shaw, WW, Watson, J, Messadi, D, Longaker, MT, Zhang, X & Ting, K 1999, 'Downregulation of apoptosis-related genes in keloid tissues', *J Surg Res*, vol. 87, no. 2, pp. 209-216.

Schadt, EE, Sachs, A & Friend, S 2005, 'Embracing complexity, inching closer to reality', *Sci STKE*, vol. 2005, no. 295, pp. pe40.

Schena, M, Shalon, D, Davis, RW & Brown, PO 1995, 'Quantitative monitoring of gene

expression patterns with a complementary DNA microarray', *Science*, vol. 270, no. 5235, pp. 467-470.

Schierle, HP, Scholz, D & Lemperle, G 1997, 'Elevated levels of testosterone receptors in keloid tissue: an experimental investigation', *Plast Reconstr Surg*, vol. 100, no. 2, pp. 390-5; discussion 396.

Schulze, A & Downward, J 2001, 'Navigating gene expression using microarrays--a technology review', *Nat Cell Biol*, vol. 3, no. 8, pp. E190-E195.

Seifert, O, Bayat, A, Geffers, R, Dienus, K, Buer, J, Löfgren, S & Matussek, A 2008, 'Identification of unique gene expression patterns within different lesional sites of keloids', *Wound Repair Regen*, vol. 16, no. 2, pp. 254-265.

Seifert, O & Mrowietz, U 2009, 'Keloid scarring: bench and bedside', *Arch Dermatol Res*, vol. 301, no. 4, pp. 259-272.

Shaffer, JJ, Taylor, SC & Cook-Bolden, F 2002, 'Keloidal scars: a review with a critical look at therapeutic options', *J Am Acad Dermatol*, vol. 46, no. 2 Suppl Understanding, pp. S63-S97.

Shedden, K, Chen, W, Kuick, R, Ghosh, D, Macdonald, J, Cho, KR, Giordano, TJ, Gruber, SB, Fearon, ER, Taylor, JMG & Hanash, S 2005, 'Comparison of seven methods for producing Affymetrix expression scores based on False Discovery Rates in disease profiling data', *BMC Bioinformatics*, vol. 6, pp. 26.

Shegogue, D & Trojanowska, M 2004, 'Mammalian target of rapamycin positively regulates collagen type I production via a phosphatidylinositol 3-kinase-independent pathway', *J Biol Chem*, vol. 279, no. 22, pp. 23166-23175.

Shukla, IM, Arora, NP & Arora, MM 1975, 'Corneal keloid', *Indian J Ophthalmol*, vol. 23, no. 2, pp. 18-19.

Shulewitz, M, Soloviev, I, Wu, T, Koeppen, H, Polakis, P & Sakanaka, C 2006, 'Repressor roles for TCF-4 and Sfrp1 in Wnt signaling in breast cancer', *Oncogene*, vol. 25, no. 31, pp. 4361-4369.

Singh, S, Zheng, JJ, Peiper, SC & Mclaughlin, BJ 2001, 'Gene expression profile of ARPE-19 during repair of the monolayer', *Graefes Arch Clin Exp Ophthalmol*, vol. 239, no. 12, pp. 946-951.

Sisco, M, Kryger, ZB, O'Shaughnessy, KD, Kim, PS, Schultz, GS, Ding, XZ, Roy, NK, Dean, NM & Mustoe, TA 2008, 'Antisense inhibition of connective tissue growth factor (CTGF/CCN2) mRNA limits hypertrophic scarring without affecting wound healing in vivo', *Wound Repair Regen*, vol. 16, no. 5, pp. 661-673.

Slemp, AE & Kirschner, RE 2006, 'Keloids and scars: a review of keloids and scars, their pathogenesis, risk factors, and management', *Curr Opin Pediatr*, vol. 18, no. 4, pp. 396-402.

Slusarski, DC, Corces, VG & Moon, RT 1997, 'Interaction of Wnt and a Frizzled homologue triggers G-protein-linked phosphatidylinositol signalling', *Nature*, vol. 390, no. 6658, pp. 410-413.

Smith, CJ, Smith, JC & Finn, MC 1987, 'The possible role of mast cells (allergy) in the production of keloid and hypertrophic scarring', *J Burn Care Rehabil*, vol. 8, no. 2, pp. 126-131.

Smith, JC, Boone, BE, Opalenik, SR, Williams, SM & Russell, SB 2008, 'Gene profiling of keloid fibroblasts shows altered expression in multiple fibrosis-associated pathways', *J Invest Dermatol*, vol. 128, no. 5, pp. 1298-1310.

Smith, TJ 2005, 'Insights into the role of fibroblasts in human autoimmune diseases', *Clin Exp Immunol*, vol. 141, no. 3, pp. 388-397.

Stadelmann, WK, Digenis, AG & Tobin, GR 1998, 'Physiology and healing dynamics of chronic cutaneous wounds', *Am J Surg*, vol. 176, no. 2A Suppl, pp. 26S-38S.

Steuer, R, Kurths, J, Daub, CO, Weise, J & Selbig, J 2002, 'The mutual information: detecting and evaluating dependencies between variables', *Bioinformatics*, vol. 18 Suppl 2, pp. S231-S240.

Sylvia, CJ 2003, 'The role of neutrophil apoptosis in influencing tissue repair', *J Wound Care*, vol. 12, no. 1, pp. 13-16.

Takada, T, Yagi, Y, Maekita, T, Imura, M, Nakagawa, S, Tsao, SW, Miyamoto, K, Yoshino, O, Yasugi, T, Taketani, Y & Ushijima, T 2004, 'Methylation-associated silencing of the Wnt antagonist SFRP1 gene in human ovarian cancers', *Cancer Sci*, vol. 95, no. 9, pp. 741-744.

Tosa, M, Ghazizadeh, M, Shimizu, H, Hirai, T, Hyakusoku, H & Kawanami, O 2005, 'Global gene expression analysis of keloid fibroblasts in response to electron beam irradiation reveals the involvement of interleukin-6 pathway', *J Invest Dermatol*, vol. 124, no. 4, pp. 704-713.

Tsibris, JCM, Segars, J, Coppola, D, Mane, S, Wilbanks, GD, O'Brien, WF & Spellacy, WN 2002, 'Insights from gene arrays on the development and growth regulation of uterine leiomyomata', *Fertil Steril*, vol. 78, no. 1, pp. 114-121.

Tuan, TL & Nichter, LS 1998, 'The molecular basis of keloid and hypertrophic scar formation', *Mol Med Today*, vol. 4, no. 1, pp. 19-24.

Tuan, TL, Wu, H, Huang, EY, Chong, SSN, Laug, W, Messadi, D, Kelly, P & Le, A 2003, 'Increased plasminogen activator inhibitor-1 in keloid fibroblasts may account for their elevated collagen accumulation in fibrin gel cultures', *Am J Pathol*, vol. 162, no. 5, pp. 1579-1589.

Tumor Analysis Best Practices Working Group 2004, 'Expression profiling--best practices for data generation and interpretation in clinical trials', *Nat Rev Genet*, vol. 5, no. 3, pp. 229-237.

Urioste, SS, Arndt, KA & Dover, JS 1999, 'Keloids and hypertrophic scars: review and treatment strategies', *Semin Cutan Med Surg*, vol. 18, no. 2, pp. 159-171.

Uyama, H, Tomita, Y, Nakamura, H, Nakamori, S, Zhang, B, Hoshida, Y, Enomoto, H, Okuda, Y, Sakon, M, Aozasa, K, Kawase, I, Hayashi, N & Monden, M 2006, 'Hepatoma-derived growth factor is a novel prognostic factor for patients with pancreatic cancer', *Clin Cancer Res*, vol. 12, no. 20 Pt 1, pp. 6043-6048.

Wang, Z, Fong, KD, Phan, TT, Lim, IJ, Longaker, MT & Yang, GP 2006, 'Increased transcriptional response to mechanical strain in keloid fibroblasts due to increased focal adhesion complex formation', *J Cell Physiol*, vol. 206, no. 2, pp. 510-517.

Werner, S & Grose, R 2003, 'Regulation of wound healing by growth factors and cytokines', *Physiol Rev*, vol. 83, no. 3, pp. 835-870.

Wicke, C, Halliday, B, Allen, D, Roche, NS, Scheuenstuhl, H, Spencer, MM, Roberts, AB & Hunt, TK 2000, 'Effects of steroids and retinoids on wound healing', *Arch Surg*, vol. 135, no. 11, pp. 1265-1270.

Witt, E, Maliri, A, McGrouther, DA & Bayat, A 2008, 'RAC Activity in Keloid Disease: Comparative Analysis of Fibroblasts from Margin of Keloid to its Surrounding Normal Skin', *Eplasty*, vol. 8, pp. e19.

Wolfram, D, Tzankov, A, Pülzl, P & Piza-Katzer, H 2009, 'Hypertrophic scars and keloids--a review of their pathophysiology, risk factors, and therapeutic management', *Dermatol Surg*, vol. 35, no. 2, pp. 171-181.

Wool, IG 1996, 'Extraribosomal functions of ribosomal proteins', *Trends Biochem Sci*, vol. 21, no. 5, pp. 164-165.

Wu, RZ, Chaivorapol, C, Zheng, J, Li, H & Liang, S 2007, 'fREDUCE: detection of degenerate regulatory elements using correlation with expression', *BMC Bioinformatics*, vol. 8, pp. 399.

Wu, Y, Zhang, Q, Ann, DK, Akhondzadeh, A, Duong, HS, Messadi, DV & Le, AD 2004, 'Increased vascular endothelial growth factor may account for elevated level of plasminogen activator inhibitor-1 via activating ERK1/2 in keloid fibroblasts', *Am J*

Physiol Cell Physiol, vol. 286, no. 4, pp. C905-C912.

Xia, W, Phan, TT, Lim, IJ, Longaker, MT & Yang, GP 2004, 'Complex epithelial-mesenchymal interactions modulate transforming growth factor-beta expression in keloid-derived cells', *Wound Repair Regen*, vol. 12, no. 5, pp. 546-556.

Yagi, KI, Dafalla, AA & Osman, AA 1979, 'Does an immune reaction to sebum in wounds cause keloid scars? Beneficial effect of desensitisation', *Br J Plast Surg*, vol. 32, no. 3, pp. 223-225.

Yamamoto, S, Tomita, Y, Hoshida, Y, Morii, E, Yasuda, T, Doki, Y, Aozasa, K, Uyama, H, Nakamura, H & Monden, M 2007, 'Expression level of hepatoma-derived growth factor correlates with tumor recurrence of esophageal carcinoma', *Ann Surg Oncol*, vol. 14, no. 7, pp. 2141-2149.

Yang, ZQ, Liu, G, Bollig-Fischer, A, Haddad, R, Tarca, AL & Ethier, SP 2009, 'Methylation-associated silencing of SFRP1 with an 8p11-12 amplification inhibits canonical and non-canonical WNT pathways in breast cancers', *Int J Cancer*, vol. 125, no. 7, pp. 1613-1621.

Yasuoka, H, Zhou, Z, Pilewski, JM, Oury, TD, Choi, AMK & Feghali-Bostwick, CA 2006, 'Insulin-like growth factor-binding protein-5 induces pulmonary fibrosis and triggers mononuclear cellular infiltration', *Am J Pathol*, vol. 169, no. 5, pp. 1633-1642.

Yeh, FL, Shen, HD & Tai, HY 2009, 'Decreased production of MCP-1 and MMP-2 by keloid-derived fibroblasts', *Burns*, vol. 35, no. 3, pp. 348-351.

York, IA & Rock, KL 1996, 'Antigen processing and presentation by the class I major histocompatibility complex', *Annu Rev Immunol*, vol. 14, pp. 369-396.

Yoshida, K, Nakamura, H, Okuda, Y, Enomoto, H, Kishima, Y, Uyama, H, Ito, H, Hirasawa, T, Inagaki, S & Kawase, I 2003, 'Expression of hepatoma-derived growth factor in hepatocarcinogenesis', *J Gastroenterol Hepatol*, vol. 18, no. 11, pp. 1293-1301.

Yu, J, Smith, VA, Wang, PP, Hartemink, AJ & Jarvis, ED 2004, 'Advances to Bayesian network inference for generating causal networks from observational biological data', *Bioinformatics*, vol. 20, no. 18, pp. 3594-3603.

Zhang, Q, Wu, Y, Ann, DK, Messadi, DV, Tuan, TL, Kelly, AP, Bertolami, CN & Le, AD 2003, 'Mechanisms of hypoxic regulation of plasminogen activator inhibitor-1 gene expression in keloid fibroblasts', *J Invest Dermatol*, vol. 121, no. 5, pp. 1005-1012.

Zimmer, J, Poli, A, Andrès, E, Hanau, D, Brons, NHC & Hentges, F 2006, 'Reduced cytokine-mediated up-regulation of HLA-DR in TAP-deficient fibroblasts', *Immunol Lett*, vol. 107, no. 2, pp. 109-118.

APPENDICES

A.1 Full list of 181 genes upregulated in keloid compared to normal fibroblasts using the MAS 5.0 summarization algorithm (P< 0.05)

Fold change	Gene Symbol	Gene Title	Corrected p-value
26.2568	POSTN	periostin, osteoblast specific factor	2.51E-04
19.88896	ZIC1	Zic family member 1 (odd-paired homolog, Drosophila)	0.046789
14.57531	HOXD10	homeobox D10	0.004014
10.38634	COL15A1	collagen, type XV, alpha 1	4.25E-04
8.887936	EGR2	early growth response 2 (Krox-20 homolog, Drosophila)	0.006852
7.998279	HOXA11	homeobox A11	0.033229
7.9743	CCDC102B	coiled-coil domain containing 102B	0.021884
6.770347	KCNJ6	potassium inwardly-rectifying channel, subfamily J, member 6	0.014057
6.420784	JUP /// KRT19	junction plakoglobin /// keratin 19	0.014453
5.80834	MAP7	microtubule-associated protein 7	0.00785
5.553577	IGFBP3	insulin-like growth factor binding protein 3	0.021725
5.515989	ADRA2A	adrenergic, alpha-2A-, receptor	0.011306
5.489978	RAC2	ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)	0.049297
5.301598	CDYL	chromodomain protein, Y-like	0.039061
5.168533	NPTX1	neuronal pentraxin I	0.015499
4.974008	PMEPA1	prostate transmembrane protein, androgen induced 1	0.037054
4.940097	EFNB2	ephrin-B2	0.022704
4.889467	ATXN1	ataxin 1	0.003232
4.783432	CADM1	cell adhesion molecule 1	0.033839
4.513727	SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	0.008288
4.476735	WNT5A	wingless-type MMTV integration site family, member 5A	0.0251
4.305683	AK5	adenylate kinase 5	0.032527
4.167339	EVI2A /// EVI2B	ecotropic viral integration site 2A /// ecotropic viral integration site 2B	0.003941
4.105316	THBS1	thrombospondin 1	0.003338
4.064728	HMGCS2	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 2 (mitochondrial)	0.038054
4.031227	UNC5B	Unc-5 homolog B (C. elegans) (UNC5B), mRNA	0.015673
4.003236	GPSM2	G-protein signaling modulator 2 (AGS3-like, C. elegans)	0.014816
3.983835	FAM155A	family with sequence similarity 155, member A	0.025904

3.896551	DYSF	dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)	0.006904
3.846086	TRIB2	tribbles homolog 2 (Drosophila)	0.049297
3.807187	TOX	thymocyte selection-associated high mobility group box	5.70E-05
3.80115	MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2	0.031156
3.658315	THBS1	thrombospondin 1	0.019972
3.650868	CADM1	cell adhesion molecule 1	0.004192
3.569428	WNT5A	wingless-type MMTV integration site family, member 5A	0.006868
3.54556	THBS1	thrombospondin 1	0.026742
3.420288	SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	0.012037
3.40271	NR2F2	nuclear receptor subfamily 2, group F, member 2	0.029412
3.210168	OXTR	oxytocin receptor	0.028325
3.02816	ARMC9	armadillo repeat containing 9	0.033839
3.028158	SHMT2	serine hydroxymethyltransferase 2 (mitochondrial)	0.011685
3.005537	MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2	0.014608
2.982032	FARP1	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)	0.004134
2.942675	HCLS1	hematopoietic cell-specific Lyn substrate 1	0.012601
2.895509	GPSM2	G-protein signaling modulator 2 (AGS3-like, C. elegans)	0.022704
2.879644	FRZB	frizzled-related protein	0.015673
2.782385	MYO1D	myosin ID	0.006852
2.772433	SEPT11	CDNA FLJ37154 fis, clone BRACE2026054, highly similar to SEPTIN 2	0.002621
2.743658	COL1A1	collagen, type I, alpha 1	0.035314
2.694598	BGN	biglycan	0.045835
2.690918	SYNGR1	synaptogyrin 1	0.025029
2.662594	MGC87895 /// RPS14	similar to ribosomal protein S14 /// ribosomal protein S14	0.024206
2.658657	SYT1	synaptotagmin I	0.035314
2.586957	SEPT11	septin 11	0.025805
2.515018	TBC1D2	TBC1 domain family, member 2	0.006247
2.506182	UBL3	ubiquitin-like 3	0.019401
2.43455	SLC1A4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	0.04681
2.39163	COL5A1	collagen, type V, alpha 1	9.92E-04
2.37384	ECM1	extracellular matrix protein 1	0.039152
2.348127	MBD3	methyl-CpG binding domain protein 3	0.022235
2.344866	IPO5	importin 5	0.010591
2.331204	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	0.021725
2.273416	BMP6	bone morphogenetic protein 6	0.03454
2.271414	COL5A1	collagen, type V, alpha 1	0.046681
2.257667	PDGFRB	platelet-derived growth factor receptor, beta	0.001677

		polypeptide	
2.251143	CTSB	cathepsin B	0.012336
2.250428	MYO19	myosin XIX	0.039896
2.231361	LAMA2	laminin, alpha 2	0.005205
2.228339	IPO5	importin 5	0.010816
2.222405	FASN	fatty acid synthase	0.031733
2.21538	NAP1L1	nucleosome assembly protein 1-like 1	0.007082
2.202431	FKBP9	FK506 binding protein 9, 63 kDa	0.049297
2.189176	FHOD1	formin homology 2 domain containing 1	0.021585
2.17394	LAMA2	laminin, alpha 2	0.008295
2.168009	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	0.023943
2.160458	MLPH	melanophilin	0.032937
2.159276	GLS	glutaminase	0.044103
2.15634	KLF5	Kruppel-like factor 5 (intestinal)	0.008645
2.119878	PABPC4	poly(A) binding protein, cytoplasmic 4 (inducible form)	0.025147
2.107291	NXN	nucleoredoxin	0.004134
2.098882	LOC652607 /// PABPC1 /// PABPC3 /// PABPCP5	similar to Polyadenylate-binding protein 1 (Poly(A)-binding protein 1) (PABP 1) /// poly(A) binding protein, cytoplasmic 1 /// poly(A) binding protein, cytoplasmic 3 /// poly(A) binding protein, cytoplasmic pseudogene 5	0.022235
2.085628	FBLN2	fibulin 2	0.008295
2.065369	SACS	spastic ataxia of Charlevoix-Saguenay (sacsin)	0.009496
2.056679	PTK7	PTK7 protein tyrosine kinase 7	0.01158
2.050869	FARP1	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)	0.012556
2.02876	SPOCK1	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1	0.028002
2.009869	DUSP7	dual specificity phosphatase 7	0.004014
1.987354	MARCKS	myristoylated alanine-rich protein kinase C substrate	0.025545
1.971692	NUAK1	NUAK family, SNF1-like kinase, 1	0.006441
1.95389	POLR1D	polymerase (RNA) I polypeptide D, 16kDa	0.041823
1.940716	CRTAP	cartilage associated protein	0.043513
1.934144	LAMA2	laminin, alpha 2	0.002955
1.918422	ATXN1	ataxin 1	0.028403
1.910812	GSTM3	glutathione S-transferase mu 3 (brain)	0.006366
1.893329	ANP32B	acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	0.002621
1.885341	EEF1D	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)	0.005205
1.877402	LOC285053 /// LOC390354 /// RPL18A	similar to ribosomal protein L18a /// ribosomal protein L18a pseudogene /// ribosomal protein L18a	0.021725
1.867147	CCNB1IP1	cyclin B1 interacting protein 1	0.031156
1.865347	LOC644191 /// LOC728937 /// RPS26	similar to hCG15685 /// similar to 40S ribosomal protein S26 /// ribosomal protein S26	0.025519
1.862667	RPL35	ribosomal protein L35	0.025168

1.861686	IPO5	importin 5	0.002103
1.853484	IPO5	importin 5	0.02642
1.825099	CTSB	cathepsin B	0.018044
1.822173	PTPRG	protein tyrosine phosphatase, receptor type, G	0.027471
1.812249	PARVB	parvin, beta	0.006852
1.811397	COL1A1	collagen, type I, alpha 1	0.020984
1.806129	GARS	glycyl-tRNA synthetase	0.03012
1.798827	SHMT2	serine hydroxymethyltransferase 2 (mitochondrial)	0.001636
1.790564	GPR137B	G protein-coupled receptor 137B	0.034176
1.780408	RPL10	ribosomal protein L10	0.019277
1.765466	PTPRK	protein tyrosine phosphatase, receptor type, K	0.047404
1.764005	RPS16	ribosomal protein S16	0.048597
1.753873	AHCYL1	S-adenosylhomocysteine hydrolase-like 1	0.031368
1.74401	MBNL1	muscleblind-like (Drosophila)	0.040962
1.735983	NONO	non-POU domain containing, octamer-binding	0.028002
1.729728	CALHM2	calcium homeostasis modulator 2	0.012441
1.717712	PPP3CA	protein phosphatase 3 (formerly 2B), catalytic subunit, alpha isoform	0.037703
1.717474	GPR153	G protein-coupled receptor 153	0.036174
1.704002	TST	thiosulfate sulfurtransferase (rhodanese)	0.01541
1.702673	DLEU1	deleted in lymphocytic leukemia 1 (non-protein coding)	0.009449
1.700546	ATF4	activating transcription factor 4 (tax-responsive enhancer element B67)	0.017043
1.698355	DOCK1	dedicator of cytokinesis 1	0.006852
1.696983	EIF3A	eukaryotic translation initiation factor 3, subunit A	0.006852
1.695664	PQBP1	polyglutamine binding protein 1	0.047404
1.690349	OSBPL3	oxysterol binding protein-like 3	0.046651
1.682196	EIF3D	eukaryotic translation initiation factor 3, subunit D	0.044103
1.669918	TMEM2	transmembrane protein 2	0.011306
1.667722	C17orf81	chromosome 17 open reading frame 81	0.033839
1.661295	C14orf139	chromosome 14 open reading frame 139	0.042746
1.660296	SPARC	secreted protein, acidic, cysteine-rich (osteonectin)	0.030453
1.660139	PNN	pinin, desmosome associated protein	0.029702
1.658402	MTFR1	mitochondrial fission regulator 1	0.016285
1.645023	RPS9	ribosomal protein S9	0.025283
1.643602	S100A13	S100 calcium binding protein A13	0.006868
1.637327	C19orf2	chromosome 19 open reading frame 2	0.036054
1.636471	SARS	seryl-tRNA synthetase	0.018299
1.636263	RPL13	ribosomal protein L13	0.010804
1.623621	FAT1	FAT tumor suppressor homolog 1 (Drosophila)	0.039953
1.621903	COL17A1	collagen type XVII alpha I	0.006247
1.61924	SOS2	son of sevenless homolog 2 (Drosophila)	0.033713
1.616134	NQO2	NAD(P)H dehydrogenase, quinone 2	0.033839
1.602853	RPS2	ribosomal protein S2	0.040053
1.602017	AP2A2	adaptor-related protein complex 2, alpha 2 subunit	0.012601
1.601743	TK2	thymidine kinase 2, mitochondrial	0.040053

1.599825	MEMO1	mediator of cell motility 1	0.010405
1.597647	RPL4	ribosomal protein L4	0.032078
1.59623	RPS8	ribosomal protein S8	0.037349
1.553895	PARVB	parvin, beta	0.032078
1.553142	HNRNPA0	heterogeneous nuclear ribonucleoprotein A0	0.033839
1.549388	EPHB4	EPH receptor B4	0.016501
1.538817	POLR1E	polymerase (RNA) I polypeptide E, 53kDa	0.021725
1.531203	SERP1	stress-associated endoplasmic reticulum protein 1	0.032078
1.529446	DNAJB6	DnaJ (Hsp40) homolog, subfamily B, member 6	0.033283
1.523152	RPL13	ribosomal protein L13	0.021854
1.518518	SEC63	SEC63 homolog (S. cerevisiae)	0.011005
1.509025	HNRNPA3	heterogeneous nuclear ribonucleoprotein A3	0.005384
1.484414	BTF3	basic transcription factor 3	0.048855
1.474366	RPS6	ribosomal protein S6	0.006198
1.458746	GRB10	growth factor receptor-bound protein 10	0.004053
1.457511	FBXL5	F-box and leucine-rich repeat protein 5	0.046681
1.449214	ACAD8	acyl-Coenzyme A dehydrogenase family, member 8	0.021854
1.44043	DCTD	dCMP deaminase	0.024538
1.43772	APRT	adenine phosphoribosyltransferase	0.048855
1.434125	PDS5A	PDS5, regulator of cohesion maintenance, homolog A (S. cerevisiae)	0.048855
1.42483	HEBP1	heme binding protein 1	0.01989
1.419617	SUV420H1	suppressor of variegation 4-20 homolog 1 (Drosophila)	0.031156
1.415472	RPL8	ribosomal protein L8	0.024538
1.407764	RECQL	RecQ protein-like (DNA helicase Q1-like)	0.031862
1.404108	KDELR2	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 2	0.047112
1.40236	MPRIP	myosin phosphatase Rho interacting protein	0.01541
1.392988	ATP5E	ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit	0.049297
1.375108	LUC7L	LUC7-like (S. cerevisiae)	0.036591
1.37272	ARMCX3	armadillo repeat containing, X-linked 3	0.048597
1.371258	SERP1	stress-associated endoplasmic reticulum protein 1	0.035533
1.365867	PPP2R3A	protein phosphatase 2 (formerly 2A), regulatory subunit B'', alpha	0.023437
1.360637	RPL13	ribosomal protein L13	0.0227
1.35391	STK24	serine/threonine kinase 24 (STE20 homolog, yeast)	0.049297
1.305478	EIF4H	eukaryotic translation initiation factor 4H	0.033851
1.275109	COX4NB	COX4 neighbor	0.045835
1.264861	FBXO34	F-box protein 34	0.026786
1.25246	USP22	ubiquitin specific peptidase 22	0.038322

A.2 Full list of 290 genes downregulated in keloid compared to normal fibroblasts using the MAS 5.0 summarization algorithm (P < 0.05)

Fold change	Gene Symbol	Gene Title	Corrected p-value
77.61498	CXCL6	chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)	1.48E-05
73.48209	CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	7.93E-05
67.89986	IL8	interleukin 8	0.006303
64.39615	CXCL11	chemokine (C-X-C motif) ligand 11	0.006303
49.49322	HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1	1.98E-06
41.41703	CCL5	chemokine (C-C motif) ligand 5	0.005142
39.93947	CXCL2	chemokine (C-X-C motif) ligand 2	0.002401
32.48277	RARRES1	retinoic acid receptor responder (tazarotene induced) 1	0.004841
29.77878	RSAD2	radical S-adenosyl methionine domain containing 2	0.019981
27.68656	PLA2G2A	phospholipase A2, group IIA (platelets, synovial fluid)	0.001149
27.15919	C2 /// CFB	complement component 2 /// complement factor B	2.15E-05
27.04305	CXCL5	chemokine (C-X-C motif) ligand 5	2.44E-04
26.4216	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	0.008288
26.13762	CXCL3	chemokine (C-X-C motif) ligand 3	0.005384
23.80846	IL32	interleukin 32	0.001636
23.43326	CP	ceruloplasmin (ferroxidase)	0.003941
23.40176	CXCL10	chemokine (C-X-C motif) ligand 10	0.032337
21.83553	CHI3L2	chitinase 3-like 2	5.70E-04
21.53614	IDO1	indoleamine 2,3-dioxygenase 1	0.004053
19.30517	NTRK2	neurotrophic tyrosine kinase, receptor, type 2	0.011306
15.52373	C3	complement component 3	2.51E-04
15.49821	SLC39A8	solute carrier family 39 (zinc transporter), member 8	3.85E-07
14.28212	G0S2	G0/G1switch 2	0.003903
14.04543	OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa	0.039061
13.35598	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	0.016414
13.23721	CCL8	chemokine (C-C motif) ligand 8	0.024538
11.90575	SLC39A8	solute carrier family 39 (zinc transporter), member 8	4.73E-05
11.36108	GCH1	GTP cyclohydrolase 1	2.44E-04
10.95266	CCL5	chemokine (C-C motif) ligand 5	0.010804
10.86976	SLC19A3	solute carrier family 19, member 3	0.031156
10.20459	HERC5	hect domain and RLD 5	0.005346
10.02452	IL6	interleukin 6 (interferon, beta 2)	5.25E-04
9.336419	NRCAM	neuronal cell adhesion molecule	1.22E-05
9.271726	CCL2	chemokine (C-C motif) ligand 2	2.18E-05
8.931117	ABCA8	ATP-binding cassette, sub-family A (ABC1),	0.006852

		member 8	
8.870272	SFRP1	secreted frizzled-related protein 1	0.013944
8.768786	TMEM100	transmembrane protein 100	0.008604
8.582171	HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	9.92E-04
8.130204	TLR2	toll-like receptor 2	0.011223
8.070362	SOD2	superoxide dismutase 2, mitochondrial	0.001796
7.84843	CTSS	cathepsin S	2.95E-04
7.806789	IFI30	interferon, gamma-inducible protein 30	0.004841
7.653281	MAOB	monoamine oxidase B	0.040797
7.647369	DTNA	dystrobrevin, alpha	0.005869
7.607011	HERC6	hect domain and RLD 6	0.031862
7.487806	TLR3	toll-like receptor 3	0.031156
7.243977	RARRES3	retinoic acid receptor responder (tazarotene induced) 3	0.004841
7.072357	SOD2	superoxide dismutase 2, mitochondrial	0.004841
7.021463	TNFAIP3	tumor necrosis factor, alpha-induced protein 3	1.48E-05
6.598767	CA12	carbonic anhydrase XII	5.70E-05
6.533486	CX3CL1	chemokine (C-X3-C motif) ligand 1	0.039558
6.41113	FGL2	fibrinogen-like 2	0.046681
6.335001	CTSS	cathepsin S	0.023641
6.331704	IFIT2	interferon-induced protein with tetratricopeptide repeats 2	0.049297
6.29937	LRRN3	leucine rich repeat neuronal 3	2.64E-04
6.179313	IFIT3	interferon-induced protein with tetratricopeptide repeats 3	0.018974
6.154707	BIRC3	baculoviral IAP repeat-containing 3	0.013996
6.106007	CA12	carbonic anhydrase XII	0.004215
6.092712	CLGN	calmegin	0.008363
6.0468	TRPA1	transient receptor potential cation channel, subfamily A, member 1	0.022447
5.951259	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa	0.013892
5.838885	TNFAIP2	tumor necrosis factor, alpha-induced protein 2	1.38E-04
5.792193	ISG20	interferon stimulated exonuclease gene 20kDa	0.048855
5.715185	HPSE	heparanase	0.036366
5.679222	ABI3BP	ABI family, member 3 (NESH) binding protein	0.036366
5.661747	CA12	carbonic anhydrase XII	4.73E-05
5.640415	PTGES	prostaglandin E synthase	0.004841
5.429267	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa	0.021585
5.347407	IGF1	insulin-like growth factor 1 (somatomedin C)	0.048855
5.341403	HLA-DRB1	MHC class II HLA-DRB3 mRNA (HLA-DRB3*01012 allele)	0.021083
5.1838	SNX10	sorting nexin 10	0.041947
5.170646	CGREF1	cell growth regulator with EF-hand domain 1	0.024396
5.018916	NAMPT	nicotinamide phosphoribosyltransferase	0.013996
4.897188	SLC39A8	solute carrier family 39 (zinc transporter), member 8	5.70E-05
4.840101	CA12	carbonic anhydrase XII	3.62E-04

4.819873	STAT4	signal transducer and activator of transcription 4	0.005418
4.81897	CA12	carbonic anhydrase XII	2.51E-04
4.798669	LRRN3	leucine rich repeat neuronal 3	4.46E-04
4.769258	MT1X	metallothionein 1X	0.035224
4.76155	NAMPT	nicotinamide phosphoribosyltransferase	0.002078
4.755867	MT1M	metallothionein 1M	0.008363
4.618283	WTAP	Wilms tumor 1 associated protein	1.48E-05
4.604176	MT1P2	metallothionein 1 pseudogene 2	0.02436
4.566699	PRRG4	proline rich Gla (G-carboxyglutamic acid) 4 (transmembrane)	0.039061
4.551031	ELF3	E74-like factor 3 (ets domain transcription factor, epithelial-specific)	0.004657
4.543112	SLC16A4	solute carrier family 16, member 4 (monocarboxylic acid transporter 5)	0.023221
4.360824	HLA-DRB1 /// HLA-DRB2 /// HLA-DRB3 /// HLA-DRB4 /// HLA-DRB5 /// LOC100133484 /// LOC100133661 /// LOC100133811 /// LOC730415 /// RNASE2 /// ZNF749	major histocompatibility complex, class II, DR beta 1 /// major histocompatibility complex, class II, DR beta 2 (pseudogene) /// major histocompatibility complex, class II, DR beta 3 /// major histocompatibility complex, class II, DR beta 4 /// major histocompatibility complex, class II, DR beta 5 /// similar to Major histocompatibility complex, class II, DR beta 4 /// similar to HLA class II histocompatibility antigen, DR-W53 beta chain /// similar to hCG1992647 /// hypothetical protein LOC730415 /// ribonuclease, RNase A family, 2 (liver, eosinophil-derived neurotoxin) /// zinc finger protein 749	0.041418
4.342693	TNFAIP3	tumor necrosis factor, alpha-induced protein 3	4.97E-04
4.308981	RRAD	Ras-related associated with diabetes	0.011306
4.278844	SOD2	superoxide dismutase 2, mitochondrial	2.51E-04
4.184589	MT1E	metallothionein 1E	0.018044
4.177286	ZC3H12A	zinc finger CCCH-type containing 12A	0.02642
4.14845	HLA-DRB1 /// HLA-DRB2 /// HLA-DRB3 /// HLA-DRB4 /// HLA-DRB5 /// LOC100133484 /// LOC100133661 /// LOC100133811 /// LOC730415 /// RNASE2 /// ZNF749	major histocompatibility complex, class II, DR beta 1 /// major histocompatibility complex, class II, DR beta 2 (pseudogene) /// major histocompatibility complex, class II, DR beta 3 /// major histocompatibility complex, class II, DR beta 4 /// major histocompatibility complex, class II, DR beta 5 /// similar to Major histocompatibility complex, class II, DR beta 4 /// similar to HLA class II histocompatibility antigen, DR-W53 beta chain /// similar to hCG1992647 /// hypothetical protein LOC730415 /// ribonuclease, RNase A family, 2 (liver, eosinophil-derived neurotoxin) /// zinc finger protein 749	0.046363
4.127018	ICAM1	intercellular adhesion molecule 1	5.70E-05
4.100276	DENND2A	DENN/MADD domain containing 2A	0.021083
4.048781	SLC39A14	solute carrier family 39 (zinc transporter), member 14	0.002948
4.039061	DENND2A	DENN/MADD domain containing 2A	0.014614
4.038336	MT1F	metallothionein 1F	0.009128
4.00928	AKR1B1	aldo-keto reductase family 1, member B1 (aldose	9.92E-04

		reductase)	
3.957591	SNCA	synuclein, alpha (non A4 component of amyloid precursor)	0.019277
3.911683	WWC1	WW and C2 domain containing 1	0.022704
3.886122	ICAM1	intercellular adhesion molecule 1	1.48E-05
3.853862	WTAP	Wilms tumor 1 associated protein	2.18E-05
3.810552	SLC15A3	solute carrier family 15, member 3	0.006868
3.690421	IFI35	interferon-induced protein 35	0.033983
3.648437	FIGF	c-fos induced growth factor (vascular endothelial growth factor D)	0.043452
3.63322	MT1E /// MT1H /// MT1M /// MT1P2	metallothionein 1E /// metallothionein 1H /// metallothionein 1M /// metallothionein 1 pseudogene 2	0.002621
3.570853	AMPD3	adenosine monophosphate deaminase (isoform E)	0.005508
3.568043	MT2A	metallothionein 2A	0.022895
3.567944	LAP3	leucine aminopeptidase 3	0.017355
3.566436	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	0.002401
3.555502	MARCH3	membrane-associated ring finger (C3HC4) 3	0.003027
3.515179	PRND	prion protein 2 (dublet)	0.023981
3.490871	HTR2A	5-hydroxytryptamine (serotonin) receptor 2A	0.006852
3.436509	CLU	clusterin	0.036591
3.41227	STEAP1	six transmembrane epithelial antigen of the prostate 1	0.002103
3.410262	IFITM1	interferon induced transmembrane protein 1 (9-27)	0.014614
3.398285	PPARG	peroxisome proliferator-activated receptor gamma	0.025147
3.384179	GFRA1	GDNF family receptor alpha 1	0.021725
3.362342	NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	0.003195
3.348567	MMD	monocyte to macrophage differentiation-associated	0.004578
3.33894	NFE2L3	nuclear factor (erythroid-derived 2)-like 3	0.047112
3.329882	MT1X	metallothionein 1X	0.032078
3.321269	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	5.85E-05
3.299193	PTGFR	prostaglandin F receptor (FP)	0.004578
3.268607	MT1G	metallothionein 1G	0.036366
3.249937	NR4A3	nuclear receptor subfamily 4, group A, member 3	0.024206
3.216968	PLSCR1	phospholipid scramblase 1	0.026199
3.160602	LGALS9	lectin, galactoside-binding, soluble, 9	0.028025
3.143586	PPP1R12B	protein phosphatase 1, regulatory (inhibitor) subunit 12B	0.045835
3.135891	UCHL1	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	2.51E-04
3.114003	CHEK2	CHK2 checkpoint homolog (S. pombe)	0.012037
3.099276	MT1F	metallothionein 1F	0.003262
3.092455	PTGES	prostaglandin E synthase	0.031156
3.076524	CCR1	chemokine (C-C motif) receptor 1	0.022235
3.03837	EDNRB	endothelin receptor type B	3.35E-04
2.998312	IL15RA	interleukin 15 receptor, alpha	0.011685

2.938915	STAT1	signal transducer and activator of transcription 1, 91kDa	0.022235
2.927115	SMC2	structural maintenance of chromosomes 2	0.002621
2.921418	NME5	non-metastatic cells 5, protein expressed in (nucleoside-diphosphate kinase)	0.044101
2.915037	SAMHD1	SAM domain and HD domain 1	0.021854
2.833072	CD82	CD82 molecule	0.003571
2.778108	ABLIM1	actin binding LIM protein 1	0.009496
2.770222	LAMB3	laminin, beta 3	0.020989
2.747634	MT1F	metallothionein 1F	0.002955
2.737904	HLA-F	major histocompatibility complex, class I, F	0.003416
2.734728	CFLAR	CASP8 and FADD-like apoptosis regulator	0.006366
2.734047	SIRPA	signal-regulatory protein alpha	5.35E-04
2.71983	PDLIM4	PDZ and LIM domain 4	0.026742
2.702095	HLA-C	major histocompatibility complex, class I, C	0.008699
2.670183	SLC25A28	solute carrier family 25, member 28	5.80E-04
2.65614	AK3L1 /// AK3L2	adenylate kinase 3-like 1 /// adenylate kinase 3-like 2	0.024767
2.648943	HLA-DMA	major histocompatibility complex, class II, DM alpha	0.038585
2.638515	TAPBPL	TAP binding protein-like	9.67E-04
2.607675	NOVA1	neuro-oncological ventral antigen 1	0.036488
2.566287	TAPBPL	TAP binding protein-like	8.05E-04
2.557721	CFLAR	CASP8 and FADD-like apoptosis regulator	0.016501
2.556034	DTNA	dystrobrevin, alpha	5.70E-05
2.552846	PDPN	podoplanin	0.008288
2.544963	LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein	0.009862
2.528988	TDRD7	tudor domain containing 7	0.01541
2.501654	PSME2	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	0.004147
2.50114	IRAK3	interleukin-1 receptor-associated kinase 3	0.01197
2.484067	SLC1A1	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	0.03058
2.461875	CFLAR	CASP8 and FADD-like apoptosis regulator	0.041418
2.458896	CFLAR	CASP8 and FADD-like apoptosis regulator	0.003262
2.452654	HIST1H2BD	histone cluster 1, H2bd	0.006247
2.44606	PROCR	protein C receptor, endothelial (EPCR)	0.004583
2.442194	CFLAR	CASP8 and FADD-like apoptosis regulator	0.005876
2.431164	CYB5A	cytochrome b5 type A (microsomal)	0.036591
2.425255	FILIP1L	filamin A interacting protein 1-like	0.003195
2.424902	PSTPIP2	proline-serine-threonine phosphatase interacting protein 2	0.040425
2.421571	HLA-F	major histocompatibility complex, class I, F	0.002401
2.408101	CACNA1A	calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	0.025861
2.38139	SP100	SP100 nuclear antigen	0.046017
2.357674	CFLAR	CASP8 and FADD-like apoptosis regulator	0.006124
2.347081	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin	2.44E-04

		beta 2)	
2.331309	PSMB10	proteasome (prosome, macropain) subunit, beta type, 10	0.049297
2.316866	RNF114	ring finger protein 114	0.006302
2.296515	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	0.006778
2.289964	PSMB8	proteasome (prosome, macropain) subunit, beta type, 8 (large multifunctional peptidase 7)	0.039558
2.28674	CFLAR	CASP8 and FADD-like apoptosis regulator	0.011567
2.254133	CLIC2	chloride intracellular channel 2	0.024261
2.246542	DRAM	damage-regulated autophagy modulator	9.40E-04
2.238959	CFLAR	CASP8 and FADD-like apoptosis regulator	0.001636
2.231546	HLA-G	major histocompatibility complex, class I, G	0.006441
2.227878	FLJ22662	hypothetical protein FLJ22662	7.60E-04
2.214319	RNF114	ring finger protein 114	0.007101
2.213672	ICAM1	intercellular adhesion molecule 1	0.00542
2.21116	HLA-G	major histocompatibility complex, class I, G	0.002621
2.193658	GFPT2	glutamine-fructose-6-phosphate transaminase 2	0.036153
2.190023	LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	0.004192
2.182933	HLA-B	major histocompatibility complex, class I, B	5.70E-05
2.166235	HTATIP2	HIV-1 Tat interactive protein 2, 30kDa	0.015877
2.153668	MICALL2	MICAL-like 2	0.004841
2.148729	ZMYM6	zinc finger, MYM-type 6	0.001636
2.143614	SPARCL1	SPARC-like 1 (hevin)	0.038945
2.137008	FAM13A1	family with sequence similarity 13, member A1	0.018784
2.133692	SIRPA	signal-regulatory protein alpha	3.77E-04
2.131292	HLA-G	major histocompatibility complex, class I, G	1.63E-04
2.120025	CYB5A	cytochrome b5 type A (microsomal)	0.010129
2.117058	TAPBP	TAP binding protein (tapasin)	0.003416
2.116495	AKAP7	A kinase (PRKA) anchor protein 7	0.039896
2.111383	PON2	paraoxonase 2	0.004134
2.100947	NT5E	5'-nucleotidase, ecto (CD73)	2.51E-04
2.076569	MDM2	Mdm2 p53 binding protein homolog (mouse)	0.043542
2.065661	CNDP2	CNDP dipeptidase 2 (metallopeptidase M20 family)	0.039558
2.047671	TRIM38	tripartite motif-containing 38	0.01996
2.042615	HLA-B	major histocompatibility complex, class I, B	7.91E-04
2.026807	CFLAR	CASP8 and FADD-like apoptosis regulator	0.035314
2.016378	FSTL3	follistatin-like 3 (secreted glycoprotein)	0.032078
2.016181	KIAA0391 /// PSMA6	KIAA0391 /// proteasome (prosome, macropain) subunit, alpha type, 6	0.006852
1.989287	TRIM38	tripartite motif-containing 38	0.017295
1.982656	ATRIP /// TREX1	ATR interacting protein /// three prime repair exonuclease 1	0.047112
1.961992	GLIPR1	GLI pathogenesis-related 1	0.011306
1.959727	NFKBIE	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon	0.022704
1.954683	C10orf26	chromosome 10 open reading frame 26	0.003262

1.951621	LY6E	lymphocyte antigen 6 complex, locus E	0.02642
1.934228	PANX1	pannexin 1	0.022964
1.92549	TBC1D9	TBC1 domain family, member 9 (with GRAM domain)	8.91E-04
1.916083	ATOX1	ATX1 antioxidant protein 1 homolog (yeast)	0.020647
1.910365	CSF1	colony stimulating factor 1 (macrophage)	0.023957
1.896993	TNIP1	TNFAIP3 interacting protein 1	5.80E-04
1.879147	LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	0.009496
1.869223	RELB	v-rel reticuloendotheliosis viral oncogene homolog B	0.017671
1.861563	BNIP3	BCL2/adenovirus E1B 19kDa interacting protein 3	0.021585
1.858279	BEX4	brain expressed, X-linked 4	0.00786
1.848508	ACP2	acid phosphatase 2, lysosomal	0.028017
1.846391	VPS13D	vacuolar protein sorting 13 homolog D (S. cerevisiae)	0.006904
1.843842	TNFSF12 /// TNFSF12-TNFSF13 /// TNFSF13	tumor necrosis factor (ligand) superfamily, member 12 /// TNFSF12-TNFSF13 readthrough transcript /// tumor necrosis factor (ligand) superfamily, member 13	0.048855
1.814648	IRF2	interferon regulatory factor 2	0.008169
1.813575	PON2	paraoxonase 2	0.008863
1.809137	TRIM38	tripartite motif-containing 38	0.025296
1.799581	BTG3	BTG family, member 3	0.024396
1.798085	HLA-C	major histocompatibility complex, class I, C	1.18E-04
1.795796	HLA-A /// HLA-A29.1 /// HLA-B /// HLA-G /// HLA-H /// HLA-J	major histocompatibility complex, class I, A /// major histocompatibility complex class I HLA-A29.1 /// major histocompatibility complex, class I, B /// major histocompatibility complex, class I, G /// major histocompatibility complex, class I, H (pseudogene) /// major histocompatibility complex, class I, J (pseudogene)	0.018539
1.795212	NNMT	nicotinamide N-methyltransferase	0.025147
1.795001	IFNGR1	interferon gamma receptor 1	0.025147
1.790361	CDC42EP4	CDC42 effector protein (Rho GTPase binding) 4	0.049112
1.773213	HLA-C	major histocompatibility complex, class I, C	0.003262
1.76255	SLC30A1	Hbc647 mRNA sequence	0.006213
1.760506	STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)	0.008288
1.752857	BASP1	brain abundant, membrane attached signal protein 1	0.014614
1.744929	RFTN1	raftlin, lipid raft linker 1	0.022704
1.743621	RAB20	RAB20, member RAS oncogene family	0.022704
1.741601	BTG3	BTG family, member 3	0.007903
1.734837	IFNGR1	interferon gamma receptor 1	0.002401
1.726347	FTH1	ferritin, heavy polypeptide 1	0.019981
1.705262	DNAJA1	DnaJ (Hsp40) homolog, subfamily A, member 1	0.020755
1.69253	STBD1	starch binding domain 1	0.049297
1.691625	HLA-C	major histocompatibility complex, class I, C	4.98E-04
1.690892	PSME1	proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	0.032811

1.678448	HLA-B /// MICA	major histocompatibility complex, class I, B /// MHC class I polypeptide-related sequence A	0.004053
1.671204	RFX5	regulatory factor X, 5 (influences HLA class II expression)	0.037349
1.65795	HSPA4L	heat shock 70kDa protein 4-like	0.024306
1.656616	P4HA2	prolyl 4-hydroxylase, alpha polypeptide II	0.049297
1.648751	CD59	CD59 molecule, complement regulatory protein	0.008363
1.647957	NNMT	nicotinamide N-methyltransferase	0.019674
1.641402	DNAJA1	HDJ2 protein	0.014305
1.606447	NOC3L	nucleolar complex associated 3 homolog (S. cerevisiae)	0.022447
1.597539	TBCC	tubulin folding cofactor C	0.008288
1.584331	TMEM22	transmembrane protein 22	0.004657
1.579374	IFITM3	interferon induced transmembrane protein 3 (1-8U)	0.008288
1.554896	DFNA5	deafness, autosomal dominant 5	0.047112
1.545585	C15orf24	chromosome 15 open reading frame 24	0.008288
1.545367	LOC284889 /// MIF	hypothetical protein LOC284889 /// macrophage migration inhibitory factor (glycosylation-inhibiting factor)	0.044672
1.541742	ENDOD1	endonuclease domain containing 1	0.040086
1.540308	BTN2A2	butyrophilin, subfamily 2, member A2	0.001322
1.53979	USP25	ubiquitin specific peptidase 25	0.033598
1.520112	HLA-A	major histocompatibility complex, class I, A	0.01797
1.50069	SGCB	sarcoglycan, beta (43kDa dystrophin-associated glycoprotein)	0.012635
1.487067	DDX18	DEAD (Asp-Glu-Ala-Asp) box polypeptide 18	0.012601
1.482268	OPTN	optineurin	0.043452
1.481941	CD59	CD59 molecule, complement regulatory protein	0.009636
1.481617	TRAF3	TNF receptor-associated factor 3	0.022704
1.470603	NDUFA9	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 9, 39kDa	0.048855
1.468841	RBX1	ring-box 1	0.011306
1.462896	PEX12	peroxisomal biogenesis factor 12	0.020864
1.455167	MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	0.01541
1.450561	ZCCHC6	zinc finger, CCHC domain containing 6	0.011728
1.397826	MPZL1	myelin protein zero-like 1	0.047736
1.365859	LARP7	La ribonucleoprotein domain family, member 7	0.049297
1.355799	ALAS1	aminolevulinate, delta-, synthase 1	0.021863
1.347957	CSTB	cystatin B (stefin B)	0.043939
1.338753	PLS3	plastin 3 (T isoform)	0.049297
1.260726	CHMP4A	chromatin modifying protein 4A	0.006124
1.258344	FTH1	ferritin, heavy polypeptide 1	0.004014
1.244612	BUD31	BUD31 homolog (S. cerevisiae)	0.006352
1.229484	B2M	beta-2-microglobulin	0.016094
1.190801	TANK	TRAF family member-associated NFKB activator	0.007903

A.3 Full list of 86 genes upregulated in keloid compared to normal fibroblasts using the RMA summarization algorithm (P < 0.05)

Fold change	Gene Symbol	Gene Title	Corrected p-value
18.0305	POSTN	periostin, osteoblast specific factor	0.006498
5.40679	IGFBP3	insulin-like growth factor binding protein 3	0.024325
3.959037	COL15A1	collagen, type XV, alpha 1	0.028774
3.21548	SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	0.013197
3.042947	SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	0.035137
2.726196	CADM1	cell adhesion molecule 1	0.001646
2.631936	ATXN1	ataxin 1	0.020822
2.531559	FARP1	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)	0.045664
2.375326	MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2	0.037055
2.176466	ECM1	extracellular matrix protein 1	0.047073
2.129065	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	0.019936
2.053115	KCNJ6	potassium inwardly-rectifying channel, subfamily J, member 6	0.021022
1.990826	TBC1D2	TBC1 domain family, member 2	0.048158
1.985587	CADM1	cell adhesion molecule 1	0.046985
1.959726	NXN	nucleoredoxin	0.002448
1.906954	MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2	0.017212
1.897066	COL1A1	collagen, type I, alpha 1	0.008827
1.878355	PDGFRB	platelet-derived growth factor receptor, beta polypeptide	0.002791
1.869649	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	0.046273
1.865537	GPSM2	G-protein signaling modulator 2 (AGS3-like, C. elegans)	0.04761
1.852414	LOC644191 /// LOC728937 /// RPS26	similar to hCG15685 /// similar to 40S ribosomal protein S26 /// ribosomal protein S26	0.038141
1.84265	CTSB	cathepsin B	0.018933
1.836179	ODZ3	odz, odd Oz/ten-m homolog 3 (Drosophila)	6.59E-04
1.835598	JUP /// KRT19	junction plakoglobin /// keratin 19	0.02721
1.807992	LAMA2	laminin, alpha 2	0.006566
1.796024	FHOD1	formin homology 2 domain containing 1	0.014803
1.764263	CTDSPL	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like	0.018992
1.749725	SHMT2	serine hydroxymethyltransferase 2 (mitochondrial)	0.038598

1.747925	HDLBP	high density lipoprotein binding protein	0.037741
1.730545	COL5A3	collagen, type V, alpha 3	0.042118
1.719725	PTK7	PTK7 protein tyrosine kinase 7	0.041423
1.686518	NONO	non-POU domain containing, octamer-binding	0.017365
1.676297	NT5DC2	5'-nucleotidase domain containing 2	0.044871
1.667339	ATF4	activating transcription factor 4 (tax-responsive enhancer element B67)	0.004086
1.664031	LOC100130624	hypothetical LOC100130624	0.039839
1.634808	FAM155A	family with sequence similarity 155, member A	0.027042
1.627513	PQBP1	polyglutamine binding protein 1	0.003057
1.615317	RPL13	ribosomal protein L13	1.97E-04
1.612627	RPS8	ribosomal protein S8	0.023227
1.597447	HOXA11	homeobox A11	0.037859
1.58056	MGC87895 /// RPS14	similar to ribosomal protein S14 /// ribosomal protein S14	0.026445
1.568977	RPL3	ribosomal protein L3	0.024636
1.562242	EIF1	eukaryotic translation initiation factor 1	0.038827
1.544172	RPL13	ribosomal protein L13	7.69E-04
1.544152	RPL8	ribosomal protein L8	0.001713
1.531647	LOC642741	similar to ribosomal protein L3	0.001728
1.521308	RPS5	ribosomal protein S5	0.010015
1.51899	C20orf149	chromosome 20 open reading frame 149	0.047073
1.513227	EPHB3	EPH receptor B3	0.037648
1.499035	RPL13	ribosomal protein L13	0.002784
1.469517	RPL10L	ribosomal protein L10-like	0.030904
1.468484	RPL3	ribosomal protein L3	0.015644
1.459514	DBN1	drebrin 1	0.021666
1.456908	RPL3	ribosomal protein L3	0.025091
1.443996	RPL10	ribosomal protein L10	0.018933
1.441773	PARVB	parvin, beta	1.48E-04
1.436187	RPS16	ribosomal protein S16	0.004886
1.43166	SHMT2	serine hydroxymethyltransferase 2 (mitochondrial)	0.002008
1.425044	RPLP2	ribosomal protein, large, P2	0.008827
1.417303	SLC1A4	solute carrier family 1 (glutamate/neutral amino acid transporter), member 4	0.046909
1.415545	RPL3	ribosomal protein L3	0.016536
1.413611	YBX1	Y box binding protein 1	0.040024
1.408501	RPL13	ribosomal protein L13	7.69E-04
1.405121	RPS6	ribosomal protein S6	0.043219
1.403469	PPP2R3A	protein phosphatase 2 (formerly 2A), regulatory subunit B'', alpha	0.011736
1.386019	EIF3B	eukaryotic translation initiation factor 3, subunit B	0.020335
1.379153	RPS17L4	ribosomal protein S17-like 4	0.04918
1.360253	RPS2	ribosomal protein S2	0.029039
1.34878	TREX2 /// UCHL5IP	three prime repair exonuclease 2 /// UCHL5 interacting protein	0.014841
1.33581	RPS13	ribosomal protein S13	0.00654
1.32547	GRB10	growth factor receptor-bound protein 10	0.023425
1.303224	RPL9	ribosomal protein L9	0.046875

1.300171	RPL38	ribosomal protein L38	0.021022
1.276438	NUDT3	nudix-type motif 3	0.04918
1.274737	PARD3	par-3 partitioning defective 3 homolog (C. elegans)	0.047977
1.27258	RPL27	ribosomal protein L27	0.040303
1.264076	RPL17	ribosomal protein L17	0.018933
1.261194	CDYL	chromodomain protein, Y-like	0.049917
1.255499	RPS9	ribosomal protein S9	0.007091
1.245831	RP5-1077B9.4	invasion inhibitory protein 45	0.017708
1.23863	hCG_21078 /// RPL27A	hCG21078 /// ribosomal protein L27a	0.002151
1.22714	EIF1	eukaryotic translation initiation factor 1	0.033374
1.216908	RPS24	ribosomal protein S24	0.026217
1.170586	LOC100130553 /// RPS18	hypothetical protein LOC100130553 /// ribosomal protein S18	0.02368
1.161676	INPPL1	inositol polyphosphate phosphatase-like 1	0.022554
1.146715	RPL28	ribosomal protein L28	0.037601

A.4 Full list of 258 genes downregulated in keloid compared to normal fibroblasts using the RMA summarization algorithm (P < 0.05)

Fold change	Gene Symbol	Gene Title	Corrected p-value
44.34992	CXCL6	chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)	9.29E-10
41.64695	CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	1.89E-05
37.03845	C2 /// CFB	complement component 2 /// complement factor B	1.75E-08
33.61232	HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1	1.50E-06
29.71772	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	0.008827
21.50683	CXCL2	chemokine (C-X-C motif) ligand 2	5.89E-05
20.34249	TNFAIP6	tumor necrosis factor, alpha-induced protein 6	0.011388
19.02187	IL8	interleukin 8	0.003057
18.80414	SLC39A8	solute carrier family 39 (zinc transporter), member 8	1.22E-05
16.76682	SLC39A8	solute carrier family 39 (zinc transporter), member 8	7.69E-04
14.08095	C3	complement component 3	2.94E-04
13.78165	RSAD2	radical S-adenosyl methionine domain containing 2	0.013835
12.01816	SOD2	superoxide dismutase 2, mitochondrial	0.002706
11.75815	IL8	interleukin 8	0.001761
11.65778	CCL2	chemokine (C-C motif) ligand 2	9.46E-07
10.88791	SFRP1	secreted frizzled-related protein 1	0.026217
9.990352	G0S2	G0/G1switch 2	0.001019
9.679501	IFI44L	interferon-induced protein 44-like	0.04903

8.604651	CHI3L2	chitinase 3-like 2	0.019936
8.378408	IL6	interleukin 6 (interferon, beta 2)	4.59E-04
8.213031	CA12	carbonic anhydrase XII	2.40E-04
8.204023	OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa	0.015085
8.195308	GCH1	GTP cyclohydrolase 1	2.29E-04
8.068323	CA12	carbonic anhydrase XII	0.00631
7.890376	SOD2	superoxide dismutase 2, mitochondrial	0.001861
7.244113	IFIT1	interferon-induced protein with tetratricopeptide repeats 1	0.035976
7.188362	TNFAIP3	tumor necrosis factor, alpha-induced protein 3	8.07E-05
6.98734	IFIT3	interferon-induced protein with tetratricopeptide repeats 3	0.014427
6.850272	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	0.03935
6.773449	OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa	0.044028
6.696858	NAMPT	nicotinamide phosphoribosyltransferase	0.001998
6.448252	IFIH1	interferon induced with helicase C domain 1	0.016536
6.327625	CA12	carbonic anhydrase XII	0.006223
6.259846	HERC6	hect domain and RLD 6	0.009306
6.225736	CTSS	cathepsin S	8.67E-05
6.107744	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa	0.003902
5.973147	IFIT2	interferon-induced protein with tetratricopeptide repeats 2	0.049917
5.971855	CA12	carbonic anhydrase XII	0.004125
5.964412	CA12	carbonic anhydrase XII	4.59E-04
5.887205	HERC5	hect domain and RLD 5	0.012534
5.848494	CXCL3	chemokine (C-X-C motif) ligand 3	5.04E-04
5.825303	SOD2	superoxide dismutase 2, mitochondrial	7.69E-04
5.720924	NAMPT	nicotinamide phosphoribosyltransferase	0.023627
5.663569	BTN3A2	butyrophilin, subfamily 3, member A2	0.042157
5.519223	WTAP	Wilms tumor 1 associated protein	9.55E-05
5.507072	WTAP	Wilms tumor 1 associated protein	0.009169
5.327916	CCL5	chemokine (C-C motif) ligand 5	0.002448
5.289413	ABCA8	ATP-binding cassette, sub-family A (ABC1), member 8	0.006956
5.146521	SLC39A14	solute carrier family 39 (zinc transporter), member 14	0.024324
4.979725	RARRES3	retinoic acid receptor responder (tazarotene induced) 3	0.003712
4.976672	IFI44	interferon-induced protein 44	0.037647
4.80152	AKR1B1	aldo-keto reductase family 1, member B1 (aldose reductase)	3.95E-04
4.67249	LAP3	leucine aminopeptidase 3	0.024205
4.650329	PTGES	prostaglandin E synthase	0.012578
4.624918	TNFAIP2	tumor necrosis factor, alpha-induced protein 2	7.11E-04
4.616492	CCL5	chemokine (C-C motif) ligand 5	0.010015
4.304027	MT1X	metallothionein 1X	0.020822
4.30355	NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	4.59E-04

4.096058	PLSCR1	phospholipid scramblase 1	0.011388
4.09106	CLU	clusterin	0.026445
4.075102	MT1E	metallothionein 1E	0.047073
4.064827	TNFAIP3	tumor necrosis factor, alpha-induced protein 3	0.001023
4.041319	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa	0.021022
3.909078	CLU	clusterin	0.033121
3.828279	STEAP1	six transmembrane epithelial antigen of the prostate 1	0.014024
3.82739	MT2A	metallothionein 2A	0.029119
3.806854	IFI35	interferon-induced protein 35	0.025091
3.680742	XAF1	XIAP associated factor 1	0.023346
3.657572	MT1E /// MT1H /// MT1M /// MT1P2	metallothionein 1E /// metallothionein 1H /// metallothionein 1M /// metallothionein 1 pseudogene 2	0.021022
3.653427	CXCL5	chemokine (C-X-C motif) ligand 5	0.004989
3.645952	PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional peptidase 2)	0.03417
3.586836	IL15RA	interleukin 15 receptor, alpha	0.011711
3.463399	ICAM1	intercellular adhesion molecule 1	1.10E-04
3.378218	UCHL1	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)	0.004886
3.338111	MT1X	metallothionein 1X	0.026959
3.33326	NFIB	nuclear factor I/B	0.026219
3.289887	TMEM100	transmembrane protein 100	0.04263
3.285261	NRCAM	neuronal cell adhesion molecule	7.94E-04
3.285236	TLR3	toll-like receptor 3	0.039578
3.279497	SLC1A1	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	7.69E-04
3.275539	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	3.06E-04
3.262109	IL32	interleukin 32	4.59E-04
3.245122	LRRN3	leucine rich repeat neuronal 3	0.00287
3.224767	MT1F	metallothionein 1F	0.023227
3.219482	PTGFR	prostaglandin F receptor (FP)	0.027574
3.194404	PARP12	poly (ADP-ribose) polymerase family, member 12	0.048335
3.130064	ICAM1	intercellular adhesion molecule 1	3.40E-05
3.075876	MID1	midline 1 (Opitz/BBB syndrome)	0.033374
3.055434	MT1P2	metallothionein 1 pseudogene 2	0.035137
3.049139	VCAM1	vascular cell adhesion molecule 1	0.047073
2.993775	CP	ceruloplasmin (ferroxidase)	8.67E-05
2.925094	AMPD3	adenosine monophosphate deaminase (isoform E)	0.010801
2.918264	IFI30	interferon, gamma-inducible protein 30	0.023627
2.903807	APOL3	apolipoprotein L, 3	0.042118
2.896697	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	2.81E-04
2.892534	STAT1	signal transducer and activator of transcription 1,	0.032961

		91kDa	
2.862865	MMD	monocyte to macrophage differentiation-associated	0.006815
2.810585	ABLIM1	actin binding LIM protein 1	0.026332
2.798357	DDX60	DEAD (Asp-Glu-Ala-Asp) box polypeptide 60	0.043219
2.72018	FILIP1L	filamin A interacting protein 1-like	0.004067
2.711745	GPRC5B	G protein-coupled receptor, family C, group 5, member B	0.016484
2.703468	HIST1H2BD	histone cluster 1, H2bd	0.015237
2.693878	PSME2	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	0.001077
2.689429	BTN3A2	butyrophilin, subfamily 3, member A2	0.025091
2.688505	CFLAR	CASP8 and FADD-like apoptosis regulator	0.015237
2.687159	LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein	0.013645
2.650053	CXCL5	chemokine (C-X-C motif) ligand 5	0.015587
2.647512	DRAM	damage-regulated autophagy modulator	0.003902
2.634461	MT1M	metallothionein 1M	0.008827
2.629179	UBE2L6	ubiquitin-conjugating enzyme E2L 6	0.03412
2.62574	MT1F	metallothionein 1F	9.22E-04
2.595849	CHEK2	CHK2 checkpoint homolog (S. pombe)	5.84E-05
2.594678	CFLAR	CASP8 and FADD-like apoptosis regulator	0.00493
2.593336	NMI	N-myc (and STAT) interactor	0.044028
2.569516	CFLAR	CASP8 and FADD-like apoptosis regulator	5.48E-04
2.54052	SIRPA	signal-regulatory protein alpha	1.38E-05
2.528263	CEBPD	CCAAT/enhancer binding protein (C/EBP), delta	0.019171
2.522403	BTN3A3	butyrophilin, subfamily 3, member A3	0.048335
2.520541	CYB5A	cytochrome b5 type A (microsomal)	0.011736
2.514324	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	0.006815
2.498402	PALM	Paralemmin	0.009971
2.490228	RNF114	ring finger protein 114	0.047073
2.454265	TRIM38	tripartite motif-containing 38	0.021022
2.411988	SLC25A28	solute carrier family 25, member 28	0.001713
2.410709	STAT4	signal transducer and activator of transcription 4	3.39E-04
2.408566	CYB5A	cytochrome b5 type A (microsomal)	0.015085
2.390128	KIAA0391 /// PSMA6	KIAA0391 /// proteasome (prosome, macropain) subunit, alpha type, 6	0.027574
2.366543	SLC15A3	solute carrier family 15, member 3	0.027138
2.361601	GHR	growth hormone receptor	0.001713
2.354626	C1QTNF1	C1q and tumor necrosis factor related protein 1	0.048335
2.351021	NFIB	nuclear factor I/B	0.040444
2.347035	PANX1	pannexin 1	0.03639
2.336691	LRRN3	leucine rich repeat neuronal 3	0.016484
2.333014	RGS3	regulator of G-protein signaling 3	0.026787
2.319342	NT5E	5'-nucleotidase, ecto (CD73)	0.002448
2.311114	MARCH3	membrane-associated ring finger (C3HC4) 3	0.001642
2.27852	CYB5A	cytochrome b5 type A (microsomal)	0.023425
2.251881	TRIM38	tripartite motif-containing 38	0.032743

2.23989	CFLAR	CASP8 and FADD-like apoptosis regulator	0.023911
2.235737	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	0.003755
2.224926	HLA-F	major histocompatibility complex, class I, F	0.009477
2.22473	TAPBP	TAP binding protein (tapasin)	0.002016
2.215581	HLA-C	major histocompatibility complex, class I, C	0.006519
2.210316	AK3L1 /// AK3L2	adenylate kinase 3-like 1 /// adenylate kinase 3-like 2	0.003202
2.20811	HLA-F	major histocompatibility complex, class I, F	0.00287
2.17376	SMC2	structural maintenance of chromosomes 2	0.002998
2.169356	CYLD	cylindromatosis (turban tumor syndrome)	0.046985
2.167533	NOVA1	neuro-oncological ventral antigen 1	0.04244
2.146688	C10orf26	chromosome 10 open reading frame 26	0.025091
2.13148	HLA-B	major histocompatibility complex, class I, B	0.006956
2.124391	HTATIP2	HIV-1 Tat interactive protein 2, 30kDa	0.026762
2.12098	SLC39A8	solute carrier family 39 (zinc transporter), member 8	0.004086
2.108726	TNIP1	TNFAIP3 interacting protein 1	0.027814
2.097694	SAMHD1	SAM domain and HD domain 1	0.047977
2.097592	PON2	paraoxonase 2	0.025091
2.092583	CFLAR	CASP8 and FADD-like apoptosis regulator	0.003645
2.089171	HLA-B	major histocompatibility complex, class I, B	0.002024
2.072831	CFLAR	CASP8 and FADD-like apoptosis regulator	0.00197
2.0599	DHRS3	dehydrogenase/reductase (SDR family) member 3	0.001019
2.055893	DKFZP586H2123	regeneration associated muscle protease	0.037791
2.048616	CFLAR	CASP8 and FADD-like apoptosis regulator	8.55E-04
2.023844	HLA-B /// MICA	major histocompatibility complex, class I, B /// MHC class I polypeptide-related sequence A	0.018933
2.021447	BIRC3	baculoviral IAP repeat-containing 3	0.001713
2.007685	PDPN	podoplanin	0.033129
2.001211	LY6E	lymphocyte antigen 6 complex, locus E	0.027138
1.998355	ZC3H12A	zinc finger CCCH-type containing 12A	5.99E-05
1.981507	HLA-A /// HLA-A29.1 /// HLA-B /// HLA-G /// HLA-H /// HLA-J	major histocompatibility complex, class I, A /// major histocompatibility complex class I HLA-A29.1 /// major histocompatibility complex, class I, B /// major histocompatibility complex, class I, G /// major histocompatibility complex, class I, H (pseudogene) /// major histocompatibility complex, class I, J (pseudogene)	0.006852
1.96592	CSF1	colony stimulating factor 1 (macrophage)	0.015085
1.942084	HTATIP2	HIV-1 Tat interactive protein 2, 30kDa	0.003902
1.935575	ACP2	acid phosphatase 2, lysosomal	0.006566
1.927046	PON2	paraoxonase 2	0.014803
1.924471	IFNGR1	interferon gamma receptor 1	0.018356
1.911548	FAM117A	family with sequence similarity 117, member A	0.001713
1.909791	C14orf159	chromosome 14 open reading frame 159	0.010796
1.896043	HLA-G	major histocompatibility complex, class I, G	0.008626
1.877315	MT1F	metallothionein 1F	1.10E-04
1.876992	HLA-C	major histocompatibility complex, class I, C	7.69E-04

1.87264	MEIS3P1	Meis homeobox 3 pseudogene 1	0.007091
1.87205	TAPBPL	TAP binding protein-like	0.001292
1.85887	HLA-G	major histocompatibility complex, class I, G	0.003202
1.857911	BASP1	brain abundant, membrane attached signal protein 1	0.026445
1.847917	MICALL2	MICAL-like 2	7.91E-04
1.844846	SIRPA	signal-regulatory protein alpha	0.002908
1.843371	GFRA1	GDNF family receptor alpha 1	0.031214
1.837641	HLA-E	major histocompatibility complex, class I, E	0.042803
1.822478	DENND2D	DENN/MADD domain containing 2D	0.009234
1.811002	TRIM38	tripartite motif-containing 38	0.014427
1.805608	NNMT	nicotinamide N-methyltransferase	0.033657
1.801476	FTH1	ferritin, heavy polypeptide 1	0.032417
1.785207	PION	pigeon homolog (Drosophila)	0.026445
1.763702	NNMT	nicotinamide N-methyltransferase	0.015237
1.752336	HLA-C	major histocompatibility complex, class I, C	0.001075
1.748331	NR4A3	nuclear receptor subfamily 4, group A, member 3	0.010796
1.718806	PROCR	protein C receptor, endothelial (EPCR)	0.027362
1.700887	PDCD5	programmed cell death 5	0.039578
1.683461	HSPB8	heat shock 22kDa protein 8	0.04761
1.668382	HLA-C	major histocompatibility complex, class I, C	0.016484
1.667567	PDCD1LG2	programmed cell death 1 ligand 2	0.004125
1.651713	SLCO3A1	solute carrier organic anion transporter family, member 3A1	0.00287
1.641853	SLC1A1	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	0.014024
1.629699	HLA-G	major histocompatibility complex, class I, G	0.006852
1.626722	BTN3A1	butyrophilin, subfamily 3, member A1	0.0442
1.625006	HLA-A	major histocompatibility complex, class I, A	0.002743
1.61725	LARGE	like-glycosyltransferase	0.013456
1.617092	AK3L1 /// AK3L2	adenylate kinase 3-like 1 /// adenylate kinase 3-like 2	0.033481
1.592655	PION	pigeon homolog (Drosophila)	0.003955
1.589381	NFKBIE	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon	0.019276
1.584692	RNF8	ring finger protein 8	0.016484
1.576378	CD59	CD59 molecule, complement regulatory protein	4.56E-04
1.571031	LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	0.029499
1.563836	MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)	5.99E-05
1.556934	NRP2	neuropilin 2	0.02645
1.533629	P2RX4	purinergic receptor P2X, ligand-gated ion channel, 4	0.021666
1.523406	NFKB1	nuclear factor of kappa light polypeptide gene enhancer in B-cells 1	0.042118
1.522656	SVEP1	sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1	0.011711
1.518637	WWC1	WW and C2 domain containing 1	0.045278

1.499967	LSAMP	limbic system-associated membrane protein	0.009696
1.497518	APOL2	apolipoprotein L, 2	0.047977
1.493991	PGK1	phosphoglycerate kinase 1	0.042095
1.489115	GBA /// GBAP	glucosidase, beta; acid (includes glucosylceramidase) /// glucosidase, beta; acid, pseudogene	0.036143
1.488427	KHDC1 /// SPA17	KH homology domain containing 1 /// sperm autoantigenic protein 17	0.004067
1.482473	PSTPIP2	proline-serine-threonine phosphatase interacting protein 2	0.008121
1.466329	POLD3	polymerase (DNA-directed), delta 3, accessory subunit	0.025058
1.449279	DTNA	dystrobrevin, alpha	0.037974
1.446802	RELB	v-rel reticuloendotheliosis viral oncogene homolog B	0.009818
1.445625	ELF3	E74-like factor 3 (ets domain transcription factor, epithelial-specific)	0.02046
1.440517	CYP27A1	cytochrome P450, family 27, subfamily A, polypeptide 1	0.04263
1.439601	HYPK	Huntingtin interacting protein K	0.019813
1.429019	ACP6	acid phosphatase 6, lysophosphatidic	0.039972
1.406325	TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)	0.006852
1.397198	CTSS	cathepsin S	0.046985
1.388266	B2M	beta-2-microglobulin	0.024654
1.38517	CSTB	cystatin B (stefin B)	0.017681
1.382151	SNCA	synuclein, alpha (non A4 component of amyloid precursor)	0.006815
1.368093	PDLIM4	PDZ and LIM domain 4	0.041423
1.367311	NFE2L1	nuclear factor (erythroid-derived 2)-like 1	0.032417
1.349224	FTH1	ferritin, heavy polypeptide 1	0.013128
1.321152	ARHGEF10L	Rho guanine nucleotide exchange factor (GEF) 10-like	0.030511
1.320407	LGALS8	lectin, galactoside-binding, soluble, 8	0.044328
1.313343	FTHP1	ferritin, heavy polypeptide pseudogene 1	0.03731
1.308686	EDNRB	endothelin receptor type B	0.046273
1.307239	HLA-F	major histocompatibility complex, class I, F	0.039578
1.272	SLC19A3	solute carrier family 19, member 3	0.031469
1.266423	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	0.018933
1.25115	RBKS	ribokinase	0.018933
1.249708	CAND2	cullin-associated and neddylation-dissociated 2 (putative)	0.020736
1.246255	SNX11	sorting nexin 11	0.020092
1.238541	C6orf64	chromosome 6 open reading frame 64	0.044823
1.232883	ZMIZ2	zinc finger, MIZ-type containing 2	0.033374
1.225532	CSF1	colony stimulating factor 1 (macrophage)	0.013754
1.196416	NFIB	nuclear factor I/B	0.047977
1.174331	TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)	0.029499
1.155878	WDR48	WD repeat domain 48	0.006116

1.146091	SP100	SP100 nuclear antigen	0.015085
1.131624	TLR1	toll-like receptor 1	0.033657
1.106074	CXCL5	chemokine (C-X-C motif) ligand 5	0.044823
1.104509	TAF1B	TATA box binding protein (TBP)-associated factor, RNA polymerase I, B, 63kDa	0.037317
1.101118	BRCA2	breast cancer 2, early onset	0.027138

A.5 List of genes differentially expressed using both the RMA and MAS 5.0 summarization algorithm ($P < 0.05$)

Regulation (Keloid vs Normal)	Gene Symbol	Gene Title
down	CXCL6	chemokine (C-X-C motif) ligand 6 (granulocyte chemotactic protein 2)
down	CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)
down	IL8	interleukin 8
down	HSD11B1	hydroxysteroid (11-beta) dehydrogenase 1
down	IL8	interleukin 8
down	CCL5	chemokine (C-C motif) ligand 5
down	CXCL2	chemokine (C-X-C motif) ligand 2
down	RSAD2	radical S-adenosyl methionine domain containing 2
down	C2 /// CFB	complement component 2 /// complement factor B
down	CXCL5	chemokine (C-X-C motif) ligand 5
down	TNFAIP6	tumor necrosis factor, alpha-induced protein 6
down	CXCL3	chemokine (C-X-C motif) ligand 3
down	IL32	interleukin 32
down	CP	ceruloplasmin (ferroxidase)
down	CHI3L2	chitinase 3-like 2
down	TNFAIP6	tumor necrosis factor, alpha-induced protein 6
down	C3	complement component 3
down	SLC39A8	solute carrier family 39 (zinc transporter), member 8
down	CXCL5	chemokine (C-X-C motif) ligand 5
down	G0S2	G0/G1switch 2
down	OAS1	2',5'-oligoadenylate synthetase 1, 40/46kDa
down	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10
down	SLC39A8	solute carrier family 39 (zinc transporter), member 8
down	GCH1	GTP cyclohydrolase 1
down	CCL5	chemokine (C-C motif) ligand 5
down	SLC19A3	solute carrier family 19, member 3
down	HERC5	hect domain and RLD 5
down	IL6	interleukin 6 (interferon, beta 2)
down	NRCAM	neuronal cell adhesion molecule
down	CCL2	chemokine (C-C motif) ligand 2
down	ABCA8	ATP-binding cassette, sub-family A (ABC1), member 8

down	SFRP1	secreted frizzled-related protein 1
down	TMEM100	transmembrane protein 100
down	SOD2	superoxide dismutase 2, mitochondrial
down	CTSS	cathepsin S
down	IFI30	interferon, gamma-inducible protein 30
down	HERC6	hect domain and RLD 6
down	TLR3	toll-like receptor 3
down	RARRES3	retinoic acid receptor responder (tazarotene induced) 3
down	SOD2	superoxide dismutase 2, mitochondrial
down	TNFAIP3	tumor necrosis factor, alpha-induced protein 3
down	CA12	carbonic anhydrase XII
down	CTSS	cathepsin S
down	IFIT2	interferon-induced protein with tetratricopeptide repeats 2
down	LRRN3	leucine rich repeat neuronal 3
down	IFIT3	interferon-induced protein with tetratricopeptide repeats 3
down	BIRC3	baculoviral IAP repeat-containing 3
down	CA12	carbonic anhydrase XII
down	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa
down	TNFAIP2	tumor necrosis factor, alpha-induced protein 2
down	CA12	carbonic anhydrase XII
down	PTGES	prostaglandin E synthase
down	GBP1	guanylate binding protein 1, interferon-inducible, 67kDa
down	NAMPT	nicotinamide phosphoribosyltransferase
down	SLC39A8	solute carrier family 39 (zinc transporter), member 8
down	CA12	carbonic anhydrase XII
down	STAT4	signal transducer and activator of transcription 4
down	CA12	carbonic anhydrase XII
down	LRRN3	leucine rich repeat neuronal 3
down	MT1X	metallothionein 1X
down	NAMPT	nicotinamide phosphoribosyltransferase
down	MT1M	metallothionein 1M
down	WTAP	Wilms tumor 1 associated protein
down	MT1P2	metallothionein 1 pseudogene 2
down	ELF3	E74-like factor 3 (ets domain transcription factor, epithelial-specific)
down	TNFAIP3	tumor necrosis factor, alpha-induced protein 3
down	SOD2	superoxide dismutase 2, mitochondrial
down	MT1E	metallothionein 1E
down	ZC3H12A	zinc finger CCCH-type containing 12A
down	ICAM1	intercellular adhesion molecule 1
down	SLC39A14	solute carrier family 39 (zinc transporter), member 14
down	MT1F	metallothionein 1F
down	AKR1B1	aldo-keto reductase family 1, member B1 (aldose reductase)
down	SNCA	synuclein, alpha (non A4 component of amyloid precursor)
down	WWC1	WW and C2 domain containing 1
down	ICAM1	intercellular adhesion molecule 1
down	WTAP	Wilms tumor 1 associated protein
down	SLC15A3	solute carrier family 15, member 3

down	IFI35	interferon-induced protein 35
down	MT1E /// MT1H /// MT1M /// MT1P2	metallothionein 1E /// metallothionein 1H /// metallothionein 1M /// metallothionein 1 pseudogene 2
down	AMPD3	adenosine monophosphate deaminase (isoform E)
down	MT2A	metallothionein 2A
down	LAP3	leucine aminopeptidase 3
down	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2
down	MARCH3	membrane-associated ring finger (C3HC4) 3
down	CLU	Clusterin
down	STEAP1	six transmembrane epithelial antigen of the prostate 1
down	GFRA1	GDNF family receptor alpha 1
down	NFKBIA	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
down	MMD	monocyte to macrophage differentiation-associated
down	MT1X	metallothionein 1X
down	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2
down	PTGFR	prostaglandin F receptor (FP)
down	NR4A3	nuclear receptor subfamily 4, group A, member 3
down	PLSCR1	phospholipid scramblase 1
down	UCHL1	ubiquitin carboxyl-terminal esterase L1 (ubiquitin thiolesterase)
down	CHEK2	CHK2 checkpoint homolog (S. pombe)
down	MT1F	metallothionein 1F
down	EDNRB	endothelin receptor type B
down	IL15RA	interleukin 15 receptor, alpha
down	STAT1	signal transducer and activator of transcription 1, 91kDa
down	SMC2	structural maintenance of chromosomes 2
down	SAMHD1	SAM domain and HD domain 1
down	ABLIM1	actin binding LIM protein 1
down	MT1F	metallothionein-1F
down	HLA-F	major histocompatibility complex, class I, F
down	CFLAR	CASP8 and FADD-like apoptosis regulator
down	SIRPA	signal-regulatory protein alpha
down	HLA-C	major histocompatibility complex, class I, C
down	SLC25A28	solute carrier family 25, member 28
down	AK3L1 /// AK3L2	adenylate kinase 3-like 1 /// adenylate kinase 3-like 2
down	TAPBPL	TAP binding protein-like
down	NOVA1	neuro-oncological ventral antigen 1
down	DTNA	dystrobrevin, alpha
down	PDPN	Podoplanin
down	LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein
down	PSME2	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)
down	SLC1A1	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1

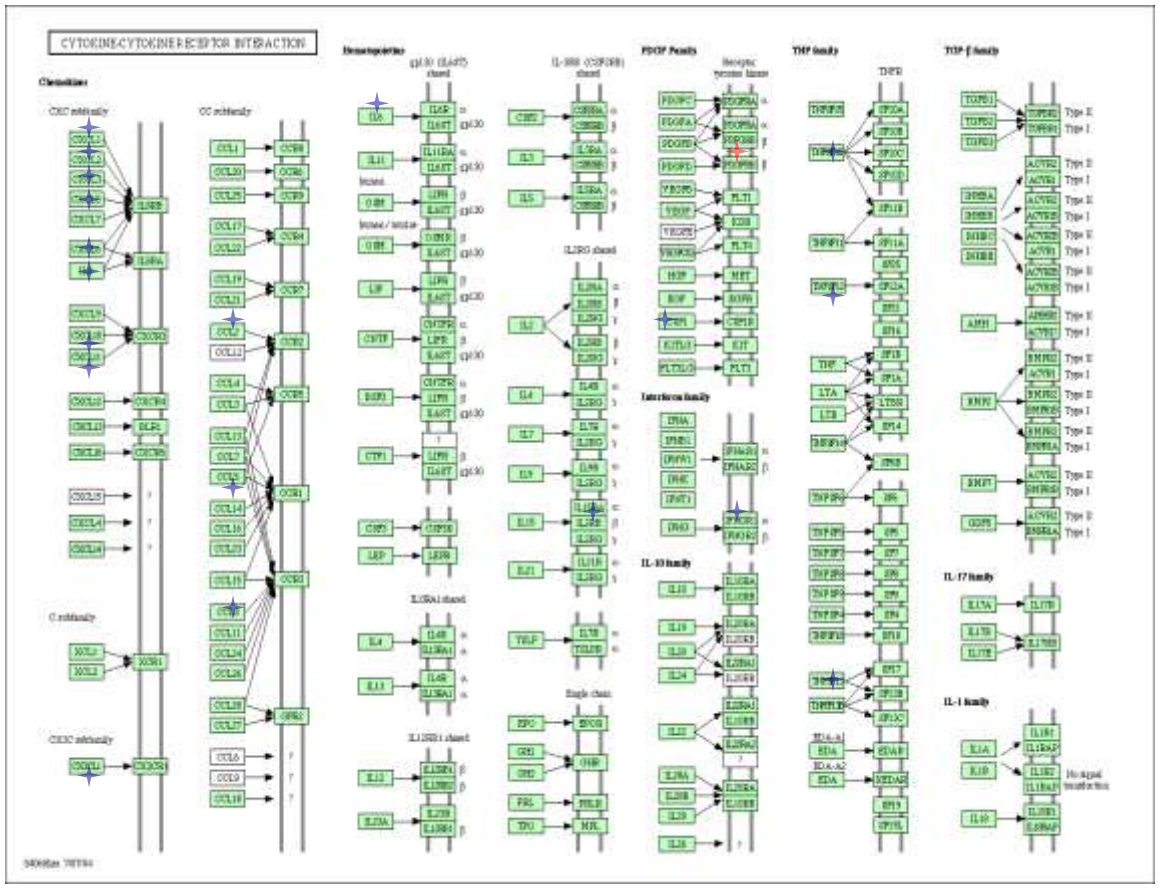
down	CFLAR	CASP8 and FADD-like apoptosis regulator
down	HIST1H2BD	histone cluster 1, H2bd
down	PROCR	protein C receptor, endothelial (EPCR)
down	CFLAR	CASP8 and FADD-like apoptosis regulator
down	CYB5A	cytochrome b5 type A (microsomal)
down	FILIP1L	filamin A interacting protein 1-like
down	PSTPIP2	proline-serine-threonine phosphatase interacting protein 2
down	HLA-F	major histocompatibility complex, class I, F
down	CFLAR	CASP8 and FADD-like apoptosis regulator
down	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)
down	RNF114	ring finger protein 114
down	SLC11A2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2
down	CFLAR	CASP8 and FADD-like apoptosis regulator
down	DRAM	damage-regulated autophagy modulator
down	CFLAR	CASP8 and FADD-like apoptosis regulator
down	HLA-G	major histocompatibility complex, class I, G
down	LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog
down	HLA-B	major histocompatibility complex, class I, B
down	HTATIP2	HIV-1 Tat interactive protein 2, 30kDa
down	MICALL2	MICAL-like 2
down	SIRPA	signal-regulatory protein alpha
down	HLA-G	major histocompatibility complex, class I, G
down	CYB5A	cytochrome b5 type A (microsomal)
down	TAPBP	TAP binding protein (tapasin)
down	PON2	paraoxonase 2
down	NT5E	5'-nucleotidase, ecto (CD73)
down	TRIM38	tripartite motif-containing 38
down	HLA-B	major histocompatibility complex, class I, B
down	CFLAR	CASP8 and FADD-like apoptosis regulator
down	KIAA0391 /// PSMA6	KIAA0391 /// proteasome (prosome, macropain) subunit, alpha type, 6
down	TRIM38	tripartite motif-containing 38
down	NFKBIE	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon
down	C10orf26	chromosome 10 open reading frame 26
down	LY6E	lymphocyte antigen 6 complex, locus E
down	PANX1	pannexin 1
down	CSF1	colony stimulating factor 1 (macrophage)
down	TNIP1	TNFAIP3 interacting protein 1
down	RELB	v-rel reticuloendotheliosis viral oncogene homolog B
down	ACP2	acid phosphatase 2, lysosomal
down	PON2	paraoxonase 2
down	TRIM38	tripartite motif-containing 38
down	HLA-C	major histocompatibility complex, class I, C
down	HLA-A /// HLA-A29.1 /// HLA-B /// HLA-G ///	major histocompatibility complex, class I, A /// major histocompatibility complex class I HLA-A29.1 /// major histocompatibility complex, class I, B /// major histocompatibility complex, class I, G /// major histocompatibility complex, class I, H

	HLA-H /// HLA-J	(pseudogene) /// major histocompatibility complex, class I, J (pseudogene)
down	NNMT	nicotinamide N-methyltransferase
down	HLA-C	major histocompatibility complex, class I, C
down	BASP1	brain abundant, membrane attached signal protein 1
down	IFNGR1	interferon gamma receptor 1
down	FTH1	ferritin, heavy polypeptide 1
down	HLA-C	major histocompatibility complex, class I, C
down	HLA-B /// MICA	major histocompatibility complex, class I, B /// MHC class I polypeptide-related sequence A
down	CD59	CD59 molecule, complement regulatory protein
down	NNMT	nicotinamide N-methyltransferase
down	HLA-A	major histocompatibility complex, class I, A
down	MMP2	matrix metallopeptidase 2 (gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase)
down	CSTB	cystatin B (stefin B)
down	FTH1	ferritin, heavy polypeptide 1
down	B2M	beta-2-microglobulin
up	POSTN	periostin, osteoblast specific factor
up	COL15A1	collagen, type XV, alpha 1
up	HOXA11	homeobox A11
up	KCNJ6	potassium inwardly-rectifying channel, subfamily J, member 6
up	JUP /// KRT19	junction plakoglobin /// keratin 19
up	IGFBP3	insulin-like growth factor binding protein 3
up	ATXN1	ataxin 1
up	CADM1	cell adhesion molecule 1
up	SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A
up	GPSM2	G-protein signaling modulator 2 (AGS3-like, C. elegans)
up	FAM155A	family with sequence similarity 155, member A
up	MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2
up	CADM1	cell adhesion molecule 1
up	SEMA5A	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A
up	SHMT2	serine hydroxymethyltransferase 2 (mitochondrial)
up	MICAL2	microtubule associated monoxygenase, calponin and LIM domain containing 2
up	FARP1	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)
up	MGC87895 /// RPS14	similar to ribosomal protein S14 /// ribosomal protein S14
up	TBC1D2	TBC1 domain family, member 2
up	ECM1	extracellular matrix protein 1
up	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6
up	PDGFRB	platelet-derived growth factor receptor, beta polypeptide

up	CTSB	cathepsin B
up	FHOD1	formin homology 2 domain containing 1
up	SLC25A6	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6
up	NXN	Nucleoredoxin
up	PTK7	PTK7 protein tyrosine kinase 7
up	LAMA2	laminin, alpha 2
up	LOC644191 /// LOC728937 /// RPS26	similar to hCG15685 /// similar to 40S ribosomal protein S26 /// ribosomal protein S26
up	PARVB	parvin, beta
up	COL1A1	collagen, type I, alpha 1
up	SHMT2	serine hydroxymethyltransferase 2 (mitochondrial)
up	NONO	non-POU domain containing, octamer-binding
up	ATF4	activating transcription factor 4 (tax-responsive enhancer element B67)
up	RPS9	ribosomal protein S9
up	RPL13	ribosomal protein L13
up	RPS8	ribosomal protein S8
up	RPL13	ribosomal protein L13
up	RPS6	ribosomal protein S6
up	RPL8	ribosomal protein L8
up	PPP2R3A	protein phosphatase 2 (formerly 2A), regulatory subunit B'', alpha
up	RPL13	ribosomal protein L13

A.6 Cytokine-cytokine receptor interaction from the KEGG database

(Benjamini corrected P-value = 0.094)



A.7 Toll-like receptor signaling pathway from the KEGG database

(Benjamini corrected P-value = 0.246)