Elucidating the Molecular Basis of a Novel Autosomal Dominant Fibrotic Syndrome

by

Lauren Watson
HRTLAU004

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfilment of the requirements for the degree

MSc (Med) in Human Genetics

Faculty of Health Sciences
UNIVERSITY OF CAPE TOWN

30 January 2009

Supervisor: Prof BM Mayosi (Dept. of Medicine)
Co-Supervisors: Prof R Ramesar (Div. of Human Genetics)
Prof T Pillay (Div. of Chemical Pathology)

Acknowledgements

Nothing significant is ever accomplished without the support and encouragement of others, and I would like to briefly acknowledge those who have played a role in making this degree a worthwhile experience.

I am indebted to my supervisors for their patience, guidance and invaluable advice. To Bongani, thank you for your insight and enthusiasm. I consider myself privileged to have been mentored by someone with such vision for the development of others. And to Raj and Prof Pillay, thank you for so willingly giving of your resources and your time.

To the many colleagues and friends who have supported me throughout this degree – know that your wisdom and generosity are greatly appreciated. To Wan Ismail, thank you for your patience and generosity, and for being willing to share your western blotting expertise. To Ingrid Baumgarten, thank you for keeping my cells alive, and for offering technical support whenever necessary. To George Rebello, thank you for the opportunity to teach, and for getting us all started on this long road. To Janine Scholefield and Jacqueline Meyer, thank you for your generous donations of biological material. To Robea Ballo, thank you for introducing me to real-time PCR, and being willing to answer all my questions. To Alvera Vorster, thank you for innumerable troubleshooting sessions, and for always being ready to share your wealth of technical expertise. And to everyone in the Human Genetics Lab – thank you for being such a supportive research "family".

This project would not have been possible without financial assistance, and in this regard, I would like to acknowledge the National Research Foundation for the Prestigious Masters Scholarship, as well as the Deutscher Akademischer Austauschdienst (DAAD), the Ernst and Ethel Eriksen Trust, and the University of Cape Town.

Finally, to my family and friends, and most especially my husband, Garth, without whose encouragement, love and unconditional support I would not have managed the past two years – thank you, as always, for carrying me through with your prayers.

Plagiarism Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the Journal of Human Genetics convention for citation and referencing. Each contribution to, and quotation in, this thesis from the work(s) of other people has been attributed, and has been cited and referenced.
3. This thesis is my own work using my own words.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Date_____

Student Name_____

Student Number_____

Signature _____

Contents

List of Abbreviations

A	Adenine
Arg	Arginine
BLAST	Basic Local Alignment Search Tool
BMP	Bone Morphogenic Protein
bp	base pairs
BSA	Bovine Serum Albumin
C	cytosine
C. elegans	*Caenorhabditis elegans*
cDNA	complementary Deoxyribose Nucleic Acid
Co-Smad	common partner Smad
C_T	crossing point/threshold cycle
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulphoxide
DNA	Deoxyribose Nucleic Acid
dNTPs	deoxyribonicleotides
E.coli	*Escherichia coli*
ECM	Extracellular Matrix
EDTA	Ethylenediamine Tetraacetic Acid
ER	Endoplasmic Reticulum
G	Guanine
hr	hour
HRP	Horseradish Peroxidase
I-Smad	inhibitory Smad
kb	Kilobases
kDa	Kilodaltons
LAP	Latency Associated Peptide
LD	Linkage Disequilibrium
Leu	Leucine
LOD	Logarithm of Odds
LTBP	Latent TGF-β Binding Protein
M	Molar (mol/dm^3)
mA	miliampere

MAP	Mitogen-Associated Protein
$MgCl_2$	Magnesium Chloride
MH1/2	Mad Homology Domain 1/2
min	minutes
ml	mililitres
mM	milimolar
mRNA	Messenger Ribose Nucleic Acid
NCBI	National Centre for Biotechnology Information
ng	nanograms
nm	nanometres
NTC	No-Template Control
PAI-1	Plasminogen Activator Inhibitor Type-1
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
pM	picomolar
Pro	Proline
PVDF	Polyvinylidene Fluoride
qPCR	quantitative Polymerase Chain Reaction
RNA	Ribose Nucleic Acid
rpm	revolutions per minute
rRNA	ribosomal RNA
R-Smad	receptor-regulated Smad
RTS	Rothmund-Thomson Syndrome
s	seconds
SARA	Smad Anchor for Receptor Activation
SBE	Smad Binding Element
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	SDS-Polyacrylamide Gel Electrophoresis
SNP	Single Nucleotide Polymorphism
Sp1	Stimulating Protein 1
T	Thymine
T_a	Annealing Temperature
TBE	Tris-Borate EDTA Electrophoresis Buffer

TBST	Tris Buffered Saline with 0.1% Tween
TGF-β1	Transforming Growth Factor β1
T_m	Melting Temperature
TβR	TGFβ Receptor
U	Units
UCT	University of Cape Town
UV	Ultraviolet
v	version
V	volts
v/v	volume per volume
w/v	weight per volume
µg	micrograms
µl	microlitres
µM	micromolar

Index of Figures

Index of Tables

1. Abstract

A novel fibrotic syndrome was recently reported in a South African family, characterised by poikiloderma, tendon contracture and progressive pulmonary fibrosis. The pathological hallmark of this autosomal dominant condition is abnormal fibrosis of the skin, tendons and viscera, with variable penetrance. A candidate gene approach was adopted to investigate the molecular basis of this disease. The *TGFB1* gene, encoding the multifunctional cytokine Transforming Growth Factor-β1, as well components of the Smad signalling cascade (Smads -2, -3 and -7) were selected as candidates, based on their established role in fibrotic pathology.

The length, sequence and expression of the *TGFB1* transcript were examined in an affected and an unaffected family member, in order to identify possible disease-causing aberrations. Additionally, expression and phosphorylation of the receptor-regulated Smads -2 and -3 were determined in skin fibroblasts from these family members, and two age- and sex-matched controls. Finally, mRNA expression levels of *SMAD7*, encoding the inhibitory Smad7, were compared between the affected and unaffected individuals.

Aside from two single nucleotide polymorphisms identified in the sequence of the *TGFB1* transcript (neither of which appeared to be disease-causing), no differences in length or sequence were identified between the two family members. Additionally, no significant difference in gene expression was observed (p=0.43). In response to TGF-β1 treatment, no significant difference in Smad2/3 expression or phosphorylation was observed between individuals. Similarly, the comparison between basal and TGF-β1-inducible *SMAD7* expression revealed no significant difference between individuals (p=0.22 and p=0.88, respectively).

These results indicate that the TGF-β1/Smad pathway is neither constitutively active nor hyperresponsive in affected individuals, and hence that the disease-causing defect responsible for the novel syndrome is unlikely to be present within this signalling cascade.

2. Introduction

2.1 Disease

2.1.1 Symptoms and Characterisation

A novel familial fibrosing syndrome, affecting two generations of a South African family, was reported in the *British Journal of Dermatology* (Khumalo et al. 2006). Major clinical manifestations of the disease include poikiloderma, tendon contracture, progressive pulmonary fibrosis, and degenerative changes in other organs (Fig. 1) (Khumalo et al. 2006). The proband, a twenty-six year old woman (II.5), was found to have sparse hair and hypohydrosis (indicative of ectodermal dysplasia), sclerosis of the digits, and restricted lung function. She also displayed skin abnormalities including mottled pigmentation, telangiectasia, and epidermal atrophy – collectively described as poikiloderma – and reported a history of tendon contractures (Khumalo et al. 2006).

Her father (I.2) and older brother (II.3), also afflicted with the disease, had shown similar skeletal and skin abnormalities, and had both died as a result of pulmonary fibrosis. The latter presented with tendon contractures at a young age, requiring tendon lengthening surgery by the age of five years. Upon postmortem examination, he was found to have tapered limbs, clubbing of fingers and toes, excessive fatty deposition in the abdominal region (including fatty infiltration of the pancreas and peripheral skeletal muscle) and elastic tissue degeneration, accompanied by calcification of the arteries (Khumalo et al. 2006).

Symptoms involving the skin appear to be the most common, affecting both a half sister (II.2), from the father's first marriage, and another brother (II.4), aged 31, who also exhibited limb abnormalities, hypertension, and left ventricular hypertrophy. In all cases, autoimmune serology was negative (Khumalo et al. 2006).

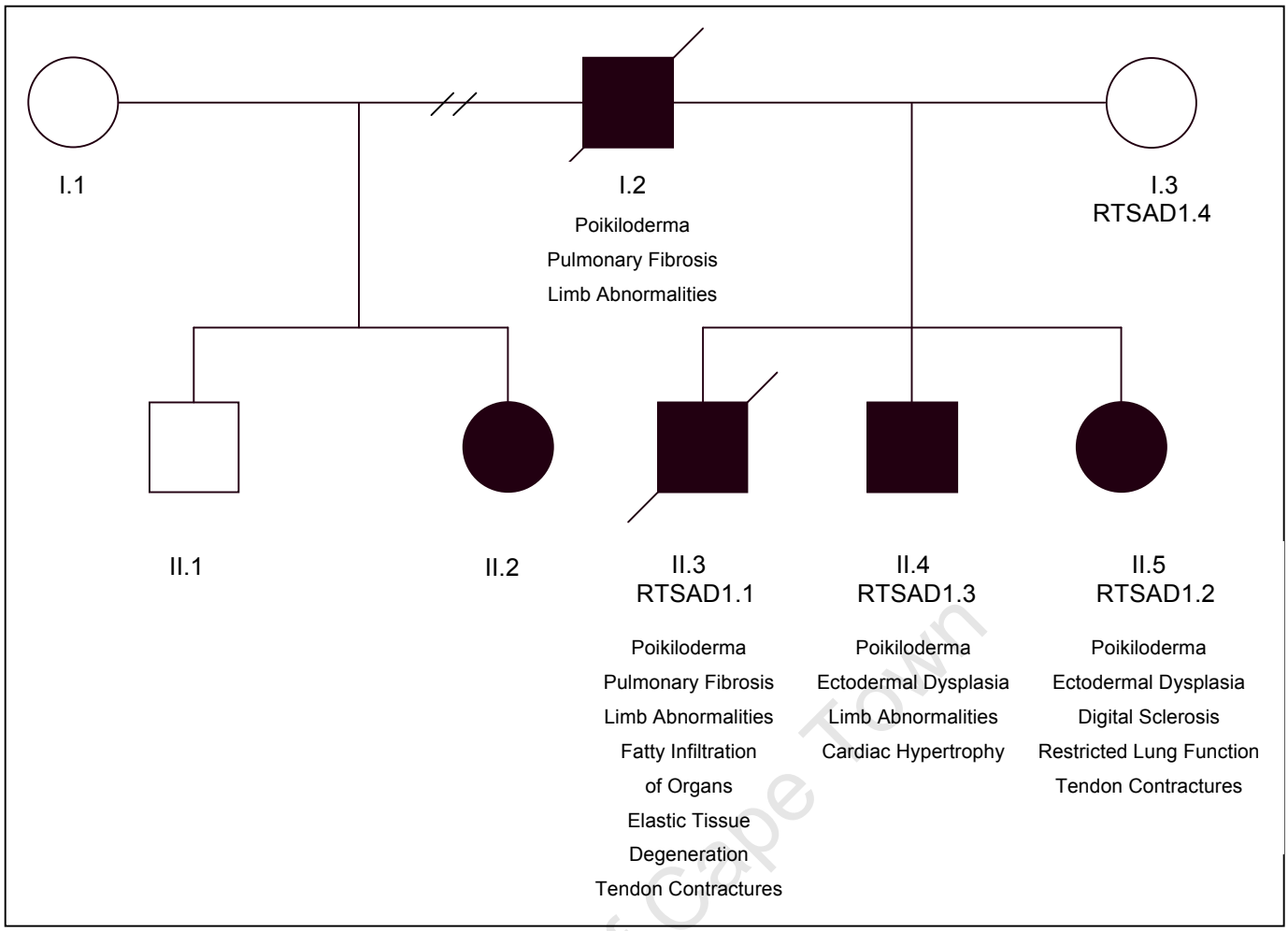


Figure 1. Pedigree of a South African family with a novel familial fibrotic syndrome. The presence of affected individuals in both generations and of both sexes is indicative of autosomal dominant inheritance (Khumalo et al. 2006). Individuals for whom biological material is available are indicated by additional coding (RTSAD1.1 – 1.4) and phenotypic abnormalities are listed where known. Males are indicated by a square and females by a circle. Black symbols represent affected individuals and clear symbols, unaffected individuals. A line drawn through a symbol indicates that the individual is deceased.

Despite the observed clinical heterogeneity, many of the symptoms (including skin pathology, hypohydrosis and tendon contractures) are believed to be secondary to excessive fibrosis in the relevant tissues, an observation that is validated by histological findings (Khumalo et al. 2006).

The mode of inheritance is likely to be autosomal dominant, since offspring of both sexes, and from both of the father's marriages, show symptoms of the disease (Fig. 1). Transmission of the disease from the affected father to his sons indicates autosomal (as opposed to X-linked) inheritance, while the

presence of affected individuals in two generations of this small family is suggestive of a highly penetrant, dominant condition. Neither of the marriages was consanguineous, and the two wives are unrelated, further supporting this assertion (Khumalo et al. 2006).

2.1.2 Comparison and Contrast with Related Diseases

Khumalo et al. (2006) suggested that the novel syndrome be classified as a member of the "hereditary sclerosing poikiloderma" group of diseases, which includes the premature aging disorders Rothmund-Thomson Syndrome (RTS) and Werner Syndrome, as well as the sclerosing poikiloderma of Weary. While the syndrome under study shares certain cutaneous manifestations with this family of diseases, and with the autoimmune disease scleroderma (systemic sclerosis), notable differences have led to its recognition as a novel disorder (Khumalo et al. 2006).

2.1.2.1 Rothmund-Thomson Syndrome

RTS, a rare syndrome characterised by a blistering rash, dermal atrophy, sparse hair and skeletal abnormalities, begins in early life as an acute-phase erythema of the cheeks, spreading eventually to the forehead, buttocks and extremities (Pesce and Rothe 1996; Wang et al. 2001). This is followed by the chronic, or poikilodermatous phase, identified by the presence of "reticulated pigmentation, telangiectasiae, and areas of punctate dermal atrophy" (Wang et al. 2001), which persists into adulthood. Additional symptoms include alopecia (affecting both scalp and facial hair), small stature, nail dystrophy, and a range of related skeletal and ocular abnormalities, including diminished or absent appendages (Kitao et al. 1999; Lindor et al. 2000; Pesce and Rothe 1996).

RTS is caused by the *RECQL4* gene, situated on chromosome 8q24.3. *RECQL4* is a member of the RecQ-like helicase family, which is hypothesized to play a "guardian" role within eukaryotic cells, maintaining genomic stability (Kitao et al. 1999; Lindor et al. 2000). Defects in *RECQL4*, leading to RTS, are inherited in an autosomal recessive fashion (Pesce and Rothe 1996).

2.1.2.2 Werner Syndrome

Werner Syndrome, which is, like RTS, an autosomal recessive premature aging disorder, differs from other disorders in this category by virtue of a later age of onset (usually in either the third or fourth decade). Symptoms typically include short stature, thin, spindly limbs coupled with a stocky trunk, scleroderma-like changes to skin, epidermal atrophy, telangiectasia, mottled pigmentation, and alopecia affecting both facial and body hair. Musculoskeletal symptoms, such as osteoporosis, sclerodactyly and calcification of ligaments and tendons, as well as cardiovascular manifestations (namely valvular calcification and cardiac hypertrophy and fibrosis) have also been observed (Pesce and Rothe 1996). Werner Syndrome is caused by mutations in the *RECQ3* gene, another member of the human RecQ helicase family (Lindor et al. 2000).

2.1.2.3 Weary Sclerosing Poikiloderma

The hereditary sclerosing poikiloderma first described by Weary and colleagues in 1969 is distinguished by the presence of five clinical features, namely: generalized poikiloderma, sclerosis of the hands and feet, hyperkeratotic and sclerotic bands in the armpits and other regions, finger clubbing, and tissue calcification (Weary et al. 1969). These findings were corroborated by Grau Salvat et al. (1999), who also observed calcification of the aorta and mitral valves in their case study.

2.1.2.4 Scleroderma/Systemic Sclerosis

Aspects of the novel syndrome have also been likened to scleroderma (systemic sclerosis), a chronic systemic autoimmune disease of unknown cause, whose symptoms include vascular dysfunction, the production of autoantibodies, and excessive deposition of extracellular matrix (ECM) components, ultimately leading to progressive fibrosis of the skin, lungs and other organs (Dong et al. 2002; Saito et al. 2002). Telangiectasiae, loss of sweat glands, and periarticular calcinosis have also been observed (Jimenez and Derk 2004; Sommer and Goodfield 2002).

2.1.2.5 Contrast with Novel Disease

While certain of these symptoms, particularly those involving the skin, and to a lesser degree, the limbs, can be identified in the South African family described by Khumalo et al. (2006), notable differences exist. For instance, two of the prominent features of the recently characterised syndrome, pulmonary fibrosis and tendon contractures, are not recognized as symptoms of the hereditary poikilodermas. Additionally, Rothmund-Thomson and Werner Syndromes are inherited in an autosomal recessive fashion, and thus do not fit the profile of the novel disease (Khumalo et al. 2006; Wang et al. 2001).

Affected members of the newly described family show no evidence of cutaneous fibrous bands or valvular heart disease (characteristic of Weary sclerosing poikiloderma), or of cataracts or hypogonadism (symptoms of RTS) (Khumalo et al. 2006). They are also autoimmune negative, ruling out systemic sclerosis as a possible cause, despite previously reported associations between pulmonary fibrosis (a prominent symptom of the novel disorder) and systemic sclerosis (Steen 2005; White 2003).

2.2 Methods of Finding Determinant Gene

The identification of gene(s) and/or protein(s) responsible for the observed phenotype represents an important first step in the molecular characterisation of a novel syndrome. The genes responsible for Rothmund-Thomson and Werner Syndromes (*RECQL4* and *RECQ3*, respectively), were deemed to be unlikely candidates, because of the discrepancies in symptoms and mode of inheritance observed between these disorders and the novel disease described by Khumalo et al. (2006). The decision was therefore made to look for other genes or proteins, which may play a causal role in this disorder. Several methods are currently employed to prioritise genes for investigation, including linkage-based methods, such as linkage analysis, positional cloning and linkage disequilibrium mapping, the candidate gene approach, and genome-wide transcriptional profiling (McCarthy et al. 2003). However, not all of these are readily applicable to small cohort sizes, such as the affected family in this study.

2.2.1 Linkage-Based Methods

Linkage-based techniques for identifying disease-causing or -susceptibility genes are based on the assumption that chromosomal regions (defined by polymorphic markers) that contain the determinant gene will be likely to be conserved in individuals affected with the disease of interest, and that the frequency of such disease-associated allele(s) will be higher in the affected population than amongst unaffected individuals (Lander and Kruglyak 1995; McCarthy et al. 2003). In practice, this amounts to a scan of the entire genome (or of a relevant chromosomal region previously identified) using a dense collection of Single Nucleotide Polymorphisms (SNPs) or microsatellite markers (e.g. dinucleotide repeats), calculation of an approximate linkage statistic at each position along the genome, and identification of regions at which the value for linkage greatly exceeds that which would be expected to occur by chance (Lander and Kruglyak 1995). There are two common types of linkage-based methods, namely linkage analysis and linkage disequilibrium mapping.

In linkage analysis, disease gene localisation is based on direct observation of recombination within families (McCarthy et al. 2003). These studies focus on the investigation of chromosomal regions that are co-inherited with the disease in families (Kruglyak 1999). Genes which are found to be present on conserved haplotypes (unaltered by meiotic recombination) in affected individuals after several rounds of linkage mapping may be screened for disease-causing mutations in a process known as positional cloning, which has proven highly successful in the identification of genes responsible for diseases with a simple Mendelian inheritance pattern (McCarthy et al. 2003).

However, because of the often weak associations between multiple susceptibility genes and the observed phenotype, linkage analysis cannot be as readily applied to the problem of complex disorders (McCarthy et al. 2003; Terwilliger and Weiss 1998). In such instances, Linkage Disequilibrium (LD) mapping has proved useful. LD refers to the non-random assortment of alleles, and appears as a difference in allele frequency distribution of one locus, dependent on a second (tightly-linked) locus present on the same

chromosome (Terwilliger and Weiss 1998). Thus, if a locus is hypothesised to have a disease-predisposing allele, and if this allele is in LD with alleles of a marker closely positioned on the same chromosome, the disease gene can be mapped.

This process is of particular relevance to complex disorders, since the nature of a LD screen does not imply the presence of a single conserved haplotype surrounding the susceptibility gene, nor does it stipulate any particular mode of inheritance (Terwilliger and Weiss 1998). Additionally, in allowing for multiple alleles of a particular marker locus, LD mapping is amenable to population-wide studies, rather than being restricted to single families at a time (Kruglyak 1999). LD is also useful in the fine-mapping of a disease-linked locus identified by linkage analysis, and as such may prove vital in refining the positional cloning approach and extending its applications to multifactorial diseases (Kruglyak 1999; McCarthy et al. 2003).

As a result of the central role of statistical inference in these two techniques, however, sample size becomes a limiting factor in determining the significance of a result. In linkage analysis, a logarithm of odds (LOD) score of 3 or greater is required to support a hypothesis of linkage in Mendelian traits, with complex traits demanding more stringent thresholds of significance. The number of meioses from which a statistically sound inference can be drawn, as well as the values of linkage which should be considered significant, also remain contentious issues in LD studies (Lander and Kruglyak 1995). Smaller sample sizes are generally acknowledged to be less informative, since their results are often indistinguishable from what may be expected purely by chance. For this reason, linkage-based techniques cannot be used for small families (such as the pedigree for the novel syndrome under investigation), unless combined with other methods of determinant gene identification.

It must however be noted that a “negative” linkage result obtained in a small family can be deemed conclusive. The abolition of a conserved haplotype in affected individuals, by meiotic recombination, provides strong evidence to

suggest that genes within that chromosomal region are not linked to the disease phenotype. In large kindreds, this is indicated by a LOD score of less than -2, while in smaller families, the evidence for meiotic recombination is based on direct observation, comparing markers present in individuals in two successive generations. This approach, known as "exclusion mapping" may be employed to rule out the involvement of certain genes or regions of the genome, particularly when inferences of causality cannot be made, or in cases when "positive" results should be treated with caution (Ott 1991).

2.2.2 Genome-Wide Transcriptional Profiling

With the advent of oligonucleotide array technology, it has become possible to examine differential gene expression in wildtype and diseased tissues on a larger scale. This is particularly useful, as it enables the visualisation of several thousand genes at a time, and allows the examination of *in vivo* or *in vitro* consequences of the disease phenotype on gene expression levels. Additionally, genome-wide analyses not only provide clues to the biological candidacy of differentially expressed genes, but may also indicate possible disease-associated pathways (McCarthy et al. 2003).

Despite the obvious advantages, there are major drawbacks to the use of such techniques, such as prohibitive costs and high margin for experimental error, which are exacerbated in the case of small cohorts or families. Additionally, the bias of transcriptional profiling towards detection of transcriptional, rather than translational or post-translational gene expression differences, may make genome-wide transcriptional profiling a less appropriate tool in certain contexts (McCarthy et al. 2003).

2.2.3 Candidate Gene Approach

A third commonly used approach to finding the genetic determinants of disease is the candidate gene approach. This technique is based on a correlation between known gene function and the known (or presumed) biological basis of the disease (McCarthy et al. 2003). The aim is to identify genes with a high prior probability of involvement, and to prioritise these genes for further investigation, either by linkage analysis, transcriptional

profiling or mutation screening. Although less successful in the discovery of disease genes than linkage-based methods, the candidate locus approach is invaluable in situations in which a whole-genome linkage analysis experiment is inappropriate (McCarthy et al. 2003; Ott 1991).

A significant downfall of this approach is that it relies on assumptions – firstly, that the gene function, and the consequence of mutations of the gene, are known and secondly, that the biological basis of the disease has been elucidated. It is often the latter that provides more of an obstacle, since the biology of many diseases, particularly the complex and syndromic cases, remains elusive. Nonetheless, the identification of possible candidate genes by this method may still be a valuable tool in guiding linkage analysis experiments, and other types of investigation (McCarthy et al. 2003; Ott 1991).

2.2.4 The Identification of a Determinant Gene in this Investigation

Because of the small number of individuals available for study in this case – four family members, of whom three were affected, and one deceased – the candidate gene approach was employed, to identify a possible determinant gene. Based on the prominence of fibrosis in the novel disorder, the *TGFB1* gene, encoding the profibrotic cytokine Transforming Growth Factor β1 (TGF-β1), had previously been suggested as a potential candidate (Khumalo et al. 2006). Additional candidates were identified based on their role in the profibrotic signalling cascade initiated by TGF-β1 – these included the signalling proteins Smad2, Smad3, Smad4 and Smad7.

2.3 The Case for the TGF-β1/Smad Signalling Pathway

TGF-β1, the prototype of the TGF-β protein superfamily, is a ubiquitously expressed, multifunctional cytokine, whose cellular effects are mediated by Type I and II serine-threonine kinase receptors and the Smad family of signalling proteins (Blobe et al. 2000; Venkatesan et al. 2004). The role of the TGF-β1/Smad pathway in the promotion of extracellular matrix (ECM)

synthesis, and hence in the development of fibrosis has been well characterised (Arany et al. 2006; Clouthier et al. 1997).

2.3.1 Structure of TGF-β1 Gene and Protein

The *TGFB1* gene is situated on chromosome 19q13, and comprises 7 exons, large introns and 5' and 3' untranslated regions (Awad et al. 1998; Blobe et al. 2000; Kim et al. 1989; Shah et al. 2006). Transcription is regulated by two major transcription start sites, with regulatory regions extending up to 2.7 kilobases upstream of the translation start codon, as determined by Kim et al. (1989) and Shah et al. (2006). These include putative binding sites for the transcription factors Stimulatory Protein 1 (Sp1) and Nuclear Factor, as well as several distal enhancers and repressors (Kim et al. 1989; Shah et al. 2006).

The C-terminal region of the protein encodes mature TGF-β1, while the N-terminal domain comprises the signal sequence (which targets the polypeptide to the Endoplasmic Reticulum (ER) and Golgi apparatus for processing), and the Latency Associated Peptide (LAP) (Munger et al. 1999) (Fig. 2). TGF-β1 is synthesised as a preproprotein, which undergoes extensive modification in the Golgi apparatus, including signal peptide cleavage, glycosylation and dimerisation by means of disulphide bonds (Janssens et al. 2000), to form a homodimer of approximately 75kDa (Annes et al. 2003). After dimerisation, the LAP is cleaved from proTGF-β1 in the Golgi by furin-type enzymes (yielding a 25kDa protein product), but remains non-covalently bound, rendering TGF-β1 inactive and incapable of binding to its receptors (Annes et al. 2003; Janssens et al. 2000; Munger et al. 1999). Association with the Latent TGF-β Binding Protein (LTBP) completes the latent TGF-β1 complex (Fig. 2), and allows it to become covalently associated with the ECM, following secretion (Annes et al. 2003). Under normal conditions, most secreted TGF-β1 can be found in this form (Tesseur et al. 2006).

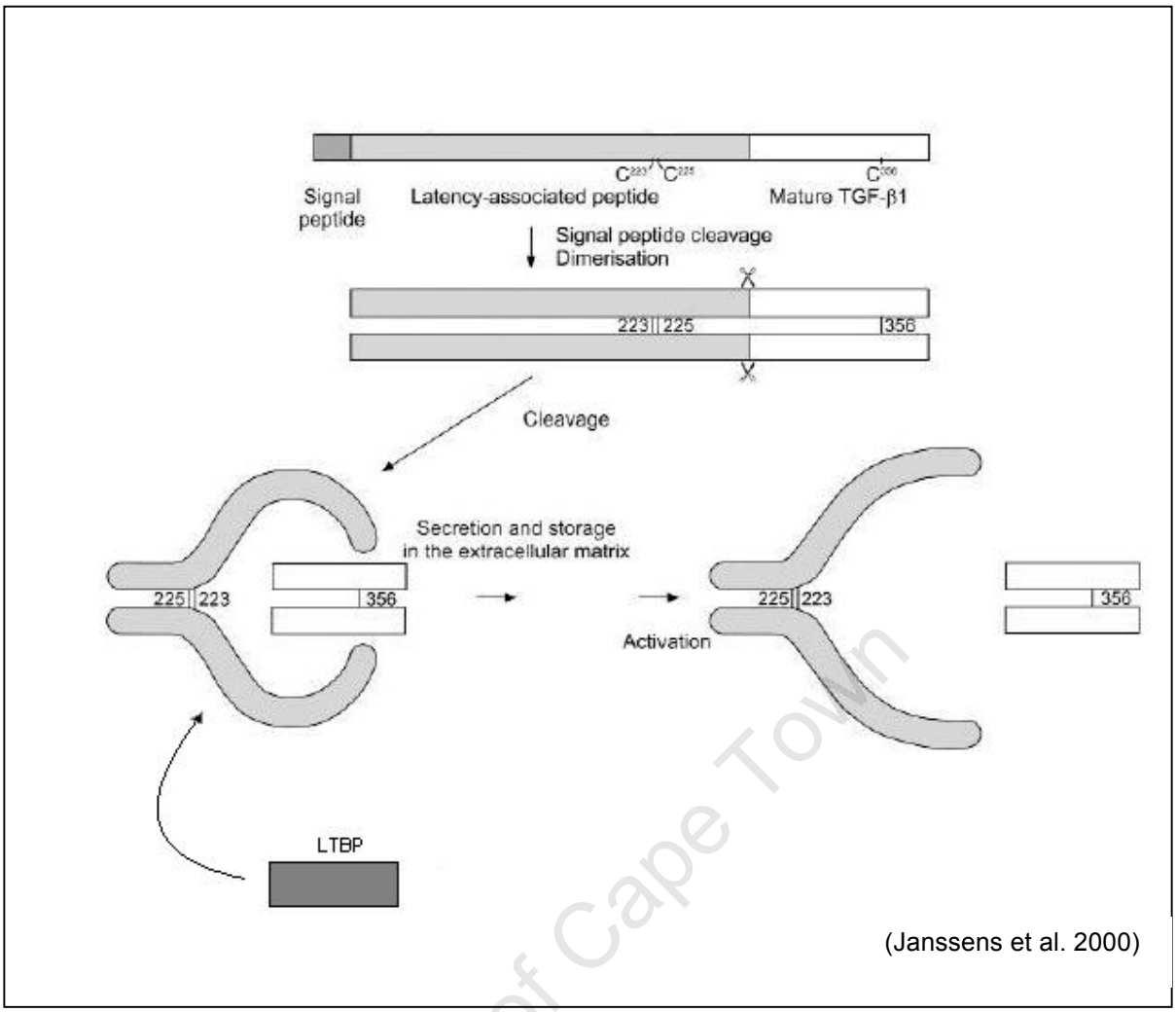


Figure 2. Structure and processing of TGF-β1 prepropeptide. The preprocytokine, comprising signal peptide, Latency Associated Peptide (LAP) and Mature TGF-β1 is directed into the Golgi for modification, at which point the signal peptide is cleaved by signal peptidase. Following dimerisation, the LAP is cleaved from mature TGF-β1, but remains non-covalently associated during secretion and binding to the ECM. Latent TGF-β1 Binding Protein (LTBP) binds to this complex and retains TGF-β1 in an inactive form until environmental stimuli trigger its release.

The mechanisms underlying TGF-β1 activation (that is, the liberation of TGF-β1 from its latent complex) are not fully understood (Annes et al. 2003; Janssens et al. 2000). However, acid activation (*in vitro*) and proteolytic cleavage (*in vivo*) have been demonstrated (Munger et al. 1999). Annes et al. (2003) propose a paradigm in which latent TGF-β1 is considered a sensor of the extracellular environment, with the LTBP acting as the localiser (by facilitating binding to the ECM), the LAP functioning as the sensor (responding to activators by releasing TGF-β1), and TGF-β1 playing the role of the effector, bringing about downstream effects.

2.3.2 Structure of Smad Family Proteins

The Smad family consists of ten proteins which are involved in TGF-β superfamily signalling pathways (Arai et al. 1998 Blobe et al. 2000). So-called because of their homology to *Drosophila* Mad (Mothers Against Decapentaplegic) and *C. elegans* Sma proteins, human Smads may be divided into TGF-β-responsive (Smad2, -3, -4, -6 and -7), and Bone Morphogenic Protein (BMP)-responsive (Smad1, -4, -5 and -8) elements (Heldin et al. 1997; Lo et al. 1998; Massagué 1998; ten Dijke et al. 2000).

Additionally, Smads are classified according to their function. Those which interact directly with, and are phosphorylated by, activated Type I Serine/Threonine Kinase Receptors are designated "Receptor Regulated-Smads" (R-Smads), while those common to both pathways are referred to as "Common Partner Smads" (Co-Smads). Those which act to downregulate signalling, meanwhile, are known as Inhibitory Smads (I-Smads) (Massagué 1998). In the absence of signalling, R-Smads are localised predominantly in the cytoplasm, while Co-Smads are distributed equally between the cytoplasm and nucleus (Ross et al. 2007; van Laar et al. 2005). In the case of TGF-β signalling, there are two receptor-interacting Smads (Smad2 and Smad3), a single Co-Smad (Smad4) and two inhibitory Smads (Smad6 and -7), the functions of which are discussed in more detail in 2.3.3 (Blobe et al. 2000).

The generalised Smad protein structure, shared by the R-Smads and Smad4, consists of two Mad-Homology (MH) domains (the N-terminal MH1 and the C-terminal MH2), which display a high degree of sequence similarity (pairwise sequence identity 40-94% for MH1, and 38-98% for MH2) and are separated by a variable, proline-rich linker region (Fig. 3) (Massagué 1998; Shi et al. 1998). The MH1 domain is responsible for DNA-binding. In the case of the R-Smads and Co-Smads, this domain interacts directly with the promoters of target genes in the nucleus, regulating their transcription in response to TGF-β signalling (Chacko et al. 2001; Lo et al. 1999). This is accomplished via a novel DNA-binding motif, an 11-residue β-hairpin, which contacts DNA in a sequence-specific manner, in the major groove (Kloos et al. 2002; Shi et al. 1998). The exception to this rule is Smad2, which lacks DNA-binding activity

as a result of a 30-bp insertion within this N-terminal region and therefore relies on the DNA-binding abilities of heteromeric partners in order to effect its function (Ross et al. 2007; Shi 1998; Zawel 1998). The MH1 domain also serves to inhibit functional activity of the MH2 domain, prior to receptor-mediated phosphorylation, and contains a nuclear import/export signal, which aids in the transport of heteromeric R-Smad/Co-Smad complexes to and from the nucleus (Ross et al. 2007) (Fig. 3).

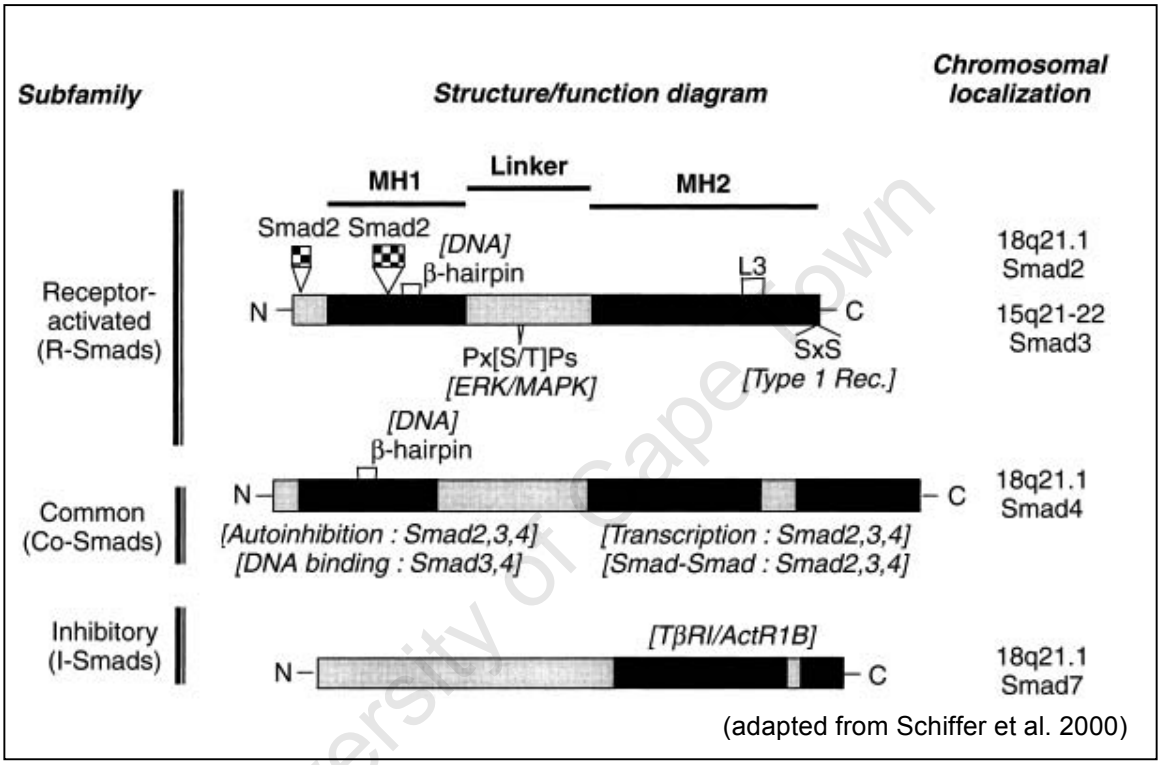


Figure 3. Generalised structure of the Smad protein family, showing representatives of the receptor-activated/receptor-regulated (Smad2 and 3), common-partner (Smad4), and inhibitory Smads (Smad7). MH1 and MH2 represent Mad homology regions 1 and 2, respectively. The functions of these domains are listed below Smad4. Phosphorylation sites for Erk/MAPK (Px[S/T]Ps) and type I TGF-β receptors (SxS) on Smad2 and -3 are indicated. Checkered boxes denote exons present in Smad2 but not Smad3. The β-hairpin DNA-binding domain (Smad2, -3 and -4) and L3 loop (Smad2 and -3) are also indicated. The chromosomal location of each of the human Smads is shown on the right.

The MH2 domain, on the other hand, is responsible for contact with receptors (in the case of R-Smads), homo- and heteromeric complex formation with other Smads and interactions with co-activator or co-repressor proteins (Lo et al. 1998; Lo et al. 1999; Massagué 1998; ten Dijke et al. 2000). Crucial to the

functions of this C-terminal domain are two highly conserved motifs – the L3 loop, and the so-called “C-tail” (Lo et al. 1998; Massagué 1998).

The first of these, the L3 loop, was identified by Lo and colleagues in 1998, as a 17-amino acid loop protruding from the β-sandwich core of the MH2 domain, in the R-Smads and Smad4 (Fig. 3). The sequence of the loop is conserved among both the TGF-β-responsive and BMP-responsive Smads, but differs between the two groups at two amino acid positions. Lo et al. demonstrated that by swopping these variant amino acids between groups, a difference in ligand sensitivity could be induced, leading to phosphorylation of Smad1 in response to TGF-β signalling and phosphorylation of Smad2 in response to BMP signalling *in vitro*. They therefore concluded that the L3 loop was a crucial determinant of the specificity of Receptor/Smad interactions, and postulated that in the case of the non-receptor-interacting Smad4, the loop might function in heteromeric association with the R-Smads (Lo et al. 1998).

The “C-tail” consists of a stretch of 11 amino acids at the extreme C-terminal of the R-Smads, which directly contacts Type I receptors during phosphorylation (Lo et al. 1998). Two serine residues within a conserved sequence motif SS[V/M]S (also designated SSXS) contained in the “tail”, act as a substrate for phosphorylation by activated receptors and it is this phosphorylation which triggers R-Smad homo- and heteromerisation (through electrostatic interactions with L3 loops of neighbouring Smads) and nuclear translocation (Chacko et al. 2001; Massagué 1998; ten Dijke et al. 2000). Thus, the L3 loop and the C-tail together are responsible for R-Smad activation (Lo et al. 1998).

The variable linker region, which joins the MH1 and MH2 domains, has been shown in R-Smads to be phosphorylated by Mitogen Activated Protein Kinases (MAP Kinases), resulting in inhibition of nuclear translocation (Lo et al. 1998; Lo et al. 1999; ten Dijke et al. 2000). The integration of such signals is important in the complex regulation of the Smad signalling pathway (see 2.3.3).

Although the MH2 sequences of Smad4 and the R-Smads are highly similar in sequence, sharing 50% identity, Smad4 is distinguishable by a 35-amino acid insertion, which extends from the core (Chacko et al. 2001). Smad4 also lacks the SSXS motif, and is therefore not phosphorylated by activated serine/threonine kinase receptors (Zhang et al. 1996).

The structure of the I-Smads is markedly different, however, as a result of their distinct function in the Smad signalling cascade. Although they share a similar MH2 domain to that of the other Smad types, these inhibitory proteins contain a more divergent N-terminal region, which lacks a DNA-binding domain (Ross et al. 2007; Takagawa et al. 2003). As with the R-Smads, the MH2 domain mediates receptor binding. However, in this case, the objective is not phosphorylation (I-Smads lack the SSXS motif), but the prevention of R-Smad binding, in order to attenuate the signal (Nakao et al. 1997a; Ross et al. 2007; Takagawa et al. 2003).

2.3.3 TGF-β1/Smad Signalling

Active TGF-β1 signals via a heterodimeric complex of type I and type II receptor serine/threonine kinases (TβRI and II). Ligand binding to TβRII leads to recruitment and phosphorylation of TβRI, which in turn results in the phosphorylation of downstream targets in the Smad signal transduction pathway, propagating the profibrotic signal (Clouthier et al. 1997; Kavsak et al. 2000; Nakao et al. 1997b; Zhang et al. 1996) (Fig. 4).

Specifically, Smad2 and -3 (the TGF-β1-responsive R-Smads) have been shown to associate with the RI-RII complex via the C-terminal L3 loop, and to be phosphorylated by TβRI upon binding (Lo et al. 1998; Zhang et al. 1996). Phosphorylation occurs at the C-terminal SSXS motif (see 2.3.3), and results in a conformational change, releasing the Smad MH2 domains from inhibition by their MH1 domains (Hill 1999). This conformational change also results in dissociation of the Smads from SARA (Smad Anchor for Receptor Activation), a FYVE-domain protein responsible for subcellular localisation of the R-Smads in the cytoplasm, and for their recruitment to activated TGF-β receptors (Chacko et al. 2001; Tsukazaki et al. 1998).

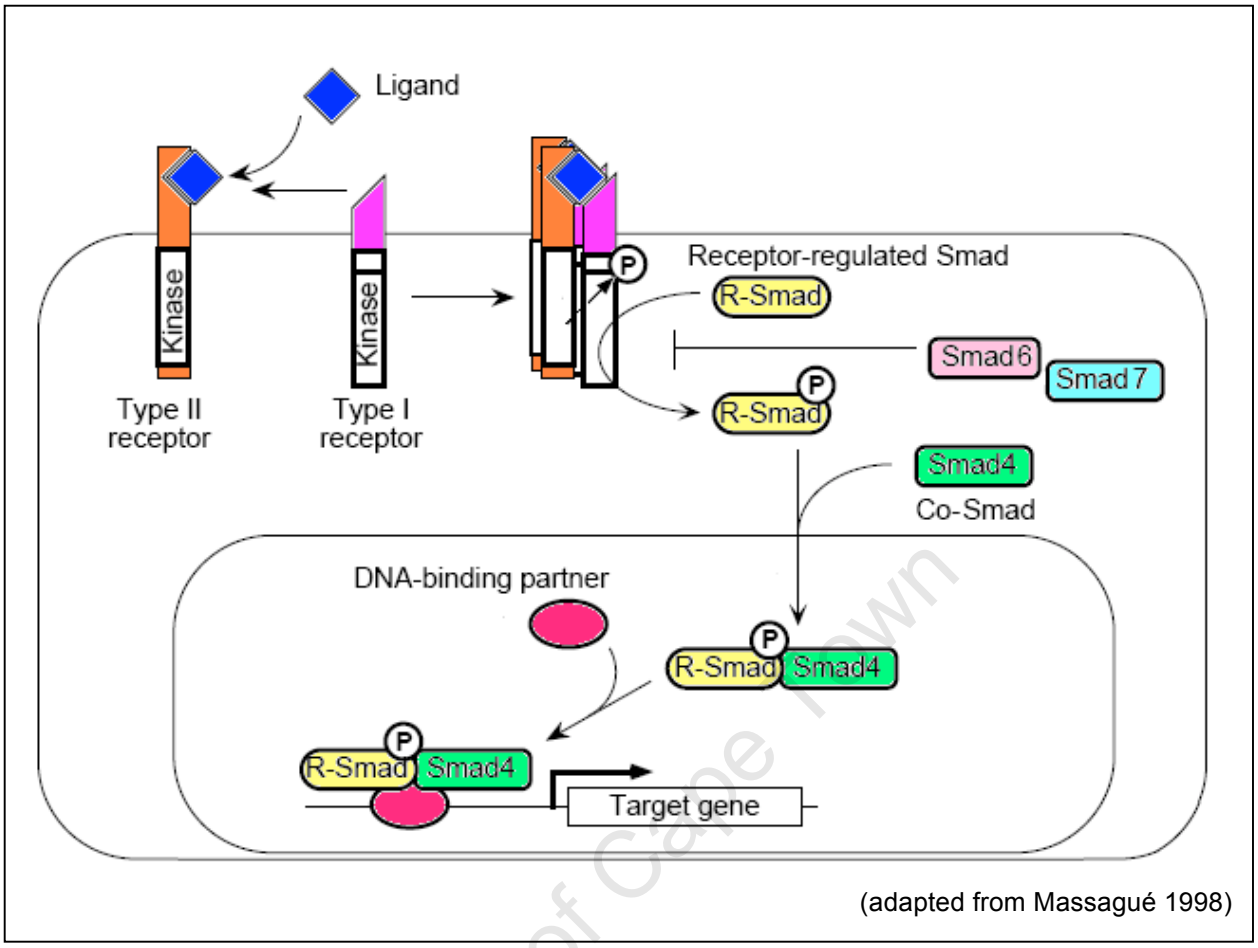


Figure 4. Proposed signalling of TGF-β1 via Smad proteins. TGF-β1 (Ligand) binding to TβRII leads to recruitment and phosphorylation of TβRI. This results in phosphorylation of the receptor regulated Smads (2 and 3), causing a conformational change, and enabling them to bind to the co-Smad, Smad 4. This heteromeric complex is translocated into the nucleus, where it activates transcription of fibrogenic genes. The inhibitory Smads (6 and 7) prevent binding of receptor-regulated Smads to the TβRI/II complex. Smad 7 has also been shown to mediate ubiquitination of the TBRI/II complex.

R-Smad release from SARA is accompanied by the concomitant formation of heterotrimers, consisting of two R-Smads and one molecule of Smad4. MH2 domain interactions act to stabilise the trimers, through the association of R-Smad phosphorylated C-tails with the L3 loops of neighbouring subunits (Chacko et al. 2001). Activated trimeric complexes then enter the nucleus to initiate transcriptional regulation of target genes, such as those involved in ECM production and wound repair (Arai et al. 1998; Chen et al. 2002; Mori et al. 2000; Nakao et al. 1997b; Tsukazaki et al. 1998; Zawel et al. 1998; Zhang et al. 1996).

Transcriptional activation is accomplished through the binding of R-Smad/Co-Smad complexes to a conserved 4-bp binding site, 5'-GTCT-3' (known as the Smad Binding Element or SBE), found in the promoters of TGF-β1 target genes, such as collagenase and collagen type VII. DNA-binding MH1 domains of the complexed Smads contact the SBE in a sequence-specific manner, via hydrogen bonds between MH1 amino acid side chains and DNA bases of this conserved recognition sequence (Johnson et al. 1999; Shi et al. 1998; Zawel et al. 1998).

However, the length of this sequence indicates that it is likely to be present in the promoters of many different genes, simply by chance. The fact that the SBE is not restricted to TGF-β1-responsive genes suggests that a combinatorial approach is required for the specific regulation of target genes in response to TGF-β1 signalling (Shi et al. 1998).

The ubiquitous, constitutive expression of the R-Smad and Co-Smad proteins also points to the importance of tissue- and cell-type specific regulatory transcription factors in determining the expression levels of TGF-β1 signalling targets (Mori et al. 2000). In reviewing the role of Smad proteins in systemic sclerosis, Varga and colleagues suggest that "the specificity of Smad transcriptional responses is achieved, in part, through selective interaction with other regulatory proteins expressed in a cell-specific manner" (Varga et al. 2002).

For instance, R-Smad/Co-Smad complexes may interact with the co-activators CBP/p300 to exert their transcriptional effects (Janknecht et al. 1998; Varga et al. 2002). Interactions are commonly mediated by the Smad MH2 domain, and affinity for co-activators (or for co-repressors) is enhanced following phosphorylation by Type I receptors (Janknecht et al. 1998).

Activation of the TGF-β1/Smad signalling pathway is tightly controlled. Inhibitory Smads, such as Smad6 and Smad7, negatively regulate TGF-β1 signalling by interacting with activated TβRI (thereby inhibiting the binding of R-Smads to receptors) or by interfering with the association of receptor-

regulated Smads with Smad4 (Nakao et al. 1997a; Nakao et al. 2002). This was demonstrated by Kavsak et al. (2000), who showed that Smad7 acts as an adaptor molecule, which recruits Smurf2, an E3 ubiquitin-protein ligase, to the TGF-β1 receptor complex, resulting in its degradation.

Unlike the R- and Co-Smads, which are constitutively expressed even in the absence of signalling, the I-Smads are rapidly induced at the transcriptional level by activated Smad3/Smad4 heteromers (Nakao et al. 2002). Elevated expression of these inhibitory components acts as a negative feedback loop, serving to attenuate signalling in normal tissues (Takagawa et al. 2003; Mori et al. 2000; Nakao et al. 1997a). Autoregulation of signalling also occurs at the level of Smad3, which is downregulated at the mRNA and protein level, following prolonged TGF-β1 signalling (Mori et al. 2000; Varga et al. 2002). In normal tissues, these regulatory mechanisms prevent constitutive signalling, which could lead to overexpression of target genes, and ultimately to disease.

2.3.4 Physiological Role of the TGF-β/Smad Signalling Pathway in the Extracellular Matrix

The TGF-β1/Smad signalling cascade plays a vital role in cell growth, differentiation and tissue repair, as well as in the stimulation and regulation of fibre synthesis in the ECM (Awad et al. 1998; Blobe et al. 2000). In response to tissue injury, increased levels of TGF-β1 contribute to wound healing by promoting transcription of structural proteins (such as collagens), suppressing the production of matrix proteases, and increasing the levels of protease inhibitors, resulting in enhanced ECM deposition (Border and Noble 1994; Varga et al. 2002).

2.3.5 Proposed Role of TGF-β1 in Disease

Although the novel disease described by Khumalo et al. (2006) is clinically heterogeneous, the common factor underlying many of the observed symptoms appears to be widespread fibrosis (Khumalo et al. 2006). The presence of fibrosis in disease is most often due to an imbalance in ECM homeostasis, leading to excessive collagen deposition (Arany et al. 2006).

This is equivalent to hyper-induction of the wound-repair pathway, in which TGF-β1 and its signalling cascade are normally involved (Border and Noble 1994). Thus, it follows that mutations in, or dysregulation of expression of the *TGFB1* gene may be responsible for the pathogenesis of fibrosis, via upregulation of profibrotic target genes in the absence of a stimulus, effectively causing constitutive activation of the TGF-β1 pathway (Dong et al. 2002). The ubiquitous nature of TGF-β1 expression may also provide an explanation for the systemic effects observed in affected individuals (Blobe et al. 2000; Varga et al. 2002).

Links between TGF-β1 overexpression and fibrosis of the lungs, skin and other organs have been demonstrated by several research groups (reviewed by Bartram and Speer 2004 and Blobe et al. 2000). Broekelmann et al. (1991) demonstrated markedly elevated levels of TGF-β1 expression in areas of fibrosis in the lung, which were associated with an increase in the production of collagen. Based on their observations, they proposed a role for TGF-β1 in human pulmonary fibrosis (Broekelmann et al. 1991). These results were supported by the histological findings of Khalil et al. (1996), who discovered increased TGF-β1 expression in epithelial cells of patients with chronic pulmonary fibrosis.

By overexpressing the *TGFB1* gene, coupled to the regulatory sequence of the phosphoenolpyruvate carboxykinase gene, Clouthier and colleagues generated a transgenic mouse model, with ectopic TGF-β1 expression localised to the liver, kidney, and white and brown adipose tissue (Clouthier et al. 1997). In organs in which TGF-β1 was overexpressed, pathological accumulation of ECM components, resulting in severe fibrosis, was observed. Additionally, transgenic mice frequently showed a loss of skin elasticity, secondary to excessive fibre deposition in the underlying adipose tissue. These findings provided strong evidence to support the link between TGF-β1 and fibrotic disease (Clouthier et al. 1997).

In the tight-skin mouse model of systemic sclerosis, Zhu et al. (2005) demonstrated elevated expression levels of TGF-β1 in sclerotic lesional

fibroblasts, which they attributed to polymorphisms in the promoter of the *TGFB1* gene. This is not the first case in which the levels of TGF-β1 protein in the body have been shown to be under genetic control (Grainger et al. 1999). Single nucleotide polymorphisms in the regulatory region of the gene, as well as in the signal peptide sequence, which regulates trafficking of TGF-β1 through the Endoplasmic Reticulum (Crilly et al. 2002), have been found to correlate with increased secretion, and hence higher serum levels, of TGF-β1 (Dunning et al. 2003; Shah et al. 2006). These include a T to C substitution at codon 10 (resulting in an amino acid change from Leucine to Proline), and a C to G (Proline to Arginine) change at codon 25 (Kim et al. 1989; Shah et al. 2006). The presence of a Proline at position 10 has been found to be associated with an increased incidence of lung fibrosis (Awad et al. 1999), poorer prognosis in Idiopathic Pulmonary Fibrosis (Xaubet et al. 2003), and the development of scleroderma (Crilly et al. 2002), all of which are symptomatically associated with the novel disease. Individuals homozygous for Arginine at position 25 have also been shown to have an increased susceptibility to pulmonary fibrosis (Awad et al. 1998).

TGF-β1 was thus concluded to be a logical candidate for investigation, since its functions, and known roles in human pathology correspond to the proposed biological basis of the novel disease.

2.3.6 Proposed Role of Smad Proteins in Disease

The downstream components of the TGF-β1/Smad pathway and their role in human disease have also been well described (see Blobe et al. 2000; and Heldin et al. 1997 for reviews). Similar to TGF-β1, aberrations leading to prolonged or constitutive signalling have been shown to result in upregulation of ECM target genes, leading ultimately to fibrotic disease (Varga et al. 2002). These aberrations include excessive phosphorylation and nuclear accumulation of R-Smads, failure to downregulate Smad3 following TGF-β1 signalling, and decreased expression of inhibitory components of the signalling pathway, such as Smad7 (Varga et al. 2002).

In a bleomycin-induced murine model of scleroderma, Takagawa et al. (2003) demonstrated sustained induction of TGF-β/Smad signalling in dermal fibroblasts, which appeared to be associated with a relative deficiency in Smad7 protein expression. These findings were corroborated by Venkatesan et al. (2004), who showed that decreased levels of Smad7 protein were associated with increased phosphorylation and nuclear translocation of the transcriptional activator R-Smads, Smad2, -3 and -4, in the lung fibroblasts of bleomycin-treated rats.

Dong and colleagues (2002) extended these experiments to human systemic sclerosis, demonstrating a decrease in both basal and TGF-β1-inducible levels of Smad7, in scleroderma lesions (*in vivo*) and cultured scleroderma fibroblasts (*in vitro*). This selective decrease in expression of the inhibitory Smad was associated with an increase in TGF-β1 signalling events, most notably, increased phosphorylation of Smad2 and -3, and elevated transcription of the profibrotic target gene, plasminogen activator inhibitor type-1 (PAI-1). Overexpression of endogenous Smad7, by means of an adenoviral vector, restored normal PAI-1 expression, and led these researchers to suggest that Smad7 deficiency may play a central role in the TGF-β1 hyperresponsiveness, and subsequent fibrosis, observed in scleroderma (Dong et al. 2002).

Removing this key negative regulator of the profibrotic pathway results in an imbalance between activating and inhibitory Smad signals, leading to prolonged signalling in response to TGF-β1, excessive deposition of ECM components and, ultimately, to fibrotic disease (Takagawa et al. 2003; Varga et al. 2002). This imbalance may also result, however, from inappropriately high levels of transcriptionally active R-Smads, such as Smad3 (Dong et al. 2002).

Because Smad proteins lack enzyme activity, the Smad cascade is not amplified. As a result, the duration and intensity of signalling in response to TGF-β1 is extremely sensitive to alterations in the levels of the different Smad proteins (Varga et al. 2002; Verecchia et al. 2001). Thus, failure of

scleroderma fibroblasts to downregulate the transcriptional activator, Smad3, in response to prolonged TGF-β signalling, results in constitutive activation of the Smad pathway, leading to the pathological accumulation of Smad-inducible gene products (Mori et al. 2000; Mori et al. 2003).

Immunohistological analysis of scleroderma skin lesions performed by Dong et al. (2002) further supported these findings, demonstrating markedly increased Smad3 staining in biopsies taken from sclerotic lesions, relative to those taken from normal tissues. These elevated expression levels were also associated with increased R-Smad phosphorylation, indicative of hyperactive TGF-β1 signalling. Taken together with the deficiency in Smad7 expression mentioned above, it seems evident that the ratio of R-Smads to I-Smads is dramatically perturbed in systemic sclerosis. This may provide a model for the molecular mechanism of fibrosis seen in other disorders (Dong et al. 2002).

Ligand-independent phosphorylation and nuclear accumulation of R-Smads have also been demonstrated to cause fibrosis – in both the liver and in fibroblast cultures taken from patients with systemic sclerosis (Varga et al. 2002). In these cases, disruption of the microtubule cytoskeletal network prevents cytosolic retention of the R- and Co-Smads, leading to transcriptional upregulation of profibrotic gene targets by R-Smad/Co-Smad complexes which accumulate in the nucleus, in the absence of TGF-β1 signalling (Dong et al. 2000).

Based on their observed role in fibrotic disease, therefore, the R-Smads (Smad2 and -3) and the inhibitory Smad7 were prioritised as candidates for investigation. The phosphorylation state of Smad3 (indicative of the level of TGF-β1 signalling) was also examined.

2.4 Previous Research and Rationalé for this study

Previous research performed on this family involved linkage-based exclusion mapping (see 2.2.1), to determine the presence of a conserved *TGFB1* haplotype in all affected individuals. Additionally, the coding region and a 538bp portion of the regulatory region of *TGFB1* (incorporating an upstream promoter and negative regulatory element) were sequenced, in an affected and an unaffected family member, to identify potential disease-causing mutations. The results indicated that the chromosomal region surrounding *TGFB1* in affected individuals had not undergone meiotic recombination, suggesting that affected individuals may share a conserved copy of the gene. However, mutation screening failed to demonstrate the presence of a causal genetic mutation in either the proximal regulatory or coding regions of *TGFB1*. Out of six polymorphisms identified, only three segregated with the disease, all of which were common variants and were therefore unlikely to be responsible for the novel phenotype observed (Hartmann 2006).

It was therefore hypothesised that the disease-causing defect may lie in a region of *TGFB1* not yet investigated (for instance, in a distal enhancer/repressor sequence or in a splice donor/acceptor site), or in a different, but related component of the fibrotic pathway. For this reason, the decision was made to examine the *TGFB1* transcript, for alterations in structure (indicative of whole exon deletions or splice-site aberrations, which could not be detected by direct exonic sequencing), and expression (indicative of alterations in distal regulatory sequences).

Downstream signalling components with a proposed role in fibrotic disease, were investigated at the protein level (in the case of Smad2 and -3) or the transcriptional level (in the case of Smad7), to determine whether alterations in expression could account for the excessive fibrosis observed. TGF-β1 responsiveness, in the form of phosphorylated R-Smad levels, was also examined.

2.5 Aim and Objectives

The aim of this project was to determine the role of *TGFB1* and the Smad proteins in the molecular basis of the novel familial fibrotic disease described by Khumalo et al. (2006).

2.5.1 Objectives

1. Determine mRNA expression levels of *TGFB1* in cultured fibroblasts from an affected and an unaffected individual, using quantitative real-time PCR.

2. Identify possible splice-site mutations and/or whole exon deletions by comparing the length and sequence of the *TGFB1* transcript in an affected and an unaffected individual.

3. Quantify basal and TGF-β1-inducible levels of downstream signaling components (Smad2/3 and 7) in fibroblast whole cell lysates from an affected and an unaffected individual by immunoblotting ("western blot") and quantitative real-time PCR.

4. Analyse the cellular response to TGF-β1 stimulation, by comparing basal level of Smad phosphorylation to the stimulated level following exposure of affected and unaffected fibroblasts to TGF-β1, by immunoblotting.

3. Materials and Methods

3.1 Analysis of *TGFB1* mRNA

The *TGFB1* transcript was examined to investigate the possibility that splice variants, or elevated expression levels of *TGFB1* could account for the observed symptoms in affected individuals. Total RNA from cultured fibroblasts of an affected (RTSAD1.3) and an unaffected individual (RTSAD1.4) (Fig.1) was reverse-transcribed to yield cDNA, from which the *TGFB1* transcript was analysed for length and sequence aberrations and for differences in expression levels between the individuals.

3.1.1 Cell Culture

Human skin fibroblasts were cultured from punch biopsies taken from family members RTSAD1.3 (affected son) and RTSAD1.4 (unaffected mother), after informed consent had been obtained (UCT Research Ethics Committee Number 492/2007). Cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) (SIGMA-ALDRICH, Missouri, USA or GIBCO®, Invitrogen Corporation, California, USA), containing 1000mg/L D-glucose, 4mM L-Glutamine, 110mg/L Sodium Pyruvate, and 10% Foetal Calf Serum (SIGMA-ALDRICH or GIBCO®) at 37°C, in a humidified atmosphere containing 10% CO_2.

3.1.2 Isolation of Total RNA from Fibroblasts

Cultured RTSAD1.3 and RTSAD1.4 fibroblasts, at passage seven, were washed with 1x sterile-filtered Phosphate Buffered Saline (PBS) (Bioshop Canada Inc., Ontario, Canada), treated with trypsin-EDTA (SIGMA-ALDRICH) for 2 minutes at 37°C, and washed again with 12ml DMEM (SIGMA-ALDRICH), followed by centrifugation for 10 minutes at 1000rpm in an MSE bench-top centrifuge (Henderson Biomedical Ltd., UK). The cell pellet was resuspended in 10ml sterile filtered 1x PBS (Bioshop Canada Inc.), and centrifuged for a further 10 minutes at 1000rpm. After aspiration of the supernatant, total RNA was isolated from the cell pellet using the High Pure RNA Isolation Kit (Roche Diagnostics, Indianapolis, USA), following

manufacturer's instructions (Appendix D). Isolated RNA was stored at -80°C in 10µl aliquots.

The concentration of RNA samples was determined using the NanoDrop Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The integrity of the RNA was also ascertained by subjecting 10µl of each sample, mixed with 3µl of loading buffer (Appendix A), to agarose gel electrophoresis for 20min at 120V, on a 1% agarose low electroendosmosis gel (Roche Diagnostics) using Tris-Borate EDTA Electrophoresis Buffer (TBE) (891.8mM Tris base; 889.4mM Boric Acid; 19.9mM EDTA, pH8), and incorporating Ethidium Bromide (SIGMA-ALDRICH) (0.5µg/ml) for visualisation (Appendix A). To determine the relative molecular mass of the bands observed, 500ng of the Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas International Inc., Hanover, USA) (Appendix B) was included as a size standard. DNA was visualised after electrophoresis using the UVIPro Gel Documentation System (UVITec, Cambridge, UK), and images were acquired using UVIPro software v12.3 for Windows (Copyright 1999-2005 UVITec).

3.1.3 cDNA Synthesis

cDNA was synthesised from 1µg of total RNA extracted from fibroblasts, using the Transcriptor First Strand cDNA Synthesis Kit (Roche). In order to maximise the possibility of generating full-length cDNA, oligo$(dT)_{18}$ primers complementary to the mRNA poly(A) tail were used, and the manufacturer's instructions were followed (Appendix D). Denaturation of the template-primer mix was performed on a Touchdown Thermal Cycler (ThermoHybaid, Middlesex, UK), while the reverse transcription and inactivation steps were performed using a Px2 Thermal Cycler (ThermoHybaid). The resulting cDNA was stored at -20°C.

The concentration of cDNA samples was determined using the NanoDrop Spectrophotometer (NanoDrop Technologies). Where necessary (for instance, in the case of cDNA synthesised from RNA retrieved after break down of the -80°C freezer), the integrity of cDNA samples was determined by

subjecting a 5µl aliquot of each sample (mixed with 3µl loading dye) to agarose gel electrophoresis on a 1% agarose gel in TBE buffer (containing 0.5µg/ml Ethidium Bromide) and visualisation under UV light (as described in 3.1.2). For approximate size determination, 500ng of the Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas) was included.

Newly-synthesised cDNA was checked for possible genomic contamination by polymerase chain reaction (PCR), using primers designed to amplify exon 3 of *TGFB1* (Forward: 5'- TCAGTTAGTGTATGTGGGGTCG -3'; Reverse: 5'-GTCTCAGCACTTTCACACCAG -3'), which had previously been synthesised using the Beckman 1000M DNA Synthesizer in the Department of Molecular and Cell Biology, UCT. These primers are complementary to intronic regions flanking exon 3; thus amplification should only occur in the presence of genomic DNA, since cDNA contains no introns.

The PCR reaction mixture contained 1X GoTaq™ Reaction Buffer (1.5mM $MgCl_2$, pH8.5) (Promega, Madison, WI, USA), forward and reverse primers (0.4µM each), deoxyribose nucleotide triphosphates (dNTPs) (Bioline, Michigan, USA) (0.2µM each), 1µl of template (either RTSAD1.4 genomic DNA at 100ng/µl, or RTSAD1.4 cDNA of variable concentration) and 0.5 Units (U) of GoTaq™ DNA polymerase (Promega), in a final volume of 25µl. PCR contamination was determined by the inclusion of a DNA-free water control, in which distilled water was substituted for the template. This reaction mix, hereafter referred to as the "Standard Reaction Mix", is listed in Table 1, and was used for all PCR reactions unless otherwise stated.

Table 1. Standard PCR mix (for a final volume of 25µl)

Reagent	Final Concentration
Forward Primer	0.4µM
Reverse Primer	0.4µM
dNTPs (Bioline)	0.2µM
GoTaq Buffer (Promega)	1X
GoTaq DNA Polymerase (Promega)	0.5U
Template (genomic DNA or cDNA)	4ng/µl

Reaction conditions were as follows: An initial denaturation step at 95°C for 5 minutes, followed by thirty cycles of 95°C for 30 seconds, 58°C for 30 seconds, and 72°C for 30 seconds, with a final 7 minute elongation step at 72°C. The reaction was performed on a Px2 Thermal Cycler (ThermoHybaid).

After the addition of 5µl loading dye, PCR products were electrophoresed on a 1% agarose gel in TBE buffer, incorporating 0.5µg/ml Ethidium Bromide (SIGMA-ALDRICH), for 20 minutes at 120V, and visualised under UV light as previously described (3.1.2). As before, 500ng of the Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas) was included as a size standard, to determine the molecular mass of the bands observed.

3.1.4 Determination of TGFB1 Transcript Length and Sequence

Since the presence of splice variants or whole exon deletions could lead to length and sequence aberrations, the length of the *TGFB1* transcript was verified by agarose gel electrophoresis, and the entire transcript sequenced in both an affected and an unaffected individual.

3.1.4.1 Primer Design

The *TGFB1* mRNA reference sequence was obtained from the NCBI database (http://www.ncbi.nlm.nih.gov/, GeneID 7040, accession number NM_000660). Information about the *TGFB1* gene, including the complete DNA sequence, had previously been retrieved from the NCBI and Ensembl (www.ensembl.org) databases (NCBI accession number NC_000019.8, sub-sequence 46523491 to 46556656), and annotated using the automatic

annotation programme, Annot v9 (Dr George Rebello, personal communication). This annotated gene sequence, which included the positions of SNPs, splice sites, exons and introns, was used to map exon boundaries onto the mRNA. Based on this gene annotation, primers were designed to amplify the shortest possible cDNA fragment that included every exon boundary.

Suitable oligonucleotide pairs, of between 18 and 22 bases in length, with a GC content of 45-65% were identified from the mRNA sequence (in the case of forward primers) and its reverse complement (in the case of reverse primers), and subjected to a sequence similarity search, using the genomic DNA and reference mRNA databases available at NCBI BLAST (Basic Local Alignment Search Tool) (www.ncbi.nlm.nih.gov/BLAST). Those oligo-nucleotides whose sequences appeared to be unique to the *TGFB1* transcript were analysed further. The melting temperatures of primer pairs were matched using OligoCalculator (http://trishul.sci.gu.edu.au/tools/OligoCalculator.html) and the potential formation of hairpin, homo- and heterodimer structures, was analysed using OligoAnalyzer (Integrated DNA Technologies, (www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx)

Two primer pairs were finally selected (Table 2). The first pair (F1 and R1), were positioned in exons 1 and 7 respectively, generating an amplicon of 1608bp in length, which covered all exon-exon boundaries. This amplicon is referred to hereafter as the “whole transcript”. Since these primers were too far apart to allow for adequate sequencing of the transcript, a second, internal primer pair (F2 and R2) was designed, amplifying a 985bp fragment, representing exons 1 to 6 of the transcript and overlapping slightly with the region which had been sequenced using F1 and R1. This amplicon is referred to hereafter as the “internal fragment”. Primers were synthesised using the Beckman 1000M DNA Synthesizer in the Department of Molecular and Cell Biology, UCT. The positions of the primers on the mRNA reference sequence, relative to the exon-exon boundaries, can be found in Appendix C.

Table 2. Primers used to amplify *TGFB1* transcript. The melting temperature of each primer and the expected amplicon sizes (based on the NCBI reference sequence) are also indicated.

Amplicon	Primers	Tm (°C)	Expected Amplicon Length
Whole Transcript	F1 5'- CTACACGGCGTCCCTCAGG-3'	64	1608bp
	R1 5'- GCAGGAGCGCACGATCATGTTG -3'	70	
Internal Fragment	F2 5'- CTACCTTTTGCCGGGAGACC-3'	64	985bp
	R2 5'- GCAGCAGTTCTTCTCCGTGG-3'	64	

3.1.4.2 PCR Optimisation

PCR reactions are optimised by the alteration of a single parameter at a time, so that the effect of each individual alteration on the efficiency of the reaction can be determined. These parameters may include annealing temperature, Magnesium Chloride ($MgCl_2$) concentration, pH, the addition of additives such as Dimethyl Sulphoxide (DMSO), or the manipulation of other reaction conditions (denaturation and elongation times and temperatures).

For both the whole transcript and internal fragment, optimisation involved determination of the optimum annealing temperature for the amplification of each fragment. For the purposes of PCR, the melting temperature (T_m) of each primer was calculated manually using the formula $T_m = 4(G+C) + 2(A+T)$, and the annealing temperature (T_a) was estimated to be 5°C lower than the lower T_m of the pair. A temperature gradient was then selected, based on these calculations. In both cases, the gradient consisted of three reaction tubes, one with an annealing temperature corresponding to the optimum calculated (59.4°C), and the other two with annealing temperatures approximately 2°C above or below the calculated optimum (57.3°C and 61.6°C respectively).

The standard reaction mix described in Table 1 was used, with the exception of the template. In this case, 1µl of cDNA from the unaffected individual (RTSAD1.4) was used for optimisation. A template-free control, in which 1µl of water was substituted for the template, was included to check for PCR

contamination, and was subjected to an annealing temperature of 59.4°C. The temperature gradient PCR was performed on the Px2 Thermal Cycler (ThermoHybaid), with the following reaction conditions: 95°C for 5 minutes, followed by 35 cycles of 95°C for 60 seconds, T_a (57.3°C, 59.4°C or 61.6°C) for 60 seconds and 72°C for 80 seconds. A final elongation step at 72°C for 7 minutes completed the reaction.

3.1.4.3 Agarose Gel Electrophoresis

Following the temperature gradient PCR, the PCR products (25µl) were each mixed with 5µl loading buffer, and electrophoresed on a 1% agarose gel in TBE buffer, containing 0.5µg/ml Ethidium Bromide (SIGMA-ALDRICH), for 40 minutes at 110V and visualised under UV light as previously described (3.1.2). To determine the molecular mass of the bands observed, 500ng of the Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas) was included as a size standard.

The optimum annealing temperature was determined to be that temperature at which the relevant band on the agarose gel showed either the brightest fluorescence, indicative of a high concentration of PCR product, or, where all bands were equally bright, the least amount of non-specific amplification.

3.1.4.4 PCR in Family Samples

Following optimisation, each amplicon was amplified from RTSAD1.3 (affected) and RTSAD1.4 (unaffected negative control) cDNA. The Standard Reaction Mix was used, with primers specific to the amplicon of interest, and 1µl of template cDNA (variable concentration) per reaction. PCR reactions were performed in duplicate, to counteract loss of PCR product sustained during gel extraction (see 3.1.4.5).

DNA Fragments were amplified on a Px2 Thermal Cycler (ThermoHybaid) using the following cycling conditions: 95°C for 5 minutes, 35 cycles of 95°C for 60 seconds, T_a (amplicon-specific, determined during optimisation) for 60 seconds, and 72°C for 80 seconds, followed by a 7 minute elongation step at 72°C.

3.1.4.5 Determination of Amplicon Length and Preparation of PCR Products for Sequencing

PCR products (25µl combined with 5µl Loading Buffer) were electrophoresed for 40 minutes at 110V on a 1% agarose gel (in 1X TBE) incorporating 0.5µg/ml Ethidium Bromide. As in previous cases, the Gene Ruler™ 100bp DNA Ladder Plus *Improved* (0.05µg/µl) (Fermentas) was included for size determination. Gels were viewed using a UV transilluminator (UVITec), and images were acquired using UVIPro software v12.3 for Windows. At this point, the size of the PCR product generated from the cDNA of the affected individual (RTSAD1.3) was compared to that of the unaffected individual, based on their electrophoretic mobility, relative to that of the Gene Ruler™ size standard. Both were then checked against the expected length of the amplicon (based on the NCBI reference sequence, accession number NM_000660) to detect possible variations.

Agarose blocks containing the PCR products were excised from the gel under UV light. Since PCR reactions had been performed in duplicate, bands containing amplified cDNA from the same individual were pooled. PCR products were purified from the agarose using the QIAquick® Gel Extraction Kit (QIAGEN, Hilden, Germany). Manufacturer's instructions for microcentrifuge extraction were followed, except for the final elution step, in which an equivalent volume of distilled water was substituted for the supplied elution buffer (Appendix D).

3.1.4.6 Agarose Gel Quantification of PCR Products

A 5µl aliquot of each PCR product extracted from agarose (mixed with 3µl loading dye) was subjected to agarose gel electrophoresis for 40 minutes at 110V, on a 1% agarose (0.5µg/ml Ethidium Bromide) gel in 1X TBE Buffer, and viewed using the UVIPro Gel Documentation System (UVITec). The gel was photographed using UVIPro software v12.3 for Windows (UVITec) and analysed using UVISoft UVIBand Windows Application v11.9 (UVITec). The amount of PCR product contained in each band in the agarose gel was determined from its fluorescent intensity, relative to the fluorescent intensity of the bands of the Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight

Marker *Improved* (Fermentas), for which the quantity of DNA was known (supplied by the manufacturer – see Appendix B). From this value, the concentration of PCR product retained after gel extraction could be calculated.

3.1.4.7 Cycle Sequencing

Cycle sequencing was performed using the ABI PRISM® BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), following manufacturer's instructions outlined in the Automated DNA Sequencing Chemistry Guide (Applied Biosystems, Copyright 1998). The reagents and their concentrations (in a final volume of 10µl) are listed in Table 3. Because of the extreme length of the fragments sequenced (~1kb), approximately 100ng of PCR product was used per reaction. The concentrations of Terminator Mix and Sequencing Buffer used were those recommended for a full reaction (for fragments >500bp in length), while 1µM of primer (forward or reverse) was used for each reaction.

Table 3. Reagents and quantities added to the cycle sequencing reaction (Final volume 10µl).

Reagent	Concentration/Amount
PCR Product	100ng
Primer (Forward or Reverse)	1µM
BigDye™ Terminator Mix	4.0µl
Sequencing Buffer	1.0 µl

Reaction conditions were as follows: an initial denaturation step at 98°C for 5 minutes, followed by 30 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes. Cycle sequencing was performed on the GeneAmp 9700 PCR System (Applied Biosystems).

3.1.4.8 Ethanol Precipitation

Following cycle sequencing, the amplified DNA was purified from contaminating reagents (such as dNTPs, excess primer or DNA polymerase) by subjecting the reaction mixture to ethanol precipitation. Two microlitres of 1.5M Sodium Acetate (pH>8, 250mM EDTA) (Atria Genetics, San Francisco,

CA, USA) and 2.5 volumes of 100% ethanol were added to the reaction mixture, which was incubated at -20°C for at least 2 hours. This mixture was centrifuged for 10 minutes at 10 000rpm on an Eppendorf 5145D microcentrifuge (Eppendorf, Hamburg, Germany). The supernatant was discarded, and the pellet resuspended in 20µl 70% ethanol, followed by a second centrifugation step for 10 minutes at 10 000rpm. The supernatant was again discarded and the pellet left to air dry for at least 1 hour at room temperature. Finally, the pellet containing purified, amplified PCR product for sequencing, was resuspended in 10µl Hi-Di™ formamide (Applied Biosystems).

3.1.4.9 Automated DNA Sequencing

Automated sequencing was performed on the ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems) at 60°C, using the POP™-6 separation polymer (Applied Biosystems) (França et al. 2002). To obtain the longest possible readable sequence, sequencing was repeated several times for each primer, extending the analysis period on the machine where necessary.

Sequencing results were collected using ABI PRISM® 3100 Data Collection software (Applied Biosystems), and analysed initially using Sequencing Analysis v3.7 software (Copyright 1989-2001, Applied Biosystems). For each amplicon, sequences generated using either the forward or reverse primers were edited, reverse complemented where applicable, and aligned with the reference sequence (obtained from NCBI) using the ClustalW Multiple Alignment accessory application available in BioEdit Sequence Alignment Editor v7.0.0.

3.1.5 Real-time (Quantitative) PCR of TGFB1 Transcript

To determine whether *TGFB1* was overexpressed in affected individuals, *TGFB1* transcript levels in cultured RTSAD1.3 and RTSAD1.4 fibroblasts were determined by real-time quantitative PCR (qPCR), using SYBR Green reaction chemistry.

SYBR Green is a DNA-binding dye, which emits fluorescent light of wavelength 530nm upon intercalation with double-stranded DNA. Thus, it can be used to measure the generation of PCR products in real time, as its fluorescent emission is proportional to the amount of double-stranded product present in the reaction tube. In qPCR, fluorescence is measured by the real-time PCR instrument (in this case, the LightCycler 1.5) at the end of the elongation step of every cycle, to monitor the increase in amplified product produced (Roche Applied Science LightCycler® DNA Master SYBR Green I Handbook, Roche Diagnostics, Mannheim, Germany).

The cycle in which the first significant increase in fluorescence (above the background fluorescence present at the start of the PCR reaction) is detected, is referred to as the crossing point, or threshold cycle (C_T), and is used as a tool for calculating the starting amount of template in each sample. Samples with a higher starting concentration of template will generate more products in a shorter number of cycles, and will therefore have a lower C_T (QuantiFast SYBR Green PCR Handbook 01/2007, QIAGEN).

However, because SYBR Green binds DNA non-specifically, the formation of primer dimers may affect the levels of fluorescence detected during the reaction. To prove that the increase in fluorescence observed is a result of specific amplification of the desired product, melting curve analysis may be performed following amplification. As the reaction mixture is heated to 95°C, double stranded DNA is denatured, and SYBR Green fluorescence decreases. This decrease in fluorescence is monitored by the qPCR instrument, which displays it as a melting curve, with the peak of the curve corresponding to the melting temperature of the DNA fragment. Since length and GC content significantly influence the melting temperature of DNA, the melting peak of any given fragment is specific to that sequence. Thus, if the PCR reaction generates only one amplicon, only one melting peak will be yielded. However, if primer dimers or non-specific products are present, additional melting peaks will be generated, and may also be present in the No-Template Control. Specificity of qPCR, and the presence of primer dimers, may also be determined by subjecting the reaction mixture to agarose

gel electrophoresis, following PCR (Roche Applied Science LightCycler® DNA Master SYBR Green I Handbook, Roche Diagnostics).

Analysis of the results generated by qPCR was performed using the $2^{-\Delta\Delta CT}$ method of relative quantification, which describes the expression levels of a target gene (in this case, *TGFB1*) in an experimental group (the affected individual), relative to a reference group (the unaffected control individual). A reference gene is incorporated to normalise the PCR reactions for the amount of cDNA used as a template (Livak and Schmittgen 2001). For these experiments, the housekeeping gene Glucuronidase Beta (*GUSB*) was used as a reference (Glenn et al. 2007; Lee et al. 2006).

3.1.5.1 Determination of Amplification Efficiencies for Target (TGFB1) and Reference (GUSB) Genes

In order for the $2^{-\Delta\Delta CT}$ method to be valid, the target and reference genes must be amplified with approximately equal efficiencies. To determine whether this was the case, a dilution series was prepared using control cDNA, previously synthesised from RNA isolated from blood taken from an unrelated individual, using the ImPromII Kit (Promega, supplied by J. Scholefield). Control cDNA was diluted to 100ng/µl, 10ng/µl and 1ng/µl in sterile distilled water, and concentrations were confirmed using the NanoDrop spectrophotometer (NanoDrop Technologies). Real-time qPCR was performed in triplicate on these cDNA dilutions, using commercially available primer assays for either *TGFB1* or *GUSB*.

A master mix, containing the QuantiFast SYBR Green PCR Master Mix (QIAGEN), the Quantitect Primer Assay specific to either the target or reference gene (QIAGEN) and RNase-free water, as described in Table 4, was prepared, mixed thoroughly by vortexing, and aliquoted into precooled LightCycler capillaries (Roche Diagnostics). To each capillary, 2µl of cDNA (at a concentration of 100, 10 or 1ng/µl) was added, to give a final reaction volume of 20µl. To determine the presence of primer dimers, or PCR contamination, a No-Template Control (NTC) was included for each primer assay, in which template cDNA was substituted with an equal volume of

RNase-free water. The capillaries were sealed and centrifuged at 3 000 rpm for 5 seconds, in an Eppendorf 5145D microcentrifuge, before being loaded into the LightCycler sample carousel (Roche Diagnostics).

Table 4. qPCR Reaction Mix (for a final volume of 20µl)

Reagent	Volume	Final Concentration
2xQuantifast SYBR Green PCR Master Mix (QIAGEN)	10µl	1x
10x Quantitect Primer Assay (QIAGEN) *GUSB* or *TGFB1*	2µl	1x
RNase-free water (QIAGEN)	6µl	-
Template cDNA (or RNase-free water)	2µl	variable

The following reaction conditions were used: activation programme (95°C for 5 minutes), amplification and quantification programme, repeated 35 times (95°C for 10 seconds, 60°C for 30 seconds, with a single fluorescence measurement, acquired during annealing), melting curve analysis programme (95°C for 0 seconds, 50°C-95°C, with a heating rate of 0.1°C per second and a continuous fluorescence measurement) and a final cooling step to 40°C. This is summarised in Table 5, and is referred to hereafter as the "standard qPCR reaction conditions". Quantitative PCR was performed using the LightCycler 1.5 (Roche Diagnostics), and amplification and melting curves were analysed using LightCycler Software v3.5 (Roche Diagnostics). Crossing points were determined using the second derivative maximum method.

Table 5. Standard qPCR Reaction Conditions

Analysis Mode	Cycles	Segment	Target Temperature	Hold Time	Acquisition Mode
Activation					
None	1	1	95°C	5min	none
Amplification					
Quantification	35	Denaturation	95°C	10s	None
		Annealing	60°C	30s	Single
Melting Curve					
Melting Curves	1	Denaturation	95°C	0s	None
		Annealing	50°C	30s	None
		Melting	95°C	0s	Continuous
Cooling					
None	1	1	40°C	30s	none

Results were exported to Microsoft Office Excel 2003 (Copyright 1983-2003 Microsoft Corporation). For each cDNA dilution, the mean crossing point (C_T) was determined, for both *GUSB* and *TGFB1*, and the ΔC_T was calculated ($C_{T,TGFB1} - C_{T,GUSB}$). The ΔC_T was then plotted against the log of cDNA concentration. If the absolute value of the slope generated is close to zero (<0.1), then the amplification efficiencies of the target and reference genes can be said to be similar, and the $2^{-\Delta\Delta CT}$ method of relative quantification can be used with confidence (Livak and Schmittgen 2001).

3.1.5.2 qPCR from Family Samples

Following efficiency calculations, qPCR amplification of both *GUSB* and *TGFB1* was performed, using cDNA prepared from the affected (RTSAD1.3) and unaffected (RTSAD1.4) family members. The qPCR master mix described in Table 4 was used, and capillaries were prepared as described in 3.1.5.1. Two microlitres of template cDNA, at a concentration of 100ng/µl, was added to each capillary except in the case of the NTCs, in which 2µl of water was substituted for cDNA. For each sample (RTSAD1.3 or RTSAD1.4), reactions were performed in triplicate and repeated twice. Amplification was performed using the standard qPCR reaction conditions, on the LightCycler 1.5, with analysis using LightCycler Software v3.5.

3.1.5.3 Analysis using the $2^{-\Delta\Delta CT}$ Method

Analysis of the results of qPCR performed on the family samples was done by means of the $2^{-\Delta\Delta CT}$ method described by Livak and Schmittgen (2001). Crossing points determined from replicate qPCR reactions were averaged, to obtain a mean crossing point and a standard deviation. A maximum standard deviation of 0.4 was deemed acceptable, and outlying values were disregarded. For each individual (RTSAD1.3 and RTSAD1.4), the average C_T for *GUSB* was subtracted from the average C_T for *TGFB1*, to obtain the ΔC_T. These values were then used to calculate $\Delta\Delta C_T$, by subtracting the ΔC_T for RTSAD1.4 from the ΔC_T for each sample (RTSAD1.3 and RTSAD1.4). The fold change in expression was determined using the formula $2^{-\Delta\Delta CT}$.

These calculations can be summarised in Equation [1] and [2] as follows:

$$\Delta\Delta C_T = (C_{T,\,TGFB1} - C_{T,\,GUSB})_{individual\ X} - (C_{T,\,TGFB1} - C_{T,\,GUSB})_{RTSAD1.4}, \quad [1]$$

where individual X may be either RTSAD1.3 or RTSAD1.4, and

$$\text{fold change in } TGFB1 \text{ expression} = 2^{-\Delta\Delta CT}, \quad [2]$$

relative to RTSAD1.4. Standard deviations were carried over from ΔC_T to $\Delta\Delta C_T$, and were used to create a range for the fold change (by calculating $2^{-\Delta\Delta CT}$ for $\Delta\Delta C_T \pm SD$).

3.1.5.4 Statistical Analysis

Statistical analysis of the fold change in *TGFB1* expression between the individuals was performed in Microsoft Excel (Copyright 2003, Microsoft Corporation), using the Student's *t*-test for unpaired samples. A two-tailed distribution, assuming unequal variances, was used to analyse the data. Significance was defined using a type I error, or *p*-value, of 0.05.

3.1.5.5 Determining Specificity of qPCR and Checking for the Presence of PCR Contamination

For each qPCR reaction performed, 10µl each of representative PCR products (one replicate per individual per gene, and one no-template control per gene) was mixed with 3µl loading buffer (Appendix A), and electrophoresed on a 1% agarose gel in TBE buffer, containing 0.5µg/ml Ethidium Bromide, for 40 minutes at 110V, and visualised under UV light as previously described (3.1.2). To determine the molecular mass of the band(s) observed, 500ng of the Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas International Inc.) was included as a size standard. The number and size of bands, as well as the presence or absence of PCR product in no-template control reactions (indicative of PCR contamination) were noted.

Additionally, melting curves for each reaction were examined, and the melting temperature (an indication of the product size) and number of peaks were determined. Where contamination was present, elimination measures were carried out. These included changing reagents, equipment and the location of pre-PCR preparation.

3.2 Analysis of the Smad Signalling Pathway

To determine whether aberrant expression or phosphorylation of components of the Smad signalling cascade could account for the fibrotic phenotype, the basal and TGF-β1-inducible levels of relevant Smads (Smad2/3 and -7, as well as phosphorylated Smad3) in skin fibroblasts were compared in affected and unaffected family members and appropriate unrelated controls, by means of western blotting and qPCR.

3.2.1 Cell Culture

As previously described, primary human fibroblast cultures were established from punch biopsies. These were taken either from unaffected, unrelated anonymised controls (in the case of optimisation experiments), or from family members (RTSAD1.3 and RTSAD1.4) and corresponding age- and sex-

matched controls (Control 2 and Control 1, respectively), after informed consent had been obtained. Cells were cultured in DMEM (SIGMA-ALDRICH or GIBCO®), supplemented with 1000mg/L D-Glucose, 4mM L-Glutamine, 110mg/L Sodium Pyruvate, and 10% Foetal Calf Serum (SIGMA-ALDRICH or GIBCO®) at 37°C, in a humidified atmosphere containing 10% CO_2.

3.2.2 Protein Extraction

In all cases, fibroblasts were cultured to passage 4-8 in culture flasks (Corning Incorporated, Corning, NY, USA), and washed with 1X sterile-filtered PBS (Bioshop Canada Inc.), before treatment with trypsin-EDTA (SIGMA-ALDRICH) for 2 minutes at 37°C. Cells were then seeded into 35mm culture dishes or 6-well culture plates at an approximate concentration of 100 000 cells per dish (or well), in DMEM, and grown to 85% confluence. After incubation at 37°C for 30-72hrs, suitably confluent cultures were washed twice with sterile filtered 1XPBS (Bioshop Canada Inc.), and incubated overnight at 37°C in serum-free DMEM (GIBCO®). Dishes were divided into "treated" and "untreated", with treated cells receiving 800pM recombinant human TGF-β1 (R&D Systems, Minneapolis, MN, USA) for 1hr at 37°C (Dong et al. 2002).

Following treatment with TGF-β1, culture medium was aspirated and cells were incubated with 100µl 1X Lysis Buffer (Appendix E) for 10 minutes on ice, with occasional scraping (cells intended for use in phospho-protein detection were flash-frozen on liquid nitrogen, prior to the addition of lysis buffer). The cell lysate was then transferred to a microcentrifuge tube, and centrifuged at 4°C in a Sigma 302K benchtop centrifuge (Sigma, Germany) for 10 minutes at 12 000rpm. The supernatant was transferred to a clean microcentrifuge tube, and mixed with 15-20µl of 5X Laemmli sample buffer containing 2-mercaptoethanol (see Appendix E), to give a final concentration of 1X. This mixture was boiled at 100°C for 5 minutes and centrifuged briefly in a Progen 24D microcentrifuge (Progen, Mexborough, UK). Samples were stored at -20°C (for use within 2 weeks) or -80°C (for long term storage).

3.2.3 Antibodies

Rabbit anti-Smad2/3 and Rabbit anti-phosphoSmad3 primary antibodies were obtained from Cell Signalling Technology (Danvers, MA, USA). Goat anti-Actin (loading control) primary antibody was obtained from Santa Cruz Biotechnology Inc. (Santa Cruz, CA, USA). Goat anti-Rabbit HRP-conjugated secondary antibody was obtained from Santa Cruz Biotechnology, while Mouse anti-Goat HRP-conjugated secondary antibody was obtained from Pierce Biotechnology Inc (Rockford, IL, USA).

3.2.4 Optimisation of Immunoblotting ("Western Blotting")

The first part of the western blot technique involves the separation of denatured proteins extracted from cultured cells, on a Polyacrylamide Gel matrix, incorporating Sodium Dodecyl Sulphate (SDS) as a denaturing agent. This is referred to as SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE). The Laemmli buffer system, a variation of SDS-PAGE, was used for all the western blot experiments described below. This is a discontinuous buffer system (different buffers are used in the gel and electrode reservoirs), and utilises two different percentage acrylamide gels – the stacking (4%) and resolving (7.5%) gels – to ensure that samples are compressed into a thin starting band and that individual proteins are finely separated and resolved. During protein extraction (see 3.2.2), proteins are heated in Laemmli buffer (Appendix E), containing SDS and the reducing agent 2-mercaptoethanol. As a result, the denatured polypeptides assume a rod-like shape, with a uniform mass:charge ratio, proportional to their molecular weights. Thus, when proteins are electrophoresed, they are separated and migrate according to their molecular weights, allowing accurate identification of proteins based on size (Mini-Protean ® 3 Cell Instruction Manual, BIO-RAD).

Proteins are then electrophoretically transferred from the gel and immobilised onto a nitrocellulose or polyvinylene difluoride (PVDF) membrane to which the proteins adhere noncovalently by adsorption. To maximise the success of protein transfer, the buffer system contains a conductive, buffering agent (Tris) at an alkaline pH, to maintain the conductivity and pH of the system; methanol, to promote the binding of proteins to the membrane; and SDS, to

promote the elution of proteins from the gel (Protein Blotting Guide: A Guide to Transfer and Detection, Third Edition, BIO-RAD).

The success of electrophoresis and transfer can be ascertained using a non-specific anionic dye to visualise all the proteins present on the membrane after transfer or by using pre-stained molecular weight markers to monitor transfer.

The specific protein of interest is detected using a four-step process. First, unoccupied binding sites on the membrane are blocked to reduce background signal, which could result from non-specific binding of either the primary or secondary antibody to the membrane. Next, the membrane is incubated with a primary antibody specific to the protein of interest. A species-specific, horseradish-peroxidase (HRP) conjugated secondary antibody is then added, which recognises and binds to the primary antibody. The addition of a suitable enzyme substrate, luminol, which is cleaved by HRP, results in the production of a chemiluminescent product through oxidation. The resultant light signal is captured by exposure to X-ray film or chemi-imager, allowing for specific and sensitive detection of the position of the protein of interest. The molecular weight of this protein can then be estimated, by comparison of its electrophoretic mobility to that of various size standards (Protein Blotting Guide: A Guide to Transfer and Detection, Third Edition, BIO-RAD).

Several of these steps presented targets for optimisation. Depending on the cross-reactivity of different antibodies, blocking solutions were chosen which minimised background signal. The type of membrane was also altered, from PVDF to nitrocellulose, to improve the efficiency of transfer and to reduce non-specific antibody binding.

Additionally, the dilution of each primary and secondary antibody (made using the appropriate diluent) was optimised, to ensure specificity and to reduce background signal. The optimal antibody concentration was considered to be the greatest dilution of antibody that yielded a positive signal, with minimal

background or non-specific reactions (Protein Blotting Guide: A Guide to Transfer and Detection, Third Edition, BIO-RAD).

Incubation time of the membrane with each primary antibody varied, with higher affinity antibodies requiring shorter periods of incubation. Finally, the optimal exposure time to either film or chemi-imager was ascertained, in order to accurately determine and compare the levels of protein present in each sample.

In general, however, the following method was used, and is referred to hereafter as the "standard western blot protocol": Twenty microlitre aliquots of the samples extracted from fibroblasts as described in 3.2.2 were electrophoresed on a 7.5% Resolving/4% Stacking Polyacrylamide Gel containing 0.1% SDS (Appendix E), at a constant voltage of 150V (maximum current 60mA) for 1 hour (or 70 minutes, if two gels were electrophoresed simultaneously) in 1X Running Buffer (Appendix E) using the Mini Protean II gel apparatus (BIO-RAD). The Precision Plus Protein Standard Dual Colour (BIO-RAD) was included for molecular weight determination (Appendix B).

Following electrophoresis, proteins were transferred to a nitrocellulose membrane (Hybond ECL, Amersham plc, Buckinghamshire, UK), at a constant voltage of 100V (maximum current 250mA) for 2 hours, in 1X Transfer Buffer (Appendix E). The membrane was washed once for 5 minutes in TBS containing 0.1% Tween20 (TBST) (Appendix E). Proteins were detected after transfer by staining briefly with Ponceau S (0.1% (w/v) Ponceau S in 5% acetic acid), and where necessary (as in the case where two or more antibodies were being optimised at one time), membranes were cut into sections using a scalpel.

After destaining with deionised water, membranes (or membrane sections) were blocked with 5% Marvel fat-free milk powder (Premier International Foods UK, Ltd.) in TBST for 1 hour at room temperature. Membranes were then incubated with primary antibody (at a dilution determined by optimisation) either overnight at 4°C or for 1 hour at room temperature.

This was followed by a brief rinse in TBST and several washes with TBST (as recommended by the antibody manufacturer), to remove excess primary antibody. Membranes were then incubated with secondary antibody (diluted in blocking solution) for 1 hour at room temperature.

Following incubation with secondary antibody, membranes were rinsed briefly in TBST and then subjected again to several washes with TBST, to remove excess antibody.

Finally, membranes were rinsed three times in deionised water and incubated in LumiGLO Reserve chemiluminescent substrate (KPL Inc., Gaithersburg, MA, USA), consisting of two volumes of Solution B, combined with one volume of Solution A for 1 minute. Excess substrate solution was drained from the membrane, which was placed between two sheets of transparency film in an X-ray cassette. Visualisation was performed by exposing the membrane to Agfa film (Agfa-Gevaert N.V., Mortsel, Belgium), for various time periods. Chemiluminescent detection was also performed using the ChemiImager 550 (Alpha Innotech Corporation, USA). Band density values were obtained using the Fluorchem 550 software (Alpha Innotech Corporation), following manufacturer's instructions (Alpha Innotech Corporation).

Membranes to be stripped and reprobed with a different antibody were kept moist after detection by vacuum sealing in a plastic bag or by placing the membrane in TBST in a sealed container and were stored at 4°C until reuse. Stripping and reprobing was performed within five days of first use and followed the "standard stripping and reprobing protocol", as follows: membranes were incubated with stripping buffer (62mM Tris-HCl pH6.8, 2% SDS, 100mM 2-mercaptoethanol, see Appendix E) for 15 minutes at 50°C. Stripped membranes were then washed twice with 200ml deionised water and the standard western blot protocol was followed from the blocking step. Individual membranes were stripped and reprobed up to three times with different antibodies.

3.2.5 anti-Smad2/3 Western Blot

Twenty microlitre aliquots of total protein isolated from treated or untreated RTSAD1.3, RTSAD1.4 and corresponding age- and sex-matched control fibroblasts (see 3.2.2) were subjected to SDS-PAGE, and the standard western blot protocol described in 3.2.4 was followed, with the following alterations: after blocking, with 5% (w/v) fat-free milk powder in TBST, membranes were rinsed with TBST, and washed 3 times, for 5 minutes each, with TBST, to remove excess blocking reagent. Membranes were then incubated overnight with primary antibody (Rabbit anti-Smad2/3, Cell Signalling Technologies) at a dilution of 1:1000 in 5% (w/v) globulin-free Bovine Serum Albumin (BSA) (SIGMA-ALDRICH) in TBST. Membranes were again rinsed, and washed 3 times for 5 minutes each with TBST. Following incubation with secondary antibody (Goat anti-Rabbit, Santa Cruz Biotechnology) at a dilution of 1:2000, this washing process was repeated.

3.2.6 anti-phosphoSmad3 Western Blot

As for Smad2/3, phosphorylated Smad3 was detected in total protein isolated from treated and untreated fibroblasts by western blot using the standard western blot protocol, with the following alterations: after blocking, with 5% (w/v) fat-free milk powder in TBST, membranes were rinsed with TBST, and washed 4 times, for 5 minutes each, with TBST. Membranes were then incubated overnight with primary antibody (Rabbit anti-phosphoSmad3, Cell Signalling Technologies) at a dilution of 1:1000, in 5% (w/v) BSA in TBST. Membranes were again rinsed, and washed 4 times for 5 minutes each with TBST. Following incubation with secondary antibody (Goat anti-Rabbit, Santa Cruz Biotechnology) at a dilution of 1:1000, this washing process was repeated.

3.2.7 anti-Actin Western Blot

Membranes which had previously been probed for one of the Smad proteins were stripped, as described in the standard stripping and reprobing protocol, and probed with anti-Actin primary antibody, as a loading control. The standard western blot protocol was then followed from the blocking step, with the following alterations: after blocking, membranes were incubated with

primary antibody (Goat anti-Actin, Santa Cruz Biotechnology), at a dilution of 1:1000 in 5% (w/v) fat-free milk powder in TBST, for one hour. Membranes were then rinsed in TBST, and washed four times, for 15 minutes each, with TBST. After incubation with secondary antibody (Mouse anti-Goat, Pierce) at a dilution of 1:2000, this washing procedure was repeated, prior to chemiluminescent detection.

3.2.8 Statistical Analysis

For a single individual, comparisons between basal and TGF-β1-inducible levels of either Smad2/3 expression or Smad3 phosphorylation were analysed using the Student's *t*-test for paired samples. A two-tailed distribution was used, assuming equal variances.

Statistical analysis of the difference in treated:untreated ratios of expression or phosphorylation between individuals was performed by means of the Student's *t*-test for unpaired samples. A two-tailed distribution was used, assuming unequal variances.

In both cases, significance was defined using a type I error, or *p*-value, of 0.05. All tests were performed on Microsoft Excel (Copyright 2003, Microsoft Corporation).

3.2.9 Determination of SMAD7 Expression Levels by qPCR

Basal and TGF-β1-inducible levels of *SMAD7* mRNA, in RTSAD1.3 and RTSAD1.4 cultured fibroblasts, were determined using qPCR, and the results analysed using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

3.2.9.1 Determination of Amplification Efficiencies for Target (SMAD7) and Reference (GUSB) genes

Amplification efficiencies of the target (*SMAD7*) and reference (*GUSB*) genes were compared as described in 3.1.5.1, using the Master Mix described in Table 3, and the standard qPCR reaction conditions (Table 4). Dilutions of control cDNA (100ng/µl, 50ng/µl and 10ng/µl), previously isolated from cultured skin fibroblasts taken from an unaffected, unrelated individual

(supplied by J. Meyer), were used to construct the standard curve. Both target and reference genes were amplified from each cDNA dilution in triplicate, by means of Quantitect Primer Assays (QIAGEN).

3.2.9.2 Basal SMAD7 Expression in Family Samples

In order to determine basal mRNA expression of *SMAD7* in skin fibroblasts, qPCR amplification of the target and reference genes was performed, using RTSAD1.3 and RTSAD1.4 cDNA, previously synthesised from RNA extracted from fibroblasts which had not been stimulated with exogenous TGF-β1 (see 3.1.2 and 3.1.3). The qPCR master mix described in Table 4 was used, and capillaries were prepared as described in 3.1.5.1. Two microlitres of template cDNA, at a concentration of 100ng/µl, was added to each capillary, except in the case of the NTCs, in which 2µl of water was substituted for cDNA. Reactions were performed in triplicate. Amplification was performed in triplicate for each individual, for each gene, using the standard qPCR reaction conditions (Table 5), on the LightCycler 1.5 (Roche), and results were analysed using LightCycler Software v3.5.

3.2.9.3 Cell Culture, RNA Isolation and cDNA Synthesis

Skin fibroblasts, taken from biopsies of an affected (RTSAD1.3) and an unaffected individual (RTSAD1.4) were cultured as previously described (3.2.1). Cells were cultured to passage 7-9, washed with 1X PBS (Bioshop Canada Inc.), and treated with trypsin-EDTA for 2 minutes at 37°C (SIGMA-ALDRICH), before being seeded into 35mm cell culture dishes. After incubation at 37°C for 48hrs, dishes which had reached 85% confluence were washed twice with 1X PBS, and incubated overnight at 37°C in serum-free DMEM (GIBCO®). Dishes were divided into “treated” and “untreated”, with treated cells receiving 500pM recombinant human TGF-β1 (R&D Systems) for 1hr at 37°C (Dong et al. 2002). Following treatment, medium was aspirated, cells were washed with 1X sterile-filtered PBS (Bioshop Canada Inc.), before treatment with trypsin-EDTA (SIGMA-ALDRICH). This was followed by a second wash, in 12ml DMEM (SIGMA-ALDRICH), and centrifugation for 10 minutes at 1000rpm in an MSE bench-top centrifuge (Henderson Biomedical Ltd.). The cell pellet was resuspended in 10ml sterile filtered 1x PBS

(Bioshop Canada Inc.), and centrifuged for a further 10 minutes at 1000rpm. The pellet was again resuspended, in 1ml PBS, and centrifuged for 10 minutes at 13 400rpm in a microcentrifuge. After aspiration of the supernatant, total RNA was isolated from the cell pellet using the High Pure RNA Isolation Kit (Roche Diagnostics), following manufacturer's instructions. Isolated RNA was stored at -80°C in 10µl aliquots.

The concentration of RNA samples was determined using the NanoDrop Spectrophotometer (NanoDrop Technologies) and the integrity of the RNA was ascertained by agarose gel electrophoresis. Ten microlitres of each sample, mixed with 3µl of loading buffer, was electrophoresed on a 1% agarose gel in TBE buffer for 20min at 120V, incorporating Ethidium Bromide (SIGMA-ALDRICH) (0.5µg/ml) for visualisation. To determine the molecular mass of the bands observed, 500ng of the Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas International Inc.) was included as a size standard. The gel was visualised under UV light as previously described (see 3.1.2).

Synthesis of cDNA was performed using the Transcriptor First Strand cDNA Synthesis Kit (Roche), following manufacturer's instructions, and using oligo$(dT)_{18}$ primers, complementary to the mRNA poly(A) tail. Denaturation of the template-primer mix was performed on a Touchdown Thermal Cycler (ThermoHybaid, Middlesex, UK), while the reverse transcription and inactivation steps were performed using a Px2 Thermal Cycler (ThermoHybaid). The resulting cDNA was stored at -20°C.

Newly-synthesised cDNA was checked for possible genomic contamination PCR, using primers designed to amplify exon 3 of *TGFB1*, as previously described (3.1.3). The standard reaction mix (Table 1) was used, and reaction conditions were as follows: An initial denaturation step at 95°C for 5 minutes, followed by thirty cycles of 95°C for 30 seconds, 58°C for 30 seconds, and 72°C for 30 seconds, with a final 7 minute elongation step at 72°C. The reaction was performed on a Px2 Thermal Cycler (ThermoHybaid).

Following the addition of 5µl loading dye, PCR products were electrophoresed on a 1% agarose gel in TBE buffer, incorporating 0.5µg/ml Ethidium Bromide (SIGMA-ALDRICH), for 20 minutes at 120V, and visualised under UV light (3.1.2). As before, 500ng of the Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas) was included as a size standard.

3.2.9.4 TGF-β1 Inducible SMAD7 Expression in Family Samples

The target (*SMAD7*) and reference (*GUSB*) genes were amplified from cDNA, prepared from RNA isolated from treated or untreated RTSAD1.3 and RTSAD1.4 fibroblasts (see 3.2.9.3). The qPCR master mix described in Table 3 was used, and capillaries were prepared as described in 3.1.5.1. To each capillary, 2µl of undiluted template cDNA was added, at a concentration of approximately 1900ng/µl, as determined by the NanoDrop Spectrophotometer (NanoDrop Technologies). In the case of the NTCs, 2µl of water was substituted for cDNA. Amplification was performed in duplicate for both treated and untreated samples, for each individual, on the LightCycler 1.5, using the standard qPCR reaction conditions. Results were analysed using LightCycler Software v3.5.

3.2.9.5 Analysis using the $2^{-\Delta\Delta CT}$ Method

Analysis of the results of qPCR performed on the family samples was performed using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001), as described in 3.1.5.3, with Equation [1] modified as follows:

To investigate differences in basal levels of expression:

$$\Delta\Delta C_T = (C_{T,\,Smad7} - C_{T,\,GUSB})_{individual\ X} - (C_{T,\,Smad7} - C_{T,\,GUSB})_{RTSAD1.4}, \quad [1a]$$

where individual X may be either RTSAD1.3 or RTSAD1.4, and, to investigate fold change in *SMAD7* expression in response to treatment:

$$\Delta\Delta C_T = (C_{T,\,SMAD7} - C_{T,\,GUSB})_{treated} - (C_{T,\,SMAD7} - C_{T,\,GUSB})_{untreated}, \quad [1b]$$

for either RTSAD1.3 or RTSAD1.4.

Thus, for each individual, basal *SMAD7* expression was reported relative to RTSAD1.4, while the fold change in expression in response to TGF-β1 treatment was reported relative to *SMAD7* expression in the untreated sample for that individual.

3.2.9.6 Statistical Analysis

Statistical analysis of the fold difference in basal or TGF-β1-inducible *SMAD7* expression between the individuals was performed by means of the Student's *t*-test for unpaired samples. A two-tailed distribution was used, assuming unequal variances.

The difference between basal and TGF-β1 inducible levels of *SMAD7* levels for a single individual was analysed using the Student's *t*-test for paired samples. A two-tailed distribution was used, assuming equal variances.

In both cases, significance was defined using a type I error, or *p*-value, of 0.05. All tests were performed on Microsoft Excel (Copyright 2003, Microsoft Corporation).

3.2.9.7 Determining Specificity of qPCR

As previously described (see 3.1.5.5), specificity of the qPCR reactions was determined by agarose gel electrophoresis and/or analysis of the melting peak produced for each reaction on the LightCycler (Roche).

4. Results

4.1 Analysis of *TGFB1* mRNA

The presence of possible disease-causing mutations in *TGFB1*, such as splice-site aberrations, whole exon deletions, or expression defects, was investigated by examining the length and sequence of the transcript, and determining expression levels of *TGFB1* in skin fibroblasts. This involved extraction of total RNA from cultured fibroblasts taken from an affected (RTSAD1.3) and an unaffected individual (RTSAD1.4), reverse transcription, amplification of the resulting cDNA by conventional PCR and real-time quantitative PCR, and analysis of the transcript by means of agarose gel electrophoresis and direct sequencing.

4.1.1 Isolation of Total RNA from Fibroblasts

Total RNA was isolated from RTSAD1.3 and RTSAD1.4 cultured skin fibroblasts at passage seven. To determine the quality and quantity of the RNA, aliquots were subjected to agarose gel electrophoresis following isolation.

Intact eukaryotic RNA exhibits bright, sharp bands corresponding to the 28S and 18S ribosomal RNAs (rRNAs), with the 28S band approximately twice as bright as the 18S band, when visualised under UV light. This 2:1 intensity ratio is a good indication of the integrity of the sample, with degraded RNA appearing as a smear, lacking both the bright 28S and 18S bands, and the 2:1 ratio.

Since the rRNA bands could clearly be seen in both samples (RTSAD1.3 and RTSAD1.4), and since no smearing was present, the RNA was deemed to be intact (Fig.5). The fluorescent intensity of the bands under UV light indicated that a sufficient amount of RNA had been isolated to proceed with further experiments.

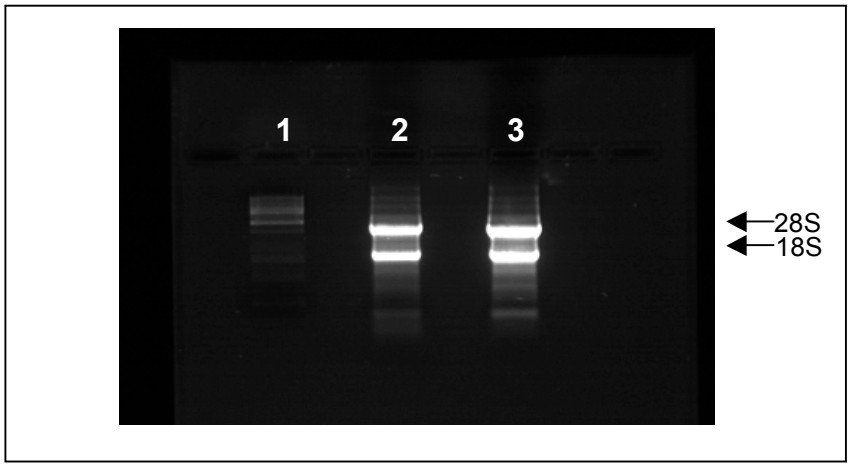


Figure 5. Total RNA extracted from RTSAD1.3 and RTSAD1.4 skin fibroblasts, electrophoresed on a 1% agarose (0.5µg/ml Ethidium Bromide) gel to determine integrity. The positions of the 28S and 18S rRNA bands are indicated with arrows. 1: Gene Ruler^{TM} 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2: RTSAD1.3 RNA, 3: RTSAD1.4 RNA.

4.1.2 cDNA Synthesis and Checking for Genomic DNA Contamination

Following isolation, RTSAD1.3 and RTSAD1.4 total RNA was reverse transcribed, and the integrity of the resulting cDNA was again checked by agarose gel electrophoresis. Because of the range of different sized fragments generated by reverse transcriptase, cDNA appears as a smear on the gel. cDNA synthesised from intact total RNA contains high-, medium- and low-molecular weight fragments, generating a longer smear than cDNA synthesised from degraded RNA, which consists mainly of low molecular weight fragments.

Results such as those shown in Figure 6 suggest sub-optimal template RNA integrity, with the breadth of the smear indicating the presence of predominantly low molecular weight fragments. However, since the average lengths of the cDNA fragments (determined by comparison with the molecular weight marker) exceeded the length of the fragment to be amplified by PCR, it was decided that these samples could be used for downstream applications.

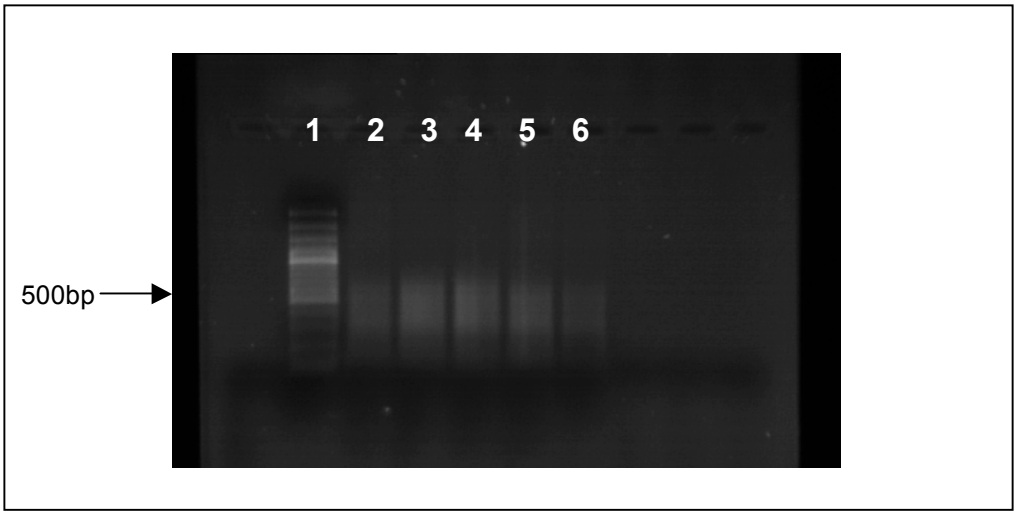


Figure 6. cDNA synthesised from isolated fibroblast RNA by reverse transcription, and subjected to agarose gel electrophoresis (1% agarose, 0.5µg/ml ethidium bromide) to determine integrity. The size of a reference band in the molecular weight marker is shown. 1: Gene Ruler^{TM} 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2 to 5: cDNA synthesised from RTSAD1.3 total RNA.

To determine whether genomic DNA was present in the cDNA samples, after isolation and reverse transcription, PCR amplification was performed on a sample of RTSAD1.4 cDNA. This PCR, which had previously been optimised on genomic DNA, employed primers complementary to intronic regions flanking exon 3 of *TGFB1* (Hartmann 2006). Since cDNA contains no introns, it was anticipated that amplification would only occur in the presence of contaminating genomic DNA (Fig 7A). A sample of RTSAD1.4 genomic DNA was included as a positive control.

As expected, amplification of exon 3 took place when genomic DNA was used as a template. In the cDNA sample, however, very weak amplification was observed (Fig. 7B). This indicated that a small amount of genomic DNA was present. However, since the primers used for sequencing and qPCR were designed to span exon/exon boundaries, making them cDNA-specific, it was concluded that the presence of genomic DNA would not affect the results of these experiments (Fig. 7C).

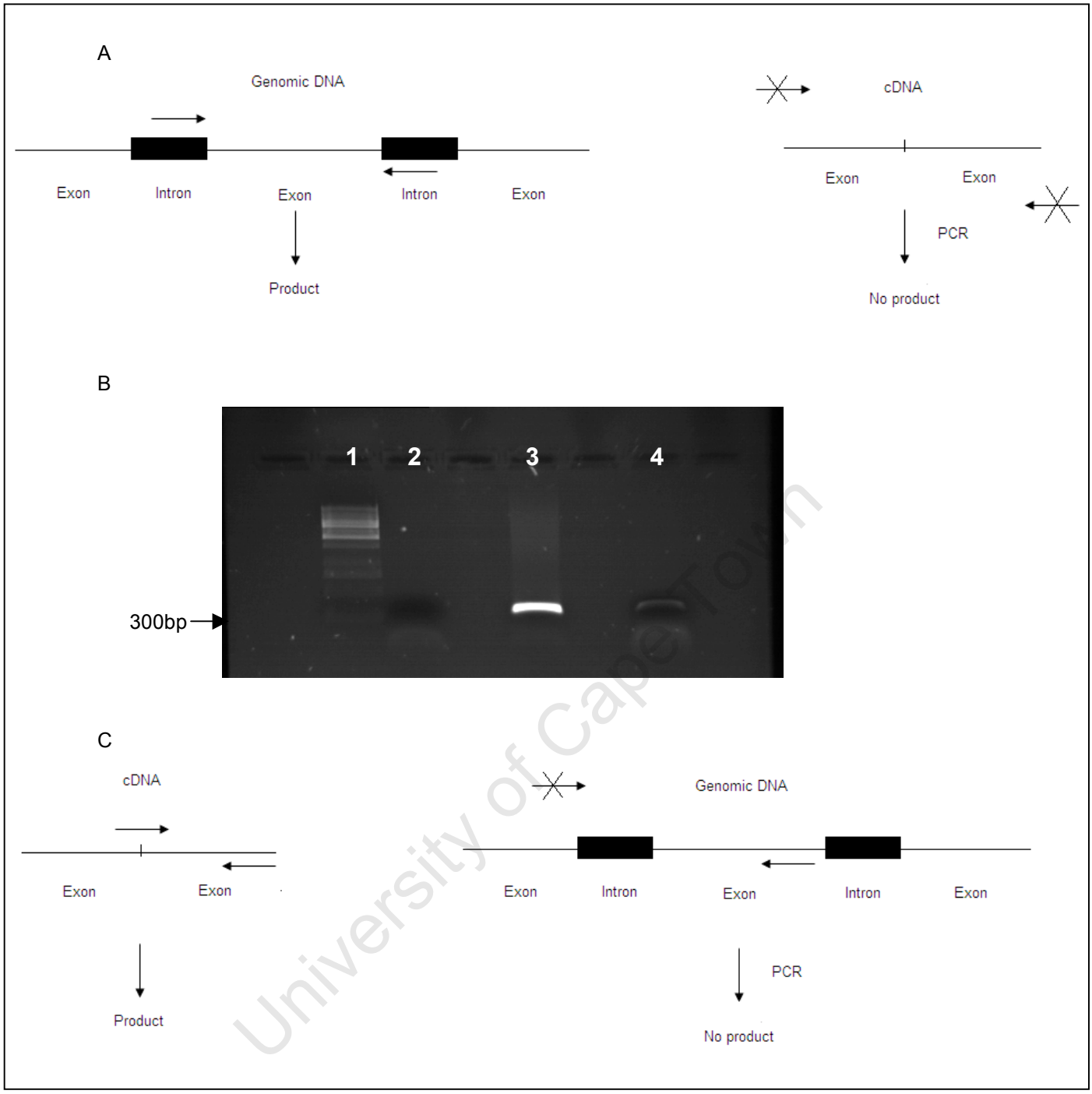


Figure 7. PCR using intronic primers to determine the presence of genomic DNA contamination. **A** Position of primers relative to the genomic and cDNA sequences. Intronic primers will only generate a product in the presence of genomic DNA, since cDNA lacks the required intronic binding sites. **B** Products of PCR using intronic primers, designed to amplify exon 3 of *TGFB1*. 1: Gene Ruler^{TM} 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2: DNA-free water control, 3: RTSAD1.4 genomic DNA positive control, 4: RTSAD1.4 cDNA. The size of a reference band in the molecular weight marker is shown. **C** Design of cDNA-specific primers, across exon/exon boundaries, to prevent erroneous amplification from contaminating genomic DNA.

4.1.3 Determination of TGFB1 Transcript Length and Sequence

The presence of splice site mutations in the *TGFB1* transcript, which could lead to whole exons being skipped or deleted, was examined by agarose gel electrophoresis and direct sequencing of cDNA from the affected and unaffected family members.

4.1.3.1 Primer Design and PCR Optimisation

Primers were designed as described in Materials and Methods, to amplify either the whole transcript (all seven exons), or an internal fragment, corresponding to exons 1 to 6. Optimisation was performed for both of the transcript amplicons, using a temperature gradient, and results were visualised by means of agarose gel electrophoresis (Fig. 8).

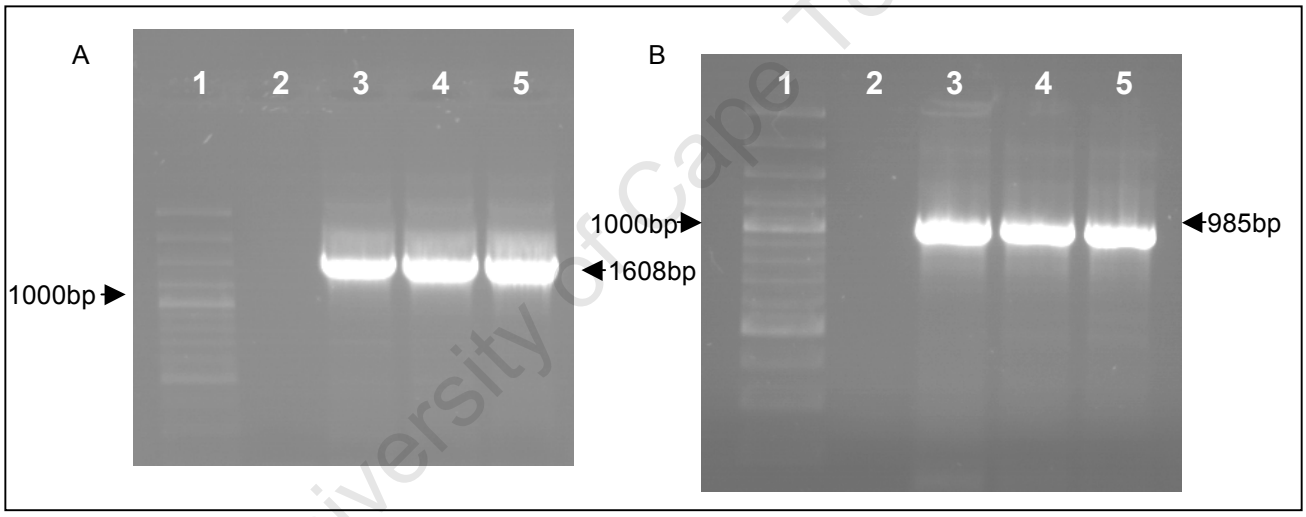


Figure 8. Results of temperature gradient PCR, following electrophoresis on 1% agarose (0.5µg/ml Ethidium Bromide) gel. Observed fragment sizes are indicated in each case by an arrow. The size of a reference band in the molecular weight marker is shown. **A** Whole Transcript amplicon, **B** Internal Fragment amplicon. 1: Gene Ruler^{TM} 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2: DNA-free water Control, 3 to 5: PCR products generated using annealing temperatures of 57.3°C, 59.4°C and 61.6°C respectively.

A bright band, corresponding to the expected size of the amplified fragment, was observed in each lane after PCR and electrophoresis. It was anticipated that the relative fluorescent intensity of the bands in different lanes, when viewed under UV light, would vary according to the efficiency of the reaction at different annealing temperatures. The annealing temperature yielding the brightest band would then be chosen as the optimum. However, in both

cases, the high intensity of the bands generated at all three of the annealing temperatures of the gradient suggested that the PCRs were extremely efficient (Fig. 8). Thus, the annealing temperature which resulted in the least non-specific amplification (i.e. the least smearing) was chosen as the optimum, in each case. For the "whole transcript" amplicon, the optimum annealing temperature was found to be 57.3°C (rounded to 57°C), while the "internal fragment" was optimally amplified at a T_a of 61.6°C (rounded to 61°C).

4.1.3.2 PCR in Family Samples, Determination of Amplicon Length and Preparation for Sequencing

Following the selection of optimum annealing temperatures for the amplification of both fragments, PCR was performed on RTSAD1.3 and RTSAD1.4 cDNA. The products of the PCRs were subjected to agarose gel electrophoresis (Fig. 9), and the length of each amplicon was compared between individuals, and to the predicted length based on the NCBI mRNA reference sequence.

Because of the efficiency of the PCR, yielding a vast amount of PCR product, the bands observed after electrophoresis were fairly diffuse, making size determination difficult. However, no apparent difference in size was detectable for either amplicon, when the two individuals were compared (Fig. 9). In the case of the whole transcript amplicon, bands of approximately 1.5kb were observed for both RTSAD1.3 and RTSAD1.4 (compared to the predicted 1608bp based on the NCBI sequence), while for the internal fragment, comparison with the molecular weight marker indicated that the PCR product obtained from both individuals was approximately 1kb in length (compared with the NCBI predicted sequence length of 985bp).

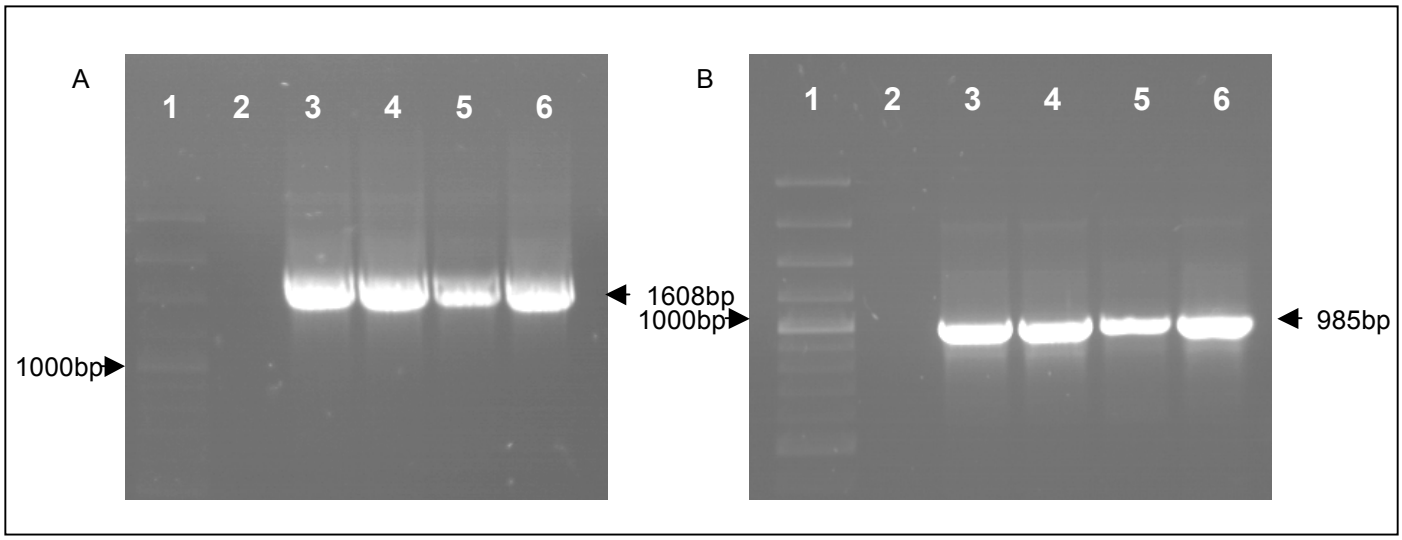


Figure 9. Results of PCR in family cDNA samples, following electrophoresis on a 1% agarose (0.5µg/ml Ethidium Bromide) gel, to determine amplicon length. Observed fragment sizes are indicated by arrows. The size of a reference band in the molecular weight marker is shown. **A** Whole Transcript amplicon. **B** Internal Fragment amplicon. 1: Gene Ruler^{TM} 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2: DNA-free water control, 3 and 4: RTSAD1.3, 5 and 6: RTSAD1.4. PCR product lengths were determined for each individual, based on their electrophoretic mobility relative to the size standard.

These bands were excised and purified from the agarose gel, and the amount of PCR product present in each was quantified, by subjecting an aliquot of the purified product to agarose gel electrophoresis. Due to the loss of PCR product incurred during clean-up, the bands on the quantification gel were much sharper, allowing confirmation of the sizes of the amplicons in each individual (Fig. 10).

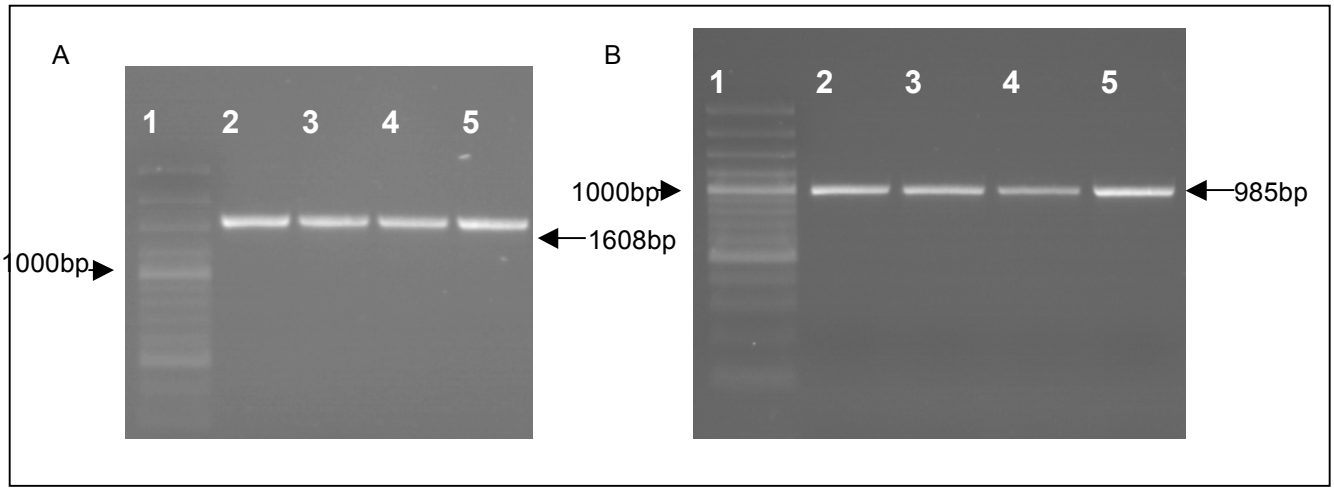


Figure 10. PCR products amplified from family cDNA samples, following gel extraction, and electrophoresis on a 1% agarose (0.5µg/ml Ethidium Bromide). Observed fragment sizes are indicated by arrows. The size of a reference band in the molecular weight marker is shown. **A** Whole Transcript amplicon. **B** Internal Fragment amplicon. 1: Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2 and 3: RTSAD1.3, 4 and 5: RTSAD1.4.

4.1.3.3 Sequencing

Sequences generated for each of the amplicons, using either the forward or reverse primers, were compared to the *TGFB1* mRNA reference sequence obtained from the NCBI database. Electropherograms were analysed for heterozygous changes and ambiguous sequence positions, which were replaced with the notation "N", and edited sequences were then aligned using ClustalW, available in BioEdit Sequence Alignment Editor. Alignments were generally created using two or more sequence reads per primer for both individuals, to ensure a greater degree of accuracy.

In the case of the "whole transcript" forward and reverse primers, sequencing quality declined rapidly after approximately 400bp, leaving a region of 748bp internal to the transcript, which could not be sequenced. Thus, the internal fragment primers were employed to sequence this region. These were designed to allow for an overlap between the whole transcript and internal fragment sequences, ensuring coverage of all the exon/exon boundaries of the transcript (see Appendix C).

Positions that differed from the reference sequence in all the forward and/or reverse reads for a particular individual were investigated further. The presence of these variants in the previously determined RTSAD1.3 and RTSAD1.4 genomic DNA sequences was ascertained and, where possible, information regarding their effect on the structure and function of *TGFB1* and its protein product was obtained.

Numerous novel single nucleotide changes were observed in single sequence reads of the whole transcript amplicon, for both individuals. However, when examining the alignment generated by several rounds of sequencing, these variants could not be confirmed. It is therefore likely that such apparent polymorphisms are in fact sequencing artefacts, or aberrant bases inserted during amplification of this particularly long sequence by a non-proofreading Taq polymerase. Excluding these artefacts, two previously published SNPs were identified, both of which had previously been found during the sequencing of genomic DNA from RTSAD1.3 and RTSAD1.4 (Hartmann 2006).

In the alignment of sequences generated using the whole transcript forward primer, a single heterozygous change (C→T) was observed in exon 1 of the transcript, in RTSAD1.3. This change, which was supported by an identical finding in the overlapping region of the internal fragment forward primer alignment, was identified as the c. +29T>C polymorphism (rs1982073). The unaffected control, RTSAD1.4, was found to be homozygous (C/C) at the equivalent position (Fig. 11.1).

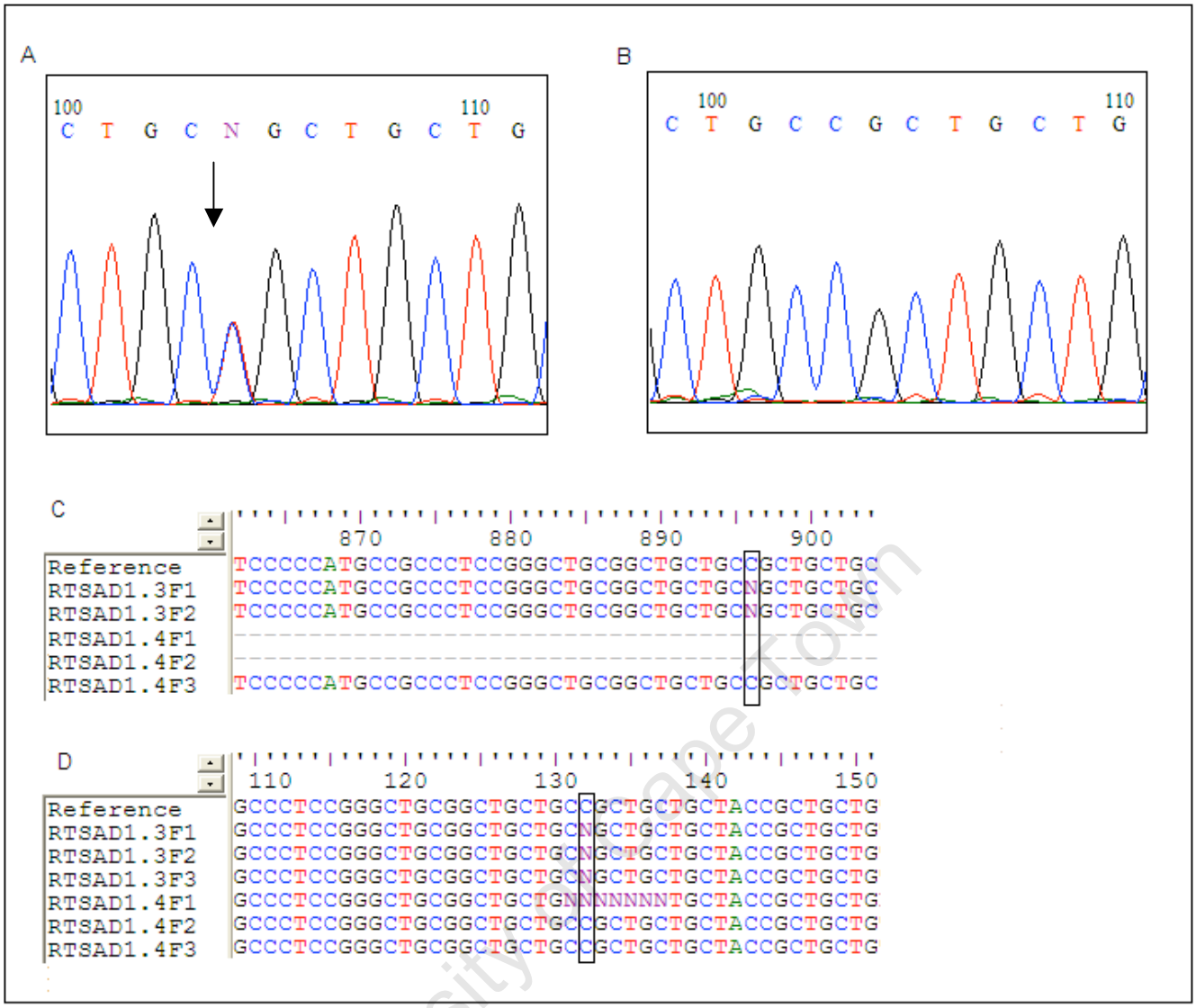


Figure 11.1 Results of sequencing using the whole transcript and internal fragment forward primers. **A** Electropherogram of RTSAD1.3, showing the heterozygous change identified as the c.+29T>C polymorphism (indicated by an arrow). **B** Electropherogram of corresponding sequence in RTSAD1.4. **C and D** BioEdit sequence alignment of sequences generated using the whole transcript, and internal fragment forward primers, respectively (Reference: NCBI reference sequence for *TGFB1* transcript, F: Forward primer, numbers indicate different sequencing runs). The position of the polymorphism is indicated by a block.

The second SNP, a G→C transversion, was identified in the heterozygous state in RTSAD1.4, in the alignment produced from sequences generated using the internal fragment reverse primer. This SNP, present in exon 1 of *TGFB1*, was determined to be the c. +74G>C polymorphism, described by Kim et al. (1989) (rs1800471). The sequence of RTSAD1.3 did not differ from

the reference sequence at this position, and was deemed homozygous (G/G) (Fig. 11.2).

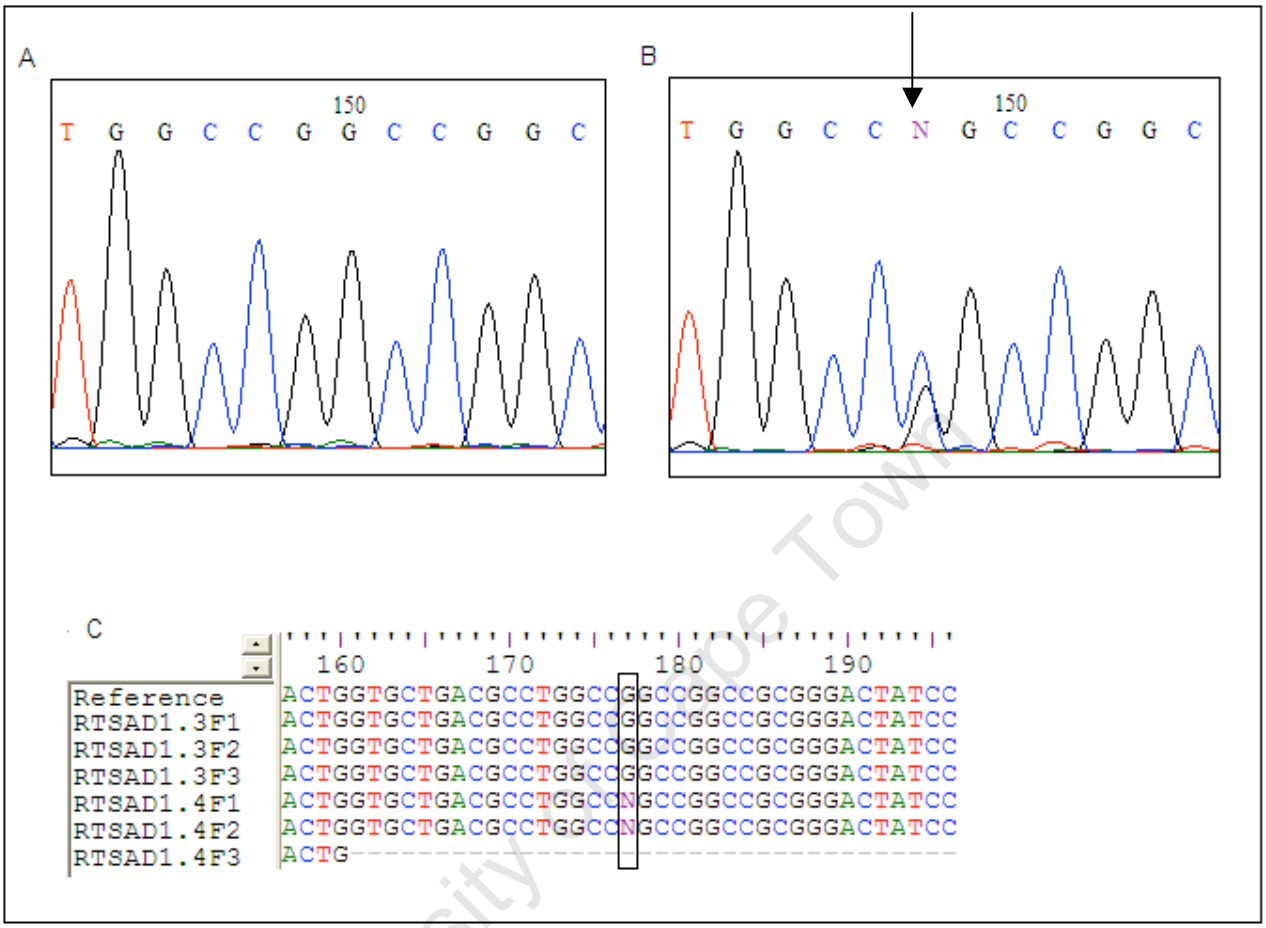


Figure 11.2 Results of sequencing using the internal fragment forward primer. **A** Electropherogram of RTSAD1.3. **B** Electropherogram of corresponding sequence in RTSAD1.4, showing the heterozygous change identified as the c.+74G>C SNP (indicated by an arrow). **C** BioEdit sequence alignment of sequences generated using the internal fragment forward primer (Reference: NCBI reference sequence for *TGFB1* transcript, F: Forward primer, numbers indicate different sequencing runs). The position of the polymorphism is indicated by a block.

Combining the sequencing results generated using the whole transcript and internal fragment primers, all the exon/exon junctions could be examined. No deviations from the reference sequence were observed in either individual, confirming the results of agarose gel electrophoresis (see 4.1.3.2), and indicating that splicing and post-transcriptional processing of the *TGFB1* mRNA had occurred normally.

4.1.4 Real-time (quantitative) PCR of the TGFB1 Transcript

Expression levels of the *TGFB1* transcript in fibroblasts taken from an affected and an unaffected individual were quantified by qPCR (using SYBR Green reaction chemistry), and the results were analysed using the $2^{-\Delta\Delta CT}$ method.

4.1.4.1 Determination of Amplification Efficiencies for Target (TGFB1) and Reference (GUSB) Genes

For the $2^{-\Delta\Delta CT}$ method to be valid, the amplification efficiencies of the target and reference gene PCRs must be approximately equal. To confirm whether this was indeed the case, the target (*TGFB1*) and reference (*GUSB*) genes were amplified from serially diluted control cDNA. For each cDNA dilution, the ΔC_T values ($C_{T,\ TGFB1} - C_{T,GUSB}$) were plotted against the logarithm of the concentration, and the slope of the linear regression line was determined, with an absolute value close to zero indicating approximately equal efficiencies. As shown in Figure 12, the gradient of the line was found to be 0.0417, indicating that the amplification efficiencies of *GUSB* and *TGFB1* were equal, and that the $2^{-\Delta\Delta CT}$ method could be used to analyse the data.

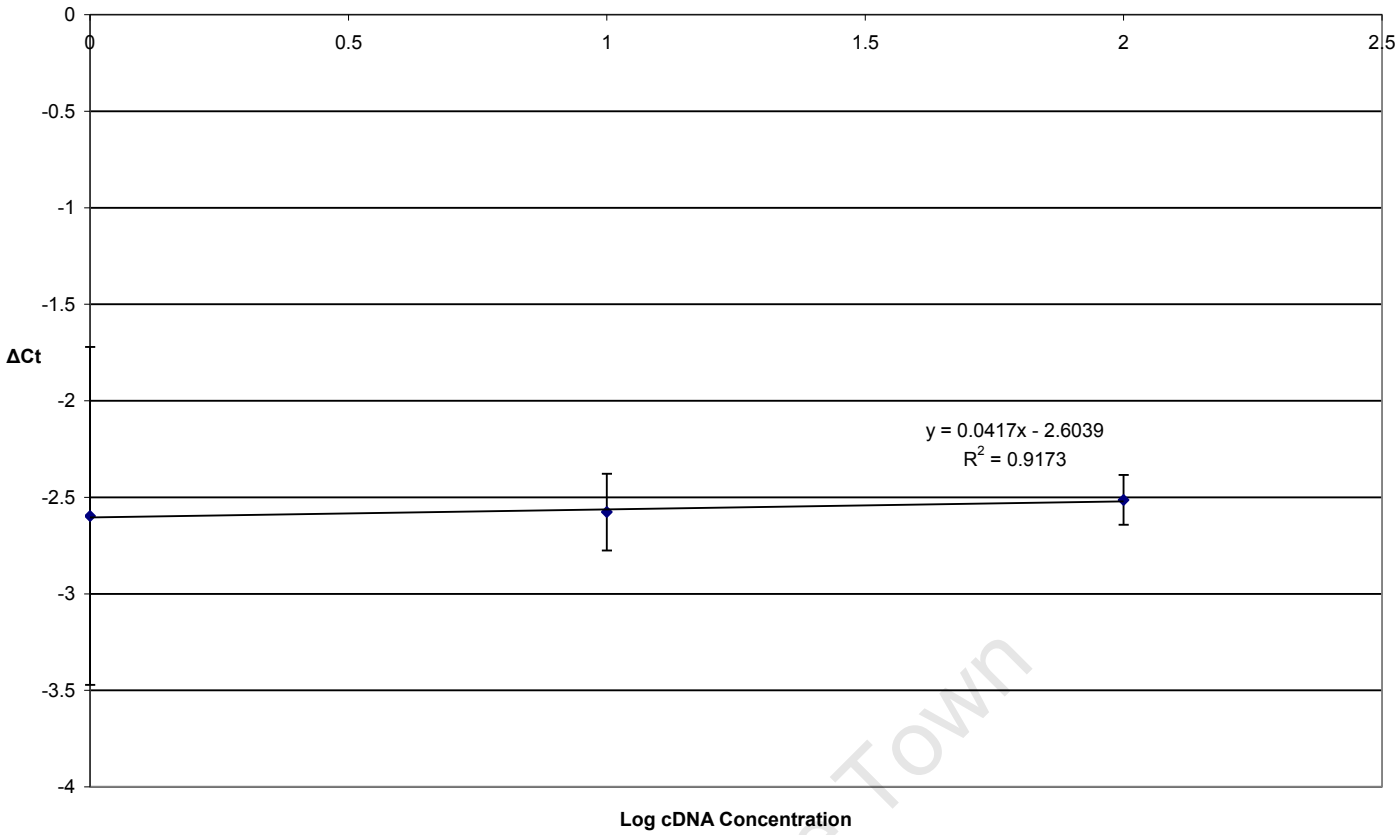


Figure 12. Validation of the $\Delta\Delta C_T$ method. The efficiency of amplification of the target (*TGFB1*) and reference (*GUSB*) genes was investigated using qPCR. Serially diluted control cDNA (100, 10 and 1ng/µl) was amplified by qPCR using gene-specific primers. The ΔC_T ($C_{T,\ TGFB1} - C_{T,GUSB}$) was calculated for each dilution, and plotted against the log of the concentration, and the linear regression line was determined using Microsoft Office Excel 2003. A gradient with absolute value approximately zero indicates comparable efficiencies.

4.1.4.2 qPCR from Family Samples and Analysis using the $2^{-\Delta\Delta CT}$ Method

Once the amplification efficiencies of the target and reference genes had been determined, *GUSB* and *TGFB1* were amplified from RTSAD1.3 and RTSAD1.4 cDNA in triplicate, and the crossing points and mean crossing points were calculated. Expression of *TGFB1* was normalised to *GUSB* for each individual, by calculating the ΔC_T ($C_{T,\ TGFB1} - C_{T,GUSB}$). The $\Delta\Delta C_T$ was then calculated using Eq. [1], and the fold change in *TGFB1* expression, relative to the unaffected individual, was determined using Eq. [2] (see 3.1.5.3).

Based on these calculations, no significant difference in normalised *TGFB1* expression was detected between RTSAD1.3 and RTSAD1.4 (p=0.43) (Fig. 13).

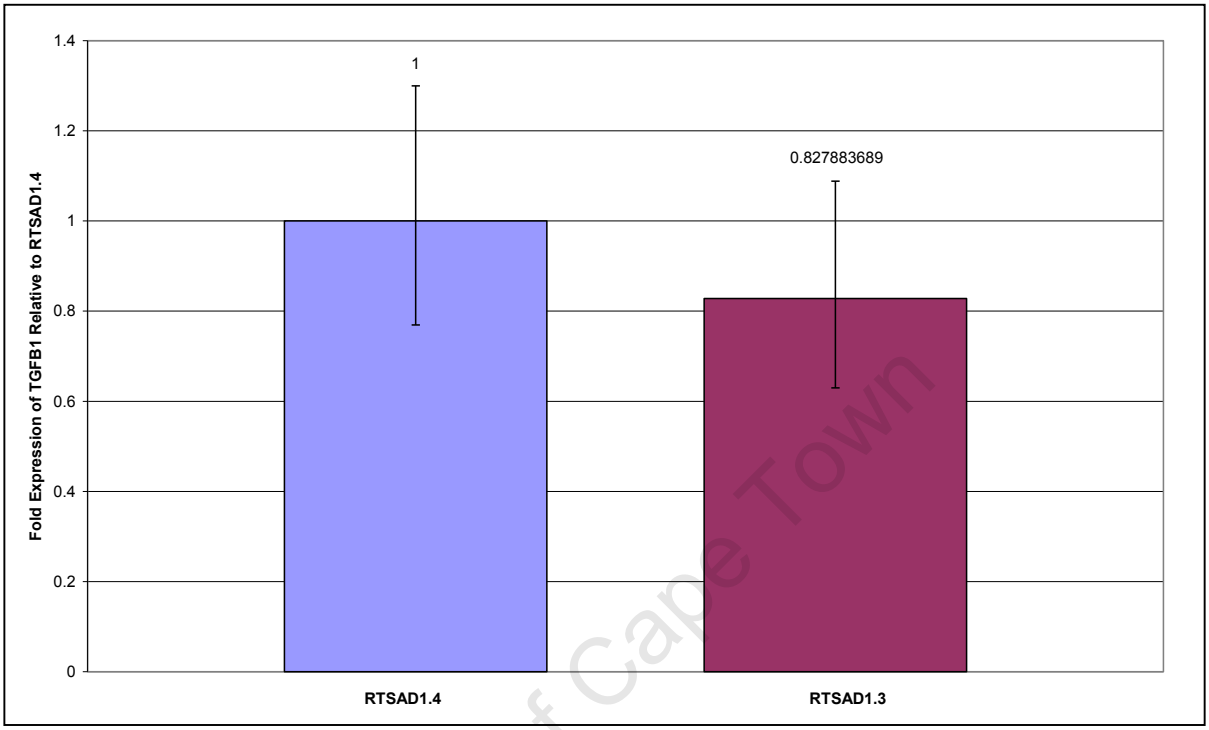


Figure 13. Normalised *TGFB1* gene expression relative to RTSAD1.4 (unaffected individual), evaluated using the $2^{-\Delta\Delta Ct}$ method. No significant difference was observed between the affected and unaffected individual (p = 0.43).

4.1.4.3 Determining Specificity of qPCR and Checking for the Presence of PCR Contamination

Because SYBR Green binds non-specifically to all double-stranded DNA in the reaction vessel, it was necessary to confirm that the increase in fluorescence observed during the amplification cycles of qPCR was a result of specific amplification of the desired fragment, and not a result of primer dimerisation. Melting curve analysis, performed on the LightCycler (Roche Diagnostics) after amplification, yielded a single peak at approximately 80°C in the cDNA-containing samples, for each primer pair. This indicated the presence of a single, specifically amplified fragment. A smaller peak, at approximately 75°C, was observed in some of the No-Template controls (NTCs), suggesting that primer dimerisation occurred only in the absence of template cDNA (data not shown).

Results of the melting curve were confirmed by subjecting representative samples (one replicate per individual per primer) to agarose gel electrophoresis (Fig. 14). A single band of the desired size was observed in all template-containing reaction tubes, demonstrating specific amplification of the desired fragment, while a faint band of low molecular weight, indicative of the formation of primer dimer, was only observed in NTCs.

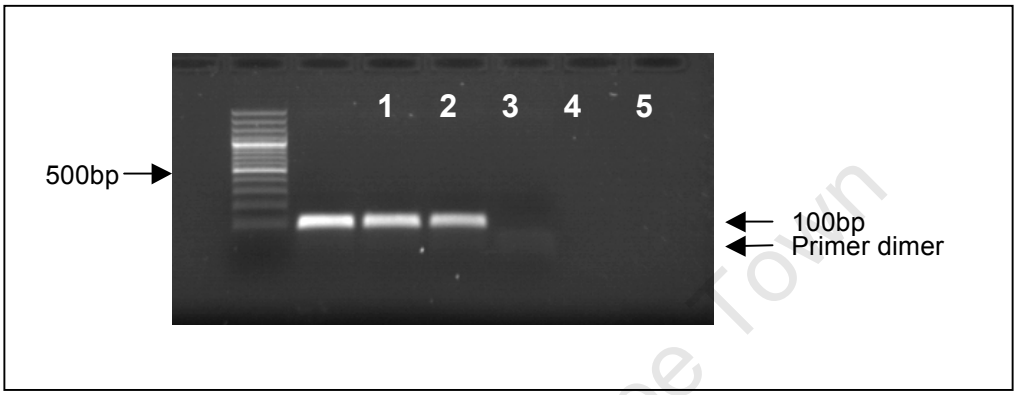


Figure 14. Products of qPCR with *GUSB* primers, following electrophoresis on a 1% agarose (0.5µg/ml ethidium bromide) gel to determine specificity of amplification. The approximate length of the PCR product, as well as the presence of primer dimer in the no template control (NTC), is indicated by an arrow. The size of a reference band in the molecular weight marker is shown. 1: Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2 to 4: Products of qPCR from template-containing reaction vessels, 5: No template control.

4.2 Analysis of the Smad Signalling Pathway

Key components of the Smad signalling cascade were investigated, to determine their contribution to the development of the fibrotic phenotype. These included the receptor-regulated Smads 2 and 3, whose expression and phosphorylation was determined by western blot, and the inhibitory Smad 7, which was investigated at the mRNA expression level by means of qPCR.

4.2.1 Western Blot Optimisation

To ensure optimal protein detection, numerous experimental parameters were adjusted, including the type of membrane used for transfer, the dilutions of primary and secondary antibody, and the duration of exposure of the membrane to either film or chemi-imager.

It was found that nitrocellulose membrane, and in particular, the type of membrane designed for enhanced chemiluminescent detection, produced the best results, allowing for maximal protein transfer, with minimal background signal. Optimal antibody dilutions generally followed those recommended by the manufacturer, with the exception of the secondary antibody concentration required for the detection of phosphoSmad3, which was increased from 1:2000 to 1:1000, in order to increase the intensity of signal. Finally, the duration of exposure of the membrane to the film or chemi-imager varied between antibodies, with optimal exposure chosen to be the time at which differences in band intensities could be detected and background signal was minimised (data not shown).

4.2.2 Smad2/3 Expression in Family and Control Samples

In order to detect possible aberrations in basal and TGF-β1-inducible receptor-associated Smad expression, western blotting was performed, using anti-Smad2/3 antibody, on extracts from cultured fibroblasts taken from the affected and unaffected family members (RTSAD1.3 and RTSAD1.4) and two age- and sex-matched controls. Comparisons were then made between extracts taken from fibroblasts stimulated with 800pM TGF-β1 for 1 hour, and unstimulated cells.

In all the sample lanes, a single band of high intensity was observed, with a molecular weight of 60kDa, corresponding to the predicted size of the Smad2 protein. An additional lighter band, of approximately 52kDa was assumed to be the Smad3 protein. Results reported refer to the analysis of the Smad2 band, as it was assumed to be representative of the expression of the receptor-regulated Smad proteins.

As can be seen in Figure 15 below, none of the samples displayed marked up- or downregulation of Smad2/3, in response to treatment. This assertion was confirmed by statistical analysis of the band densities captured by means of the chemi-imager. Using the Student's *t*-test for paired samples, no statistically significant difference was observed between the treated and untreated samples, for each individual (all *p*-values ≥ 0.1).

To determine whether Smad2/3 expression differed between individuals, the band density ratio of treated:untreated was calculated for each individual, and the results compared between the two family members (RTSAD1.3 and RTSAD1.4) and between each family member and their respective matched control (Control 2 and Control 1, respectively), using the Student's *t*-test for unpaired samples, assuming unequal variances. Again, no significant difference was detected.

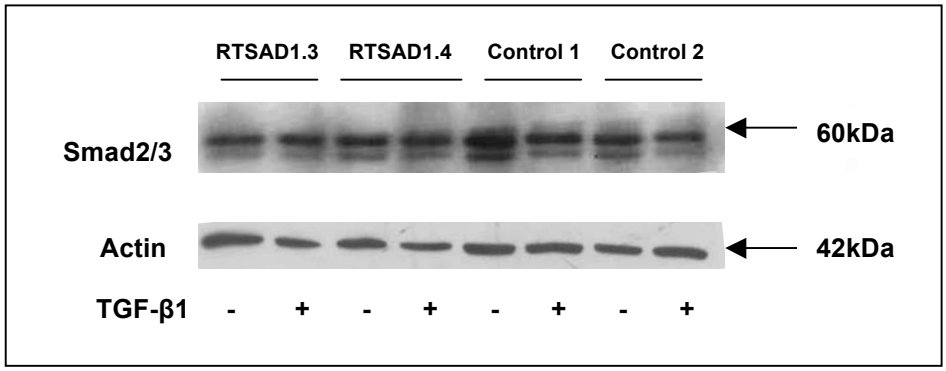


Figure 15. Basal and TGF-β1-inducible Smad2/3 expression in RTSAD1.3, RTSAD1.4, and two age- and sex-matched controls (Control 2 and Control 1, respectively). Whole cell lysates from treated (with 800pM TGF-β1 for 1 hour) and untreated fibroblasts were separated by SDS-PAGE and transferred to a nitrocellulose membrane. Actin was used as a loading control. Bound antibody was detected by chemiluminescence as described in Materials and Methods. No significant difference in expression was observed between treated and untreated samples for each individual, or between the family samples and their respective controls.

4.2.3 Smad3 Phosphorylation in Family and Control Samples

Smad signalling was also examined by means of western blot, using a monoclonal rabbit antibody specific to the phosphorylated form of the Smad3 protein. As for Smad2/3, western blotting was performed using whole cell lysates from treated and untreated fibroblasts, for each of the family members and controls. Comparisons were made between treated and untreated samples for each individual, and between individuals, to determine the level of Smad3 phosphorylation in response to TGF-β1 treatment.

In all sample lanes, the brightest band detected, at approximately 52kDa, was assumed to be phosphorylated Smad3 (Fig. 16). However, the low concentration of this protein in the whole cell lysate necessitated the use of a high secondary antibody concentration, in order to ensure detection. This resulted in the detection of numerous non-specific bands and a high background signal, making analysis using the chemi-imager less reliable.

Nonetheless, a distinct difference can be seen when comparing the treated and untreated lanes for each individual. Despite the high background signal observed in some of the lanes, it appears that phosphorylated Smad3 is virtually absent from the unstimulated fibroblasts of all individuals, regardless of disease status. In comparison, however, stimulation with TGF-β1 appears to result in a rapid increase in Smad3 phosphorylation (observed as an approximate two-fold increase in band density between treated and untreated samples) in all individuals, the effects of which are visible after 1 hour of treatment.

As for Smad2/3, the ratios of band densities between treated and untreated lanes, generated by the chemi-imager, were compared between individuals, using the Student's *t*-test for unpaired samples, assuming unequal variances. Although there appeared to be a difference in levels of phosphorylated protein between each family member (RTSAD1.3 and RTSAD1.4) and their matched controls (Control 2 and Control 1, respectively), these differences were not statistically significant (all *p*-values ≥ 0.19).

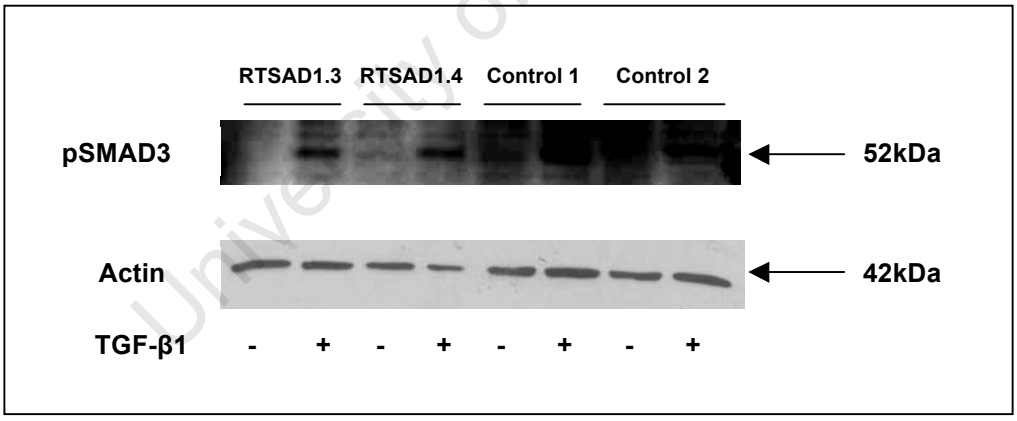


Figure 16. Basal and TGF-β1-inducible Smad2/3 phosphorylation in RTSAD1.3, RTSAD1.4, and two age- and sex-matched controls (Control 2 and Control 1, respectively). Whole cell lysates from treated (with 800pM TGF-β1 for 1 hour) and untreated fibroblasts were separated by SDS-PAGE and transferred to a nitrocellulose membrane. Actin was used as a loading control. Bound antibody was detected by chemiluminescence as described in Materials and Methods. A noticeable increase in phosphorylation in response to treatment was observed for each individual. No significant difference in phosphorylation was observed between the family samples and their respective controls.

4.2.4 Determination of SMAD7 mRNA Expression Levels by qPCR

The basal and TGF-β1-inducible expression levels of *SMAD7* in cultured skin fibroblasts of the affected son and his unaffected mother were quantified by qPCR (using SYBR Green reaction chemistry), and the results were analysed using the $2^{-\Delta\Delta CT}$ method, as previously described.

4.2.4.1 Determination of Amplification Efficiencies for Target (SMAD7) and Reference (GUSB) Genes

The amplification efficiencies of the target (*SMAD7*) and reference (*GUSB*) genes were compared, in order to confirm the validity of the $2^{-\Delta\Delta CT}$ method. As described in 4.1.4.1, both genes were amplified in triplicate from serially diluted control cDNA, and the ΔC_T values ($C_{T,SMAD7} - C_{T,GUSB}$) were plotted against the logarithm of the concentration. The slope of the linear regression line was determined to be approximately -0.04, the absolute value of which was sufficiently close to zero to assume that the efficiency of amplification of the two genes was equal, and hence that the $2^{-\Delta\Delta CT}$ method could be used to analyse the qPCR results generated using these primers.

4.2.4.2 Basal SMAD7 Expression in Family Samples and Analysis Using the $2^{-\Delta\Delta CT}$ Method

Basal *SMAD7* expression was determined in cultured fibroblasts taken from the two family members, which had not been stimulated with exogenous TGF-β1. Both *GUSB* and *SMAD7* were amplified from RTSAD1.3 and RTSAD1.4 cDNA in triplicate, and the crossing points and mean crossing points were calculated. Expression of *SMAD7* was normalised to *GUSB* for each individual, by calculating the ΔC_T ($C_{T,\ TGFB1} - C_{T,GUSB}$). The $\Delta\Delta C_T$ was calculated using Eq. [1a] (see 3.2.8.5), and the fold change in *SMAD7* expression, relative to the unaffected individual, was determined using Eq. [2] (3.1.5.3).

Based on these calculations, no significant difference in normalised *SMAD7* expression was observed between RTSAD1.3 and RTSAD1.4 (p=0.22), indicating that basal levels of the *SMAD7* transcript in the affected individual

are neither abnormally elevated nor depressed, in relation to the unaffected control individual (Fig. 17).

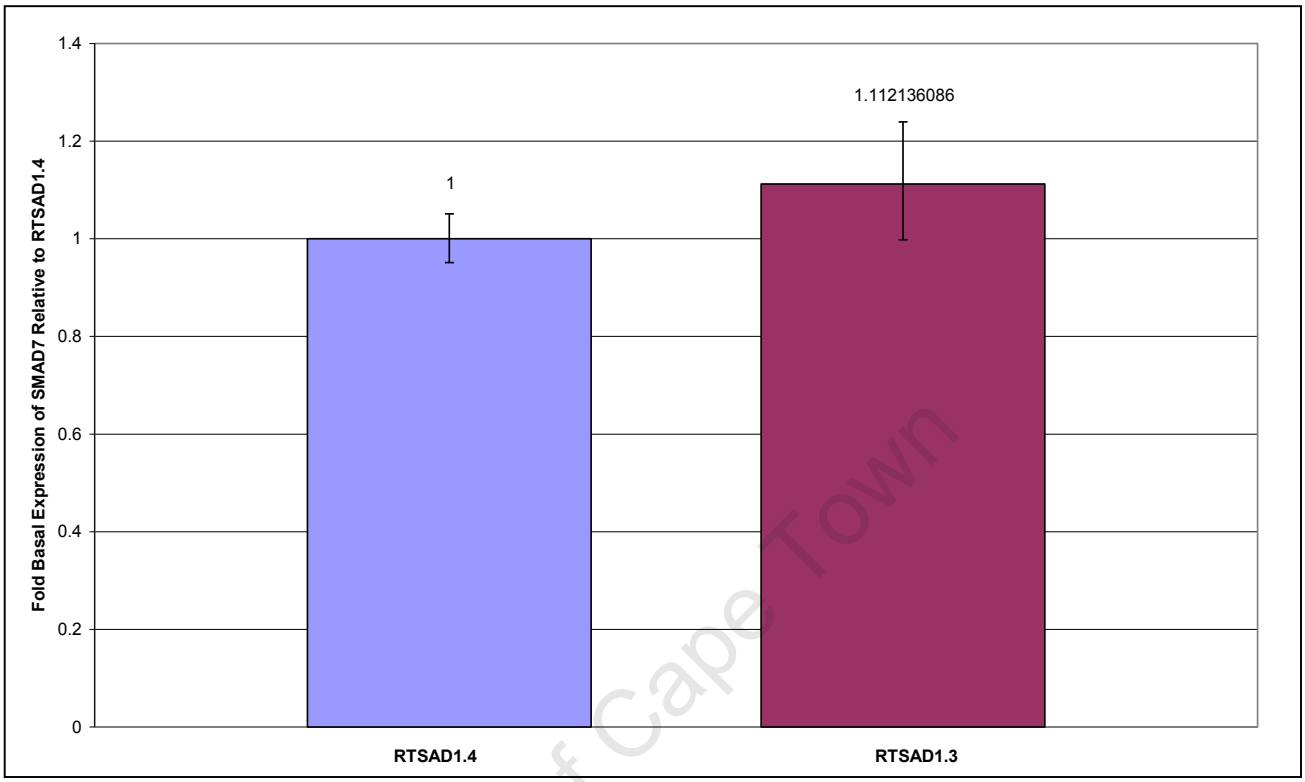


Figure 17. Normalised *SMAD7* gene expression in unstimulated fibroblasts, relative to RTSAD1.4 (unaffected individual), evaluated using the $2^{-\Delta\Delta Ct}$ method. No significant difference in expression was observed between the affected and unaffected individual (p = 0.22).

4.2.4.3 RNA Isolation and cDNA Synthesis

Total RNA was isolated from treated (with 500pM recombinant human TGF-β1 for 1 hour) and untreated RTSAD1.3 and RTSAD1.4 cultured skin fibroblasts at passage seven to nine, and subjected to agarose gel electrophoresis to determine the quality and quantity of the RNA.

The 18 and 28S rRNA bands could clearly be seen in all samples, at a fluorescent intensity ratio of 2:1, and since no smearing was present, the RNA was deemed to be intact (Fig. 18). Although the fluorescent intensity of the bands under UV light indicated a low yield after extraction (particularly when compared to the bands visible in Fig. 5), sufficient RNA was present to proceed with further experiments.

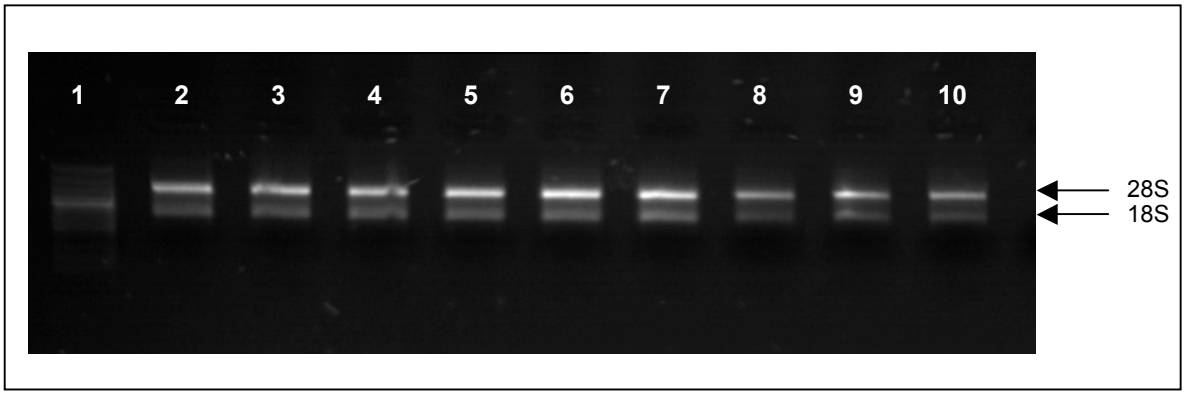


Figure 18. Total RNA extracted from treated and untreated RTSAD1.3 and RTSAD1.4 skin fibroblasts, electrophoresed on a 1% agarose (0.5µg/ml Ethidium Bromide) gel to determine integrity. The positions of the 28S and 18S rRNA bands are indicated with arrows. 1: Gene Ruler^{TM} 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2 to 4: RNA isolated from RTSAD1.3 fibroblasts treated with 500pM TGF-β1, 5 to 7: RNA isolated from untreated RTSAD1.3 fibroblasts, 8 to 10: RNA isolated from RTSAD1.4 fibroblasts treated with 500pM TGF-β1.

To detect possible genomic DNA contamination in the cDNA samples, after isolation and reverse transcription, PCR amplification was performed on a sample of RTSAD1.4 cDNA, using primers complementary to the intronic regions flanking exon 3 of *TGFB1*. Amplification was anticipated only in the presence of genomic DNA, since cDNA contains no introns. A sample of RTSAD1.4 genomic DNA was included as a positive control.

As predicted, amplification of exon 3 took place when genomic DNA was used as a template. In the PCR containing cDNA, however, no product was generated, indicating the absence of genomic DNA in the sample (Fig. 19).

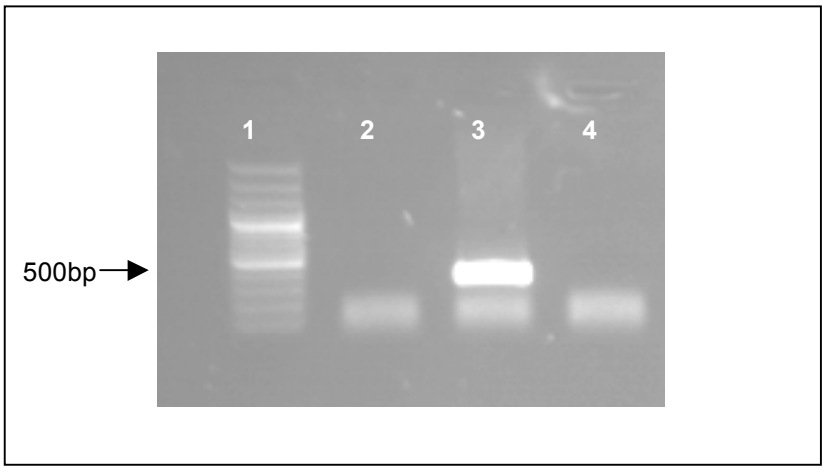


Figure 19. PCR using intronic primers, designed to amplify exon 3 of *TGFB1*, to determine the presence of genomic DNA contamination. The size of a reference band in the molecular weight marker is shown. 1: Gene Ruler^{TM} 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2: DNA-free water control, 3: RTSAD1.4 genomic DNA positive control, 4: RTSAD1.4 cDNA.

4.2.4.4 TGF-β1-Inducible SMAD7 Expression in Family Samples and Analysis Using the $2^{-\Delta\Delta CT}$ Method

The effect of TGF-β1 on *SMAD7* gene expression was examined by means of qPCR, using cDNA synthesised from RNA isolated from treated and untreated RTSAD1.3 and RTSAD1.4 fibroblasts as a template. The response to TGF-β1 treatment was then compared between individuals, to determine whether differences in induction of this inhibitory Smad could account for the observed phenotype.

Both *GUSB* and *SMAD7* were amplified from treated and untreated RTSAD1.3 and RTSAD1.4 cDNA samples in duplicate, and the crossing points and mean crossing points were calculated. The expression of *SMAD7* was normalised to *GUSB*, by calculating the ΔC_T ($C_{T,\ TGFB1} - C_{T,GUSB}$). For each individual, the $\Delta\Delta C_T$ was calculated using Eq. [1b] {$\Delta\Delta C_T = (C_{T,\ SMAD7} - C_{T,\ GUSB})_{treated} - (C_{T,\ SMAD7} - C_{T,\ GUSB})_{untreated}$}, and the fold change in *SMAD7* expression in response to treatment was determined using Eq. [2] (3.1.5.3).

Based on these results, it was evident that treatment with 500pM TGF-β1 for 1 hour caused a significant upregulation of *SMAD7* mRNA expression in both RTSAD1.3 (p=0.0002) and RTSAD1.4 (p=0.005) fibroblasts, resulting in a

two-fold increase in expression relative to the basal level of transcription of the gene. However, when comparing the response to TGF-β1 treatment between individuals, no significant difference was observed (p=0.88). Thus it appears that fibroblasts from both the affected and the unaffected individual are equally capable of *SMAD7* upregulation in response to TGF-β1 signalling (Fig. 20).

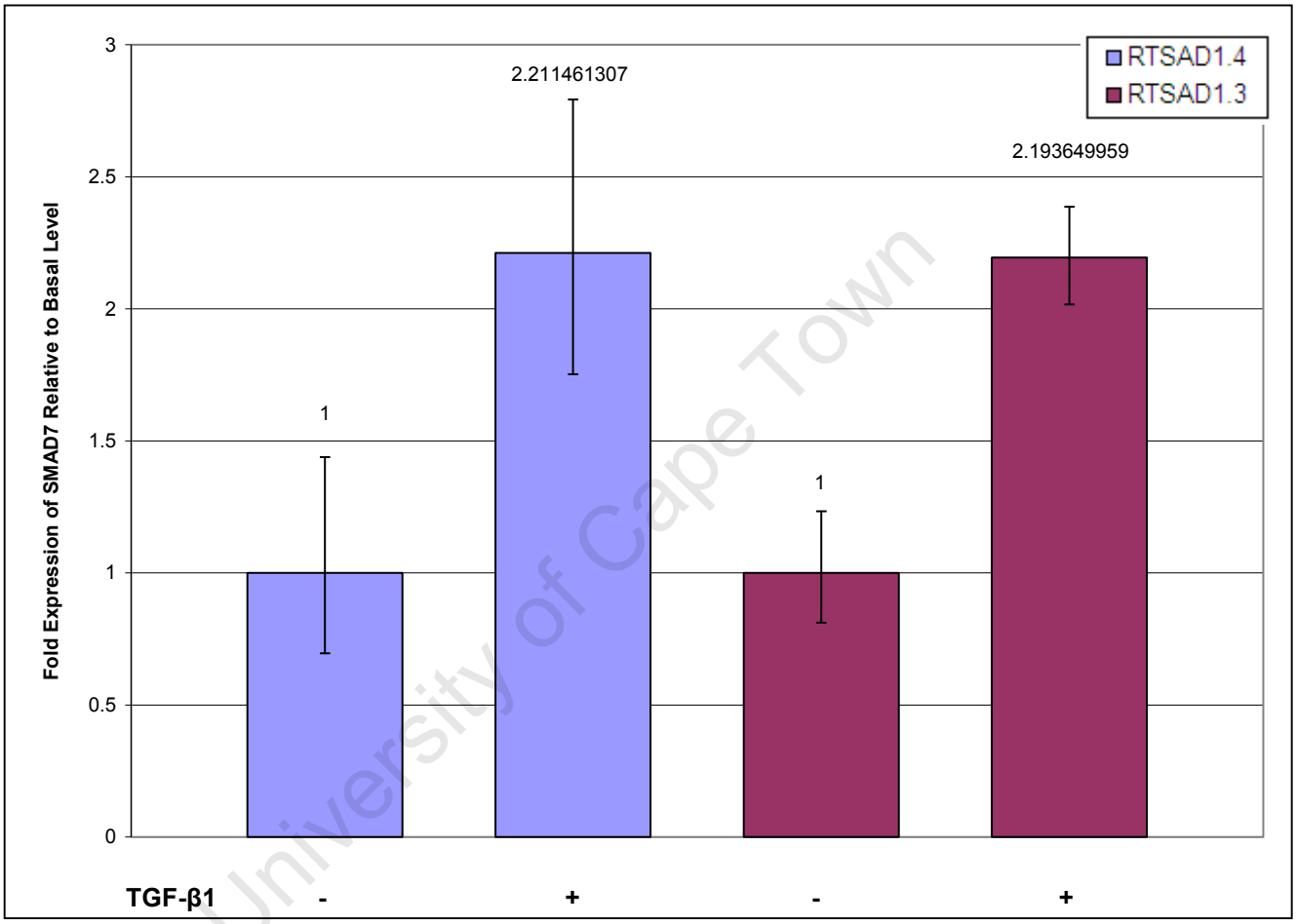


Figure 20. Fold change in *SMAD7* expression in response to treatment with 500pM TGF-β1 for 1 hour, relative to the untreated sample (-) for each individual. A significant two-fold upregulation of *SMAD7* transcription was observed in both RTSAD1.3 and RTSAD1.4 (p=0.0002 and p=0.005, respectively), however this TGF-β1-inducible response did not differ between individuals (p=0.88).

4.2.4.5 Determining Specificity of qPCR

To determine the specificity of each of the *SMAD7* qPCRs, melting curve analysis was performed on the LightCycler (Roche Diagnostics). For each primer pair (*SMAD7* or the reference gene, *GUSB*), a single peak was generated, with a fragment-specific melting temperature of around 80°C, in

the cDNA-containing samples. This indicated the presence of a single, specifically amplified fragment. A smaller peak, at approximately 75°C, was observed in some of the No-Template controls (NTCs), suggesting that primer dimerisation occurred only in the absence of template cDNA (data not shown).

Results of the melting curve were confirmed by subjecting representative samples (one replicate per individual per primer) to agarose gel electrophoresis (Fig. 21). A single band of the desired size was observed in all template-containing reaction tubes, demonstrating specific amplification of the desired fragment.

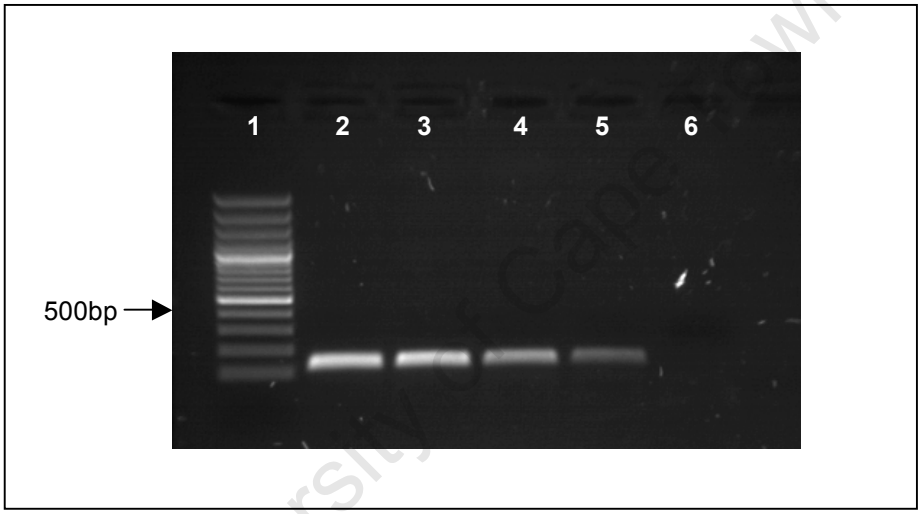


Figure 21. Products of qPCR with *SMAD7* primers, following electrophoresis on a 1% agarose (0.5µg/ml ethidium bromide) gel to determine specificity of amplification. The size of a reference band in the molecular weight marker is shown. 1: Gene Ruler™ 100bp DNA Ladder Plus Molecular Weight Marker *Improved* (Fermentas), 2 to 5: Products of qPCR from template-containing reaction vessels, 6: No template control.

5. Discussion

The autosomal dominant fibrotic disorder described by Khumalo et al. (2006), characterised by poikiloderma, tendon contractures and progressive pulmonary fibrosis, was accorded novel status based on its unique phenotype and mode of inheritance, when compared to other related disorders (Khumalo et al. 2006). As the biological basis of the disease remained unknown, the molecular characterisation of this disorder began with the identification of genes and/or proteins which could be responsible for the observed phenotype.

A candidate gene approach was adopted to prioritise genes for investigation, since the small number of individuals available for study prevented the use of statistically-based methods, such as genome-wide analyses (McCarthy et al. 2003; Ott 1991). Based on the prominence of fibrosis in affected individuals, genes encoding proteins with a proven role in ECM deposition and fibrogenesis were chosen as candidates. These included *TGFB1*, which encodes the profibrotic cytokine TGF-β1; and several components of the Smad pathway, namely Smads 2, 3 and 7, which regulate the expression of ECM proteins in response to TGF-β1 signalling (Clouthier et al. 1997; Kavsak et al. 2000; Nakao et al. 1997b; Zhang et al. 1996). Mutations in each of these candidates have been previously shown to contribute to the development of fibrotic disease (Arany et al. 2006; Blobe et al. 2000; Clouthier et al. 1997; Nakao et al. 2002). The aim of this investigation was thus to determine the role of *TGFB1* and the abovementioned Smad proteins in the molecular basis of this novel disorder.

Since the manifestations of this novel familial disorder are thought to be, at least in part, a consequence of excessive fibrosis, it was anticipated that disease-causing defects in *TGFB1* should lead to overexpression, or constitutive activity, of its profibrotic protein product (Dong et al. 2002; Khumalo et al. 2006). Thus, mutations were anticipated to primarily affect the signal sequence (disrupting trafficking and processing of the protein through the Golgi or Endoplasmic Reticulum) (Awad et al. 1998) or the Latency-

Associated Peptide, resulting in dysregulation of activity of the TGF-β1 protein (Annes et al. 2003; Janssens et al. 2000; Munger et al. 1999).

Previous research performed on this family had focused on the confirmation of *TGFB1* as a candidate gene using linkage-based exclusion mapping (Ott 1991), and on mutation screening, by means of direct sequencing of the exons of *TGFB1* in genomic DNA of an affected and an unaffected family member. Based on these results, *TGFB1* could not be excluded as a candidate. However, the single nucleotide polymorphisms identified during mutation screening were deemed not to account for the phenotype observed (Hartmann 2006).

In this study, therefore, the investigation of *TGFB1* was expanded to examine the length and sequence of the transcript, since whole-exon deletions, produced as a result of aberrant splicing, could not be detected using the original method of sequencing single exons at a time. The length of the *TGFB1* transcript in skin fibroblasts from RTSAD1.3 and RTSAD1.4 was determined by PCR amplification and agarose gel electrophoresis, in order to ascertain the presence of large-scale deletions. The lengths of the PCR products generated for each individual, determined by comparison to the molecular weight marker, appeared sufficiently close to the predicted lengths based on the NCBI reference sequence (www.ncbi.nlm.nih.gov) to rule out the possibility of whole-exon deletions. Agarose gel electrophoresis was selected as a means of analysis because of its greater range of separation, when compared to polyacrylamide gel electrophoresis, allowing detection of large variations in fragment length (0.5 to 7kb for 1% agarose) (Sambrook et al. 1989). However in comparison to polyacrylamide, the resolving power of agarose is limited, preventing accurate detection of deletions of the smaller exons (Sambrook et al. 1989). For this reason, the results of electrophoresis were confirmed by direct sequencing of the *TGFB1* cDNA.

It should be noted that although five members of the family had been diagnosed with the disease, the decision was taken to use only two individuals for this study, namely an affected brother, RTSAD1.3, and the

unaffected mother, RTSAD1.4, who served as a negative control. RTSAD1.3 was used as a representative of the affected offspring, in the interests of efficiency and economy, since this individual could most readily be contacted for the donation of blood and skin biopsies (RTSAD1.1 being deceased, and RTSAD1.2 having made herself unavailable for further study). It was assumed that disease-causing variants would be present in all affected individuals, and thus it was not necessary to screen RTSAD1.1 or RTSAD1.2 until such a variant had been identified.

Alignment of the *TGFB1* transcript sequences generated for both RTSAD1.3 and RTSAD1.4 with the NCBI reference sequence confirmed that all splice junctions were intact, and that the *TGFB1* mRNA had been correctly processed in both individuals. Thus it was concluded that aberrant splicing of *TGFB1* could not account for the symptoms observed.

Although no whole-exon deletions were observed in the *TGFB1* cDNA for either individual, two single nucleotide polymorphisms (SNPs) were identified in the transcript, during the course of sequencing. Both of these were located in the signal sequence (at codons 10 and 25) and have been extensively described in the literature.

The first, a T→C substitution located in exon 1 of the transcript, was found to be present in the heterozygous (T/C) state in RTSAD1.3, and in the homozygous state in RTSAD1.4 (C/C), and was identified as the c.+29T>C SNP (rs1982073) M(numbered relative to the translation start site). This nonsynonymous C to T substitution results in an amino acid change from Leucine to Proline at codon 10, in the signal sequence of the TGF-β1 prepropeptide (Awad et al. 1998; Dunning et al. 2003).

The signal sequence of a protein is a stretch of 15 to 25 amino acids comprised of three regions: a short N-terminal made up of positively charged residues, a central hydrophobic core (8 to 12 residues long) and a polar region, which defines the cleavage site. The hydrophobic core, which adopts an alpha-helical conformation, directs transport of the protein across the Golgi

and ER (Awad et al. 1998). The replacement of Leucine (encoded by the T variant at this SNP) with Proline (encoded by the C variant) has been demonstrated to cause an increase in *TGFB1* transcription and translation (Suthanthiran et al. 2000; Yamada et al. 1998), leading to a 2.8 fold increase in circulating TGF-β1 levels (Dunning et al. 2003; Shin et al. 2005).

The elevation of TGF-β1 expression associated with Pro^{10} is postulated to be a result of altered protein export across the ER. The cyclical structure of Proline may alter the alpha-helical proportions of the signal sequence backbone, in contrast to the aliphatic Leucine side chain, which favours the formation of alpha helices (Crilly et al. 2002). This structural alteration may, in turn, influence protein secretion and, since TGF-β1 is capable of regulating its own transcription via AP-1 sites within its promoter region, retained TGF-β1 may upregulate its own expression by means of intracrine signalling (Suthanthiran et al. 2000).

Increased levels of TGF-β1, corresponding to either a C/C or T/C genotype, have been associated with a number of disease phenotypes. Crilly et al. (2002) found that significant numbers of systemic sclerosis sufferers were heterozygous (T/C) at codon 10 ($p<0.0001$), while Xaubet et al. (2003) demonstrated a correlation between Pro^{10} and rapid deterioration in patients with Idiopathic Pulmonary Fibrosis ($p=0.002$).

Thus it was anticipated that individuals showing symptoms of the novel disorder in this study, characterised by excessive fibrosis, particularly of the lungs and skin, should carry a Pro^{10} variant, indicative of increased expression of profibrotic TGF-β1. While RTSAD1.3, as a T/C heterozygote, did indeed carry one allele coding for Proline at this position, the unaffected mother, RTSAD1.4, was homozygous C/C (i.e. Pro/Pro) at the equivalent position. This appears counterintuitive, particularly since a dose effect has been demonstrated for this codon, indicating that C/C individuals produce higher levels of TGF-β1 than heterozygotes, or T/T homozygotes (Yamada et al. 1998).

However, the role of this SNP in fibrotic disease remains controversial. While most studies agree on the role of the "C" (Pro) allele in the elevation of TGF-β1 expression, some research groups have found the opposite (Mattey et al. 2005). Among them are Awad and colleagues, who reported an increased frequency of the Leu^{20} variant in cystic fibrosis patients with lung fibrosis (p<0.01), a fact which they hypothesise to be due to increased expression of TGF-β1 in individuals with the T/T genotype (Awad et al. 1998).

Additionally, the relatively high allele frequencies recorded for this SNP in Caucasians (C=0.549, T=0.451) on the NCBI SNP database (www.ncbi.nlm.nih.gov) indicate that neither the T nor the C allele at this codon alone could account for the symptoms and inheritance pattern of this rare, novel disease. Together with the conflicting reports regarding the effect of the c.+29T>C variant on TGF-β1 expression levels, these facts suggest that this is a common polymorphism that is unlikely to play a role in the pathogenesis of the disease under study.

The second SNP, identified as the c.+74G>C polymorphism, was found in the homozygous (G/G) state in the affected individual, and in the heterozygous (G/C) state in his unaffected mother. This variant is also located in the signal sequence of TGF-β1, at codon 25, proximal to the cleavage site at codon 29 (Carturan et al. 2004). The change from a G to a C at this position corresponds to an amino acid change from an Arginine to a Proline. This results in the exchange of the large, polar side chain of Arginine for the small non-polar Proline side chain, which has been postulated to either disrupt the hydrophobic core of the signal sequence, or to affect the adjacent cleavage site (Stoll et al. 2004; Awad et al. 1998; Cambien et al. 1996). Protein function may also be altered, due to a predicted change in the folding of the amino acid chain (Awad et al. 1998).

The c.+74 G/G (Arg/Arg) genotype has been shown to result in increased production of TGF-β1 by stimulated lymphocytes when compared either to heterozygotes, or to C/C homozygotes (Awad et al. 1998). Thus it follows that homozygosity for Arginine at codon 25, correlating to higher *in vitro* levels of

TGF-β1, has been significantly associated with systemic sclerosis (p=0.027) (Crilly et al. 2002) and pretransplant fibrotic lung pathology (p<0.02) (Awad et al. 1998). Notably, the combination of an Arg/Arg genotype at codon 25 and a codon 10 Leucine allele has been identified as a marker for poor post-transplant prognosis due to fibrosis (Awad et al. 1998).

The codon 25 genotypes of RTSAD1.3 (G/G) and RTSAD1.4 (G/C) therefore fit the hypothesised model of higher levels of TGF-β1 (leading to increased fibrogenesis) in individuals affected by the novel disorder, when compared to unaffected individuals. Additionally, the strong association of the Arg^{25} variant with pulmonary fibrosis, and its proposed combinatorial effect with Leu^{10} to bring about the fibrotic phenotype, raises the possibility that these variants, acting in combination with others, may play a role in the molecular basis of this disease. However, such considerations must be treated with caution, since the high frequency of the G allele in Caucasians (0.887) (www.ncbi.nlm.nih.gov) implies that this variant is common in the background population and hence cannot be solely responsible for the observed fibrotic phenotype and inheritance pattern of this disorder.

In addition to the length and sequence of the transcript, *TGFB1* mRNA expression levels were examined, in fibroblasts taken from the affected and unaffected family members, by means of real-time quantitative PCR (qPCR). It was anticipated that quantification of the transcript would provide information about the regulation of the gene, while eliminating the need to sequence the extended regulatory region, which comprises more than 2.6kb of sequence upstream of the transcription start site (Shah et al. 2006). Alterations in various components of this regulatory region have previously been shown to result in constitutive or abnormally elevated expression of the gene, resulting in the development of fibrosis (Shah et al. 2006; Shin et al. 2005; Silverman et al. 2004; Grainger et al. 1999). Additionally, it was expected that qPCR analysis of the transcript would clarify the effect of the abovementioned polymorphisms on the expression of *TGFB1*.

Analysis of the results of qPCR was performed by means of relative quantification, using the $2^{-\Delta\Delta CT}$ method described by Livak and Schmittgen (2001). Relative quantification describes the change in expression of a target gene (*TGFB1*) in an experimental group, relative to a reference group. An endogenous reference gene is included, to normalise for the amount of RNA added to the initial reverse transcription reactions (Livak and Schmittgen 2001). In this case, the affected individual was treated as the experimental group, while the unaffected individual served as the reference. The housekeeping gene *GUSB* was chosen as an endogenous reference gene, once the efficiencies of amplification of both genes had been calculated, and found to be approximately equal.

The results of the qPCR experiment showed no significant difference in *TGFB1* expression between RTSAD1.3 and RTSAD1.4 (p=0.43). Thus, it could be inferred that overexpression of profibrotic *TGFB1* was not responsible for the disease phenotype observed in the affected individual.

This finding was significant, since it implied that the disease-causing defect was not present in either the extended 5' regulatory region of *TGFB1*, or in any other genetic component regulating the expression of the gene. Elimination of *TGFB1* as a candidate was an important step in the characterisation of this fibrotic disease, since overexpression of this gene, and its protein product, are widely regarded as central to the development of fibrosis in numerous diseases, affecting most tissues of the body (Verrecchia et al. 2007; Shah et al. 2006; Bartram and Speer 2004; Nakao et al. 2002; Blobe et al. 2000; Awad et al. 1998).

Additionally, the results of qPCR indicate that neither of the signal sequence polymorphisms described above significantly modifies the expression of the *TGFB1* transcript in either individual. The *p*-values calculated for this experiment must however be treated with caution, since the small sample size used makes detection of small variations in expression difficult (Whitley and Ball 2002). Nevertheless, it appears that the biological basis of this disease

cannot be attributed either to structural defects in the *TGFB1* transcript, or to aberrant regulation of its expression.

The focus of the investigation therefore shifted to the examination of various Smad proteins, which had been identified as candidate genes based on their role in the transduction of TGF-β1 signalling and the regulation of genes involved in fibre deposition (Dong et al. 2002; Nakao et al. 2002; Varga et al. 2002; Venkatesan et al. 2004). The receptor-regulated Smads, namely Smad2 and Smad3, are phosphorylated by ligand-bound TGF-β1 receptors, resulting in a conformational change allowing recruitment of the common Smad, Smad4. This complex is then translocated to the nucleus, where it regulates the expression of various fibrogenic target genes in response to TGF-β1 signalling (Arai et al. 1998; Chen et al. 2002; Mori et al. 2000; Nakao et al. 1997b; Tsukazaki et al. 1998; Zawel et al. 1998; Zhang et al. 1996).

In this study, expression and phosphorylation of the R-Smads were quantified in TGF-β1-stimulated and unstimulated fibroblasts from both individuals. Since the excessive fibrosis observed in affected individuals is thought to result from hyperinduction of the profibrotic signalling cascade, the hypothesis was formulated that overexpression or constitutive phosphorylation of the R-Smads could account for the development of the disease (Dong et al. 2002). In structural terms, this could result from a mutation in the MH1 domain, affecting its ability to inhibit the receptor-binding and oligomerisation activity of the MH2 domain (Lo et al. 1998; Lo et al. 1999; Massagué 1998; Ross et al. 2007; ten Dijke et al. 2000) and thus producing an active protein conformation in the absence of a stimulus. Alternatively, mutations producing constitutively-active forms of either the C-tail (containing the SS[X]S motif phosphorylated by TGF-β receptors) or the linker region (phosphorylated by other Smad pathway regulators, such as MAP kinases) could result in ligand-independent signalling (Lo et al. 1998; Lo et al. 1999; ten Dijke et al. 2000). Finally, inactivation of the SARA complex, responsible for subcellular localisation of the receptor-regulated Smads in the absence of signal (Chacko et al. 2001; Tsukazaki et al. 1998), could also disrupt control of the pathway.

However, when the expression of Smad2/3 was examined, no significant difference was observed between fibroblasts which had been treated with TGF-β1, and those that had not, for each of the family members (RTSAD1.3 or RTSAD1.4), and their age- and sex-matched controls (Control 2 and Control 1, respectively). Thus it would appear that stimulation with TGF-β1 has no effect on the expression of Smad2/3. This is to be expected since, under normal physiological conditions, the induction of the Smad signalling cascade by TGF-β1 results in a change in the phosphorylation state and subcellular compartmentalisation of the R-Smads, but does not affect their expression (Mori et al. 2000; Varga et al. 2002).

Of greater interest, therefore, were the inter-individual differences in Smad2/3 expression, since previous research has demonstrated an association between elevated Smad3 expression and the development of fibrosis in the skin lesions of systemic sclerosis patients (Dong et al. 2002). In order to compare R-Smad expression between individuals, the ratio of treated:untreated band densities was calculated for each individual, and comparisons were then made between the two family members, and between each family member and his/her matched control. Again, no significant difference was observed between any of the individuals compared, suggesting that Smad2/3 is found at normal levels in the skin fibroblasts of affected individuals, and hence that overexpression of this profibrotic signalling molecule does not play a role in the development of the disorder.

It should however be noted that several technical errors could have influenced the results of this experiment. For instance, during cell culture, the speed and homogeneity of growth varied between cell lines and between replicates, resulting in differing degrees of confluence at the time of treatment with TGF-β1. This could have affected the concentration of TGF-β1 relative to cell count, the level of signalling (depending on the proximity of the cells), and the concentration of total protein present in the cell lysate after extraction. This could explain the discrepancy in band density between Control 1 and the other three cell lines, observed in Figure 15. Several measures were therefore taken in order to counteract these errors, including repeating the

extractions in triplicate and the western blots in duplicate, probing for the housekeeping protein Actin (as a loading control) and allowing the cells to grow to uniform (85%) confluence whenever possible. Normalisation of treated to untreated samples, prior to inter-individual comparisons, also minimised the discrepancies resulting from different speeds of growth between cell lines.

There are two further considerations which must be made when interpreting these results. Firstly, it has been demonstrated that the development of fibrosis may result from a failure to downregulate Smad3, in response to prolonged TGF-β1 signalling (Dong et al. 2002). Thus, although no difference in Smad2/3 expression was observed in this experiment, after one hour of treatment, results may differ as the duration of TGF-β1 exposure is increased. A time-course experiment is therefore proposed to more accurately determine the affected individual's ability to regulate Smad2/3.

Secondly, the role of the receptor-regulated Smads in perpetuating TGF-β1 signalling is strongly linked to their subcellular localisation. In scleroderma patients, for instance, excessive nuclear accumulation of R- and Co-Smads has been linked to the development of fibrosis (Dong et al. 2002). Immunohistochemical analysis is therefore recommended to determine the subcellular location of R- and Co-Smads in the fibrotic tissue of individuals affected with this novel disease.

R-Smad phosphorylation, indicative of the level of TGF-β1/Smad signalling, was quantified in fibroblasts of the affected and unaffected individual, as well as their matched controls, by means of western blot. Whole cell lysates from treated (with TGF-β1) and untreated fibroblasts of each individual were probed for phosphorylated Smad3, giving an indication of the basal and TGF-β1-inducible signalling response. It has been previously reported that constitutive Smad2 or Smad3 phosphorylation, indicative of hyperactivity of the Smad signalling cascade in the absence of a stimulus, can lead to the development of pathological fibrosis (Venkatesan et al. 2004). Similarly, an abnormally elevated phosphorylation response to TGF-β1 may also indicate

hyperactivity of the pathway and hence, a greater propensity for the accumulation of fibrous deposits (Dong et al. 2002).

Based on this hypothesis, it was anticipated that phosphorylated Smad3 should be observed in the whole cell lysates of both the treated and untreated fibroblasts of the affected individual, compared to the unaffected controls, in which Smad3 phosphorylation should only be detected in response to TGF-β1 treatment. Alternatively, the levels of Smad3 phosphorylation in stimulated fibroblasts were expected to be elevated in the affected individual, relative to the control individuals.

However, this was not the case. In the whole cell lysates of untreated fibroblasts of all four individuals, phosphorylated Smad3 was virtually absent, while treatment with TGF-β1 resulted in marked upregulation of signalling, yielding a clearly visible band corresponding to the phosphorylated R-Smad. When comparing differences in response to TGF-β1 treatment between individuals, no significant difference in the treated:untreated band density ratio was observed between any of the individuals. These results indicated that phosphorylation of Smad3 in the affected individual (RTSAD1.3) was neither constitutive, nor abnormally elevated, in comparison to phenotypically normal individuals and appeared to be physiologically responsive to ligand stimulation. Thus it was concluded that aberrant phosphorylation of the receptor-regulated Smad proteins could not account for the excessive fibrosis characteristic of this novel disorder.

Again, there were several limitations which should be taken into account when interpreting the results of this experiment. As previously mentioned, inter-individual variations in speed and homogeneity of cell growth may have affected the response of different cell lines to treatment with TGF-β1. However, as for western blot detection of Smad2/3, triplicate isolation, duplicate detection and the inclusion of an anti-Actin probe as a loading control were designed to counteract this.

However, an additional difficulty was encountered during western blot optimisation using the anti-phosphoSmad3 primary antibody. Due to the relatively low abundance of the phosphorylated protein in the whole cell lysate and the high selectivity of the monoclonal primary antibody used in detection, a secondary antibody dilution of 1:1000 (2-fold higher than the manufacturer's recommended dilution) in combination with a super-sensitive chemiluminescent substrate solution was required to detect the target protein. This resulted in high background signal, caused by non-specific binding of the secondary antibody. Various optimisation steps were therefore employed in an attempt to increase the intensity of the phospho-protein band and eliminate background signal, including increasing the concentration of TGF-β1 used in treatment {to 800pM, compared with the 500pM recommended by Dong et al. (2002)} and increasing the frequency, duration and volume of washes. However, reduced background could not be achieved without compromising detection of the phosphoSmad3 band. The results presented in Figure 16 therefore represent the best signal:noise ratio obtained with this combination of antibodies. In order to confirm these results, it is recommended that the reaction be further optimised (by adjusting various parameters such as antibody dilutions, washes or blocking reagents, or by the inclusion of an immunoprecipitation step to concentrate the target protein) to enable more accurate discrimination between specific bands and non-specific signal, thereby improving analysis.

The final component of the Smad pathway investigated in this study was the gene encoding the inhibitory Smad, *SMAD7*. In response to TGF-β1 signalling, *SMAD7* transcription is rapidly induced by activated R-Smad/Smad4 heteromers which have been translocated to the nucleus (Mori et al. 2000; Nakao et al. 1997a; Nakao et al. 2002; Takagawa et al. 2003). This initiates a negative feedback loop, in which the Smad7 protein downregulates TGF-β1 signalling by either binding to activated TGF-β1 receptors directly, preventing their association with R-Smads (Kloos et al. 2002; Nakao et al. 1997a), or by recruiting an E3 ubiquitin-protein ligase to the receptor complex, resulting in its degradation (Kavsak et al. 2000).

Several studies have shown that the expression level of *SMAD7* influences cellular responsiveness to TGF-β1 (Nakao et al. 1997a; Venkatesan et al. 2004). Specifically, decreased expression of *SMAD7*, resulting in failure to terminate Smad signalling in response to TGF-β1, has been shown to be associated with the development of fibrosis in the skin and lungs of bleomycin-treated mice (Takagawa et al. 2003; Venkatesan et al. 2004) and human systemic sclerosis patients (Dong et al. 2002). Thus it was expected that basal *SMAD7* mRNA expression would be lower in affected individuals, when compared to unaffected individuals, and that induction of *SMAD7* transcription in affected individuals would be depressed (Dong et al. 2002).

To test this hypothesis, basal and TGF-β1-inducible *SMAD7* mRNA expression was evaluated in skin fibroblasts of an affected and an unaffected family member by means of real-time qPCR. Analysis was performed using the $2^{-\Delta\Delta CT}$ method described previously (Livak and Schmittgen 2001), incorporating *GUSB* as an internal reference gene. Basal *SMAD7* levels were presented as the fold change in normalised expression relative to the expression of the transcript in the unaffected individual. Based on these results, it was concluded that *SMAD7* transcription did not differ significantly between individuals, in fibroblasts which had not been stimulated with exogenous TGF-β1.

For each individual, a statistically significant two-fold increase in *SMAD7* expression, relative to the basal level, was observed in response to treatment with recombinant TGF-β1 for 1 hour. However, when the TGF-β1-inducible response was compared between individuals, no significant difference was found. This indicates that the negative regulatory mechanism controlling profibrotic signalling in response to TGF-β1 is intact in affected individuals.

It must be noted, however, that the observed two-fold induction of *SMAD7* expression in response to stimulation with TGF-β1 differs substantially from previous results reported by similar studies. Dong and colleagues, whose 2002 investigation highlighted the role of depressed *SMAD7* expression in systemic sclerosis, found a 6.3- and 2.7-fold induction of *SMAD7* expression

after 1 hour of TGF-β1 treatment, in normal and systemic sclerosis cultured fibroblasts, respectively. Since both the affected and unaffected individual in this study show a markedly lower induction of expression, this result is unlikely to be related to the pathogenesis of the disease. It is possible that the difficulties encountered in cell culture described above may have influenced the cells' response to TGF-β1. Additionally, the small surface area of the dishes used to culture the fibroblasts limited the total cell count, leading to a decrease in quantity and quality of RNA after extraction and hence, to a decrease in quality and quantity of template cDNA (necessitating the use of a much higher template cDNA concentration for qPCR than the generally recommended 100ng/µl). A combination of these factors may therefore have resulted in relatively lower TGF-β1-inducible *SMAD7* expression levels in the family members investigated in this study. Nevertheless, the fact that TGF-β1 treatment consistently resulted in a 2-fold induction of expression between replicates and between individuals suggests that although these results differ from previous reports, reliable conclusions can nevertheless be drawn from the data.

Taken together with the results of the Smad2/3 expression and phosphorylation experiments, it appears that excessive Smad signalling is unlikely to account for the development of fibrosis observed in affected individuals. Although several components of the TGF-β-responsive signalling cascade (namely the TGF-β-receptors I and II, and the co-Smad, Smad4) were not investigated in this study, it was assumed that an investigation of R-Smad phosphorylation would serve as an indicator of aberrations in upstream genes/proteins resulting in excessive signalling.

It therefore seems reasonable to conclude that defects in the Smad pathway are not responsible for the disease described by Khumalo et al. (2006). This is in contrast to scleroderma, a related disorder described in 2.1.2.4, in which decreased basal and TGF-β1-inducible *SMAD7* expression, in combination with elevated Smad3 expression, results in unchecked activation of the TGF-β1 autocrine signalling loop, a factor identified as important in the pathogenesis of fibrotic disease (Dong et al. 2002). This underlines the novel

genetic basis of the disease under investigation, and highlights the need for a new approach to identify the molecular basis of this disease.

6. Conclusion

In this study, various components of the TGF-β/Smad signalling cascade were investigated at the transcriptional, translational and post-translational levels, to determine their role in the molecular basis of a novel, autosomal dominant fibrotic disease. These included *TGFB1*, Smad2, Smad3 and *SMAD7*.

Examination of the length and sequence of the *TGFB1* transcript, encoding the profibrotic cytokine TGF-β1, revealed no large-scale deletions, indicating correct post-transcriptional processing. Although two SNPs were identified during sequencing, both of which had previously been reported to affect the expression of TGF-β1, the high frequencies of these variants in the background population suggest that they do not account for the rare phenotype observed. Additionally, no difference was found when comparing *TGFB1* expression in the skin fibroblasts of an affected and an unaffected family member, indicating that aberrant expression and/or processing of *TGFB1* does not appear to account for the symptoms of the disease.

With respect to Smad2/3 expression and phosphorylation, no difference was observed between the cultured fibroblasts of affected and unaffected family members, or between family members and their matched controls, following treatment with TGF-β1. This indicated that the TGF-β1 signalling response was neither constitutively active, nor hyperresponsive, in the affected individual studied. Finally, basal and TGF-β1 inducible *SMAD7* mRNA expression was found not to differ between skin fibroblasts of the affected and unaffected family member, suggesting that the observed symptoms could not be attributed to prolonged activity of the Smad pathway in the absence of a negative regulator.

The combination of these results, obtained from different levels of regulation of the TGF-β1/Smad pathway, suggest that dysregulation of this signalling cascade is not responsible for the development of fibrosis in the novel syndrome; a significant finding, given its central role in the pathogenesis of

related disorders. Continued research is therefore required in order to identify the molecular basis of this disease.

7. Future Work

Further investigation into the molecular basis of this disease should follow one of two approaches, depending on the availability of biological material from the extended family. In the absence of additional blood and/or tissue samples, it seems prudent to continue the investigation using the candidate gene approach. Results of this study strongly suggest that the TGF-β1/Smad signalling pathway does not play a causal role in the development of fibrosis in affected family members. Thus, it seems logical to extend the search to include other likely candidates. These may include genes whose physiological function matches the proposed disease mechanism (that is, genes involved in the deposition, regulation of expression or degradation of extracellular matrix components), or those genes previously implicated in related disorders, such as Rothmund-Thomson syndrome.

Should biological material from additional affected family members become available, however, it may be feasible to employ whole-genome techniques, such as comparative genomic hybridisation (CGH) or expression arrays. Since these techniques enable the detection of sequence aberrations ranging from single nucleotide polymorphisms to large scale deletions across the entire genome, it is likely that mutations would be identified at a far more rapid rate using these methods, compared to the laborious, hypothesis-driven candidate gene approach.

In addition, future work should entail the confirmation and improvement of results obtained in this study. Due to the high degree of clinical heterogeneity observed in this disorder, it is recommended that the experiments be expanded to include the remaining affected individuals if possible, to confirm the absence of disease-causing mutations in the genes/proteins studied.

The Smad2/3 western blot results should be expanded to include a time-course experiment, in which expression of the R-Smads is monitored following increasing durations of TGF-β1 treatment, to determine the ability of the affected individual to downregulate these signalling molecules over time.

Additionally, immunohistochemical analysis should be performed on fibrotic lesions from the affected individual, if possible, to determine the subcellular localisation of the R- and Co-Smads before and after treatment with TGF-β1. Further work is also needed to improve the result of the anti-phosphoSmad3 western blot, in order to obtain an optimal signal:noise ratio, allowing the experiment to be repeated and analysed reliably. This may involve the alteration of western blot parameters (such as antibody concentrations, wash times and volumes), as well as optimisation of cell culture conditions.

However, since no pathogenic defect has been identified in the TGF-β1/Smad cascade thus far, research in the future is likely to extend beyond this pathway, to the identification of new candidates, or to the whole-genome techniques described above. Additionally, a larger-scale morphological approach may be taken, involving the examination of cultured fibroblast morphology and behaviour using fluorescent and/or electron microscopy. This is likely to reveal details such as rate of growth, cell mortality and organelle morphology, all of which will shed light on the cellular effects of this disease, and aid in the identification of candidate genes for study.

Finally, it has been postulated that the syndromic nature of this disease, characterised by a severe systemic phenotype, points to the possible involvement of DNA instability, a phenomenon observed in other related disorders, such as RTS. This may also be investigated, using cytogenetic techniques.

It is hoped that the recruitment of additional family members to the study will enable the use of more sophisticated methods of mutation screening. Until then, however, it is likely that the future of research into the molecular basis of this disease will be dictated by the limitations of the candidate gene approach.

8. References

Annes JP, Munger JS, Rifkin DB (2003) Making sense of latent TGF-β activation. J Cell Sci 116:217-224 doi:10.1242/jcs.00229

Arai T, Akiyama Y, Okabe S, Ando M, Endo M, Yuasa Y (1998) Genomic structure of the human *Smad3* gene and its infrequent alterations in colorectal cancers. Cancer Lett 122:157-163

Arany PR, Flanders KC, Kobayashi T, Kuo CK, C, Desai KC, Tuan R, Rennard SI, Roberts AB (2006) Smad3 deficiency alters key structural elements of the extracellular matrix and mechanotransduction of wound closure. Proc Natl Acad Sci 103:9250-9255

Awad MR, El-Gamel A, Hasleton P, Turner DM, Sinnott PJ, Hutchinson IV (1998) Genotypic variation in the transforming growth factor-[beta]1 gene: association with transforming growth factor-[beta]1 production, fibrotic lung disease, and graft fibrosis after lung transplantation. Transplantation 66:1014-1020

Bartram U, Speer CP (2004) The role of Transforming Growth Factor β in lung development and disease. Chest 125:754-765

Blobe GC, Schiemann WP, Lodish HF (2000) Role of transforming growth factor β in human disease. N Engl J Med 342:1350-1358

Border WA, Noble NA (1994) Transforming Growth Factor β in Tissue Fibrosis. N Engl J Med 331:1286-1292

Broekelmann TJ, Limper AH, Colby TV, McDonald JA (1991) Transforming growth factor β1 is present at sites of extracellular matrix gene expression in human pulmonary fibrosis. Proc Natl Acad Sci USA 88:6642-6646

Cambien F, Ricard S, Troesch A, Mallet C, Generenaz L, Evans A, Arveiler D, Luc G, Ruidavets J-B, Poirier O (1996) Polymorphisms of the Transforming Growth Factor-β1 gene in relation to myocardial infarction and blood pressure. The Etude Cas-Temoin de l'Infarctus du Myocarde (ECTIM) study. Hypertension 28:881-887

Carturan S, Roccatello D, Menegatti E, Di Simone D, Davit A, Piazza A, Sena LM, Matullo G (2004) Association between transforming growth factor beta1 gene polymorphisms and IgA nephropathy. J Nephrol 17: 786-793

Chacko BM, Qin B, Correia JJ, Lam SS, de Caestecker MP, Lin K (2001) The L3 loop and C-terminal phosphorylation jointly define Smad protein trimerization. Nat Struct Biol 8:248-253

Chen C-R, Kang Y, Siegel PM, Massagué J (2002) E2F4/5 and p107 as Smad cofactors linking the TGF-β receptor to c-myc repression. Cell 110:19-32

Clouthier DE, Comerford SA, Hammer RE (1997) Hepatic fibrosis, glomerulosclerosis, and lipodystrophy-like PEPCK-TGF-β-1 transgenic mice. J Clin Invest 100:2697-2713

Crilly A, Hamilton J, Clark CJ, Jardine A, Madhok R (2002) Analysis of transforming growth factor β1 gene polymorphisms in patients with systemic sclerosis. Ann Rheum Dis 61:678-681 DOI 10.1136/ard.61.8.678

Dong C, Li Z, Alvarez, R Jr., Feng X-H, Goldschmidt-Clermont PJ (2000) Microtubule Binding to Smads May Regulate TGF-β Activity. Mol Cell 5:27–34

Dong C, Zhu S, Wang T, Yoon W, Li Z, Alvarez RJ, ten Dijke P, White B, Wigley FM, Goldschmidt-Clermont PJ (2002) Deficient Smad7 expression: a putative molecular defect in scleroderma. Proc Natl Acad Sci 99:3908-3913 DOI 10.1073/pnas.062010399

Dunning AM, Ellis PD, McBride S, Kirschenlohr HL, Healey CS, Kemp PR, Luben RN, Chang-Claude J, Mannermaa A, Kataja V, Pharoah PDP, Easton DF, Ponder BAJ, Metcalfe JC (2003) Transforming Growth Factor β1 signal peptide variant increases secretion *in vitro* and is associated with increased incidence of invasive breast cancer. Cancer Res 63:2610–2615A

França LTC, Carrilho E, Kist TBL (2002) A review of DNA sequencing techniques. Q Rev Biophys 35:169–200 DOI 10.1017/S0033583502003797

Glenn ST, Jones CA, Liang P, Kaushik D, Gross KW, Kim HL (2007) Expression profiling of archival renal tumors by quantitative PCR to validate prognostic markers. BioTechniques 43:639-647

Grainger DJ, Heathcote K, Chiano M, Snieder H, Kemp PR, Metcalfe JC, Carter ND, Spector TD (1999) Genetic control of the circulating concentration of transforming growth factor type β1. Hum Mol Genet 8:93-97

Grau Salvat C, Pont V, Cors JR, Aliaga A (1999) Hereditary sclerosing poikiloderma of Weary: report of a new case. Brit J Dermatol 140:358–377

Hartmann LM (2006) Investigation of the role of the Transforming Growth Factor β1 gene in a novel familial fibrosing syndrome. Honours Thesis submitted to the Division of Human Genetics, University of Cape Town

Heldin C-H, Miyazono K, ten Dijke P (1997) TGF-β signalling from cell membrane to nucleus through SMAD proteins. Nature 390:465-471

Janknecht R, Wells NJ, Hunter T (1998) TGF-β-stimulated cooperation of Smad proteins with the coactivators CBP/p300. Genes Dev 12:2114–2119

Janssens K, Gershoni-Baruch R, Guañabens N, Migone N, Ralston S, Bonduelle M, Lissens W, Van Maldergem L, Vanhoenacker F, Verbruggen L, Van Hul W (2000) Mutations in the gene encoding the latency-associated peptide of TGF-β1 cause Camurati-Engelmann disease. Nature Genet 26:273-275

Jimenez SA, Derk CT (2004) Following the molecular pathways toward an understanding of the pathogenesis of systemic sclerosis. Ann Intern Med 140:37-50

Johnson K, Kirkpatrick H, Comer A, Hoffmann FM, Laughon A (1999) Interaction of Smad complexes with tripartite DNA-binding sites. J Biol Chem 274: 20709–20716

Kavsak P, Rasmussen RK, Causing CG, Bonni S, Zhu H, Thomsen GH, Wrana JL (2000) Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF-β receptor for degradation. Molec Cell 6:1365-1375

Khalil N, O'Connor RN, Flanders KC, Unruh H (1996) TGF-beta 1, but not TGF-beta 2 or TGF-beta 3, is differentially present in epithelial cells of advanced pulmonary fibrosis: an immunohistochemical study. Am J Respir Cell Mol Biol 14:131-138

Khumalo NP, Pillay K, Beighton P, Wainwright H, Walker B, Saxe N, Mayosi BM, Bateman ED (2006) Poikiloderma, tendon contracture, and pulmonary fibrosis: a new autosomal dominant syndrome? Brit J Dermatol. 155:1057-1061

Kim S-J, Glick A, Sporn MB, Roberts AB (1989) Characterization of the promoter region of the human transforming growth factor-β1 gene. J Biol Chem 264: 402-408

Kitao S, Shimamoto A, Goto M, Miller RW, Smithson WA, Lindor NM, Furuichi Y (1999) Mutations in *RECQL4* cause a subset of cases of Rothmund-Thomson syndrome. Nature Genet 22:82-84

Kloos DU, Choi C, Wingender E (2002) The TGF-β-Smad network: introducing bioinformatic tools. Trends Genet 18:96-103

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genet 22:139-144

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685 doi:10.1038/227680a0

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genet 11:241-247

Lee JW, Chen Q, Knowles DM, Cesarman E, Wang YL (2006) β - Glucuronidase is an optimal normalization control gene for molecular monitoring of chronic myelogenous leukemia. J Mol Diagn 8:385–389

Lindor NM, Furuichi Y, Kitao S, Shimamoto A, Arndt C, Jalal S (2000) Rothmund-Thomson Syndrome due to *RECQ4* helicase mutations: report and clinical and molecular comparisons with Bloom Syndrome and Werner Syndrome. Am J Med Genet 90:223–228

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 25:402–408

Lo RS, Chen Y-G, Shi Y, Pavletich NP, Massagué J (1998) The L3 loop: a structural motif determining specific interactions between SMAD proteins and TGF-β receptors. EMBO J 17:996-1005

Lo RS, Massagué J (1999) Ubiquitin-dependent degradation of TGF-β-activated Smad2. Nat Cell Biol 1:472-478

Massagué J (1998) TGF-β signal transduction. Annu Rev Biochem 67:753–791

Mattey DL, Nixon N, Dawes PT, Kerr J (2005) Association of polymorphism in the transforming growth factor β1 gene with disease outcome and mortality in rheumatoid arthritis. Ann Rheum Dis 64:1190-1194 DOI 10.1136/ard.2004.031674

McCarthy MI, Smedley D, Hide W (2003) New methods for finding disease-susceptibility genes: impact and potential. Genome Biol 4:119.1-119.8

Mori Y, Chen S-J, Varga J (2000) Modulation of endogenous Smad expression in normal skin fibroblasts by Transforming Growth Factor-β. Exp Cell Res 258:374–383

Mori Y, Chen S-J, Varga J (2003) Expression and regulation of intracellular SMAD signaling in scleroderma skin fibroblasts. Arthritis Rheum 48:1964–1978

Munger JS, Huang X, Kawakatsu H, Griffiths MJD, Dalton SL, Wu J, Pittet J-F, Kaminski N, Garat C, Matthay MA, Rifkin DB, Sheppard D (1999) The integrin αvβ6 binds and activates latent TGF β1: a mechanism for regulating pulmonary inflammation and fibrosis. Cell 96:319-328

Nakao A, Afrakhte M, Morén A, Nakayama T, Christian JL, Heuchel R, Itoh S, Kawabata M, Heldin N-E, Heldin C-H, ten Dijke P (1997a) Identification of Smad7, a TGFβ-inducible antagonist of TGF-β signaling. Nature 389:631-635

Nakao A, Imamura T, Souchelnytskyi S, Kawabata M, Ishisaki A, Oeda E, Tamaki K, Hanai J-I, Heldin C-H, Miyazono K, ten Dijke P (1997b) TGF-β

receptor-mediated signalling through Smad2, Smad3 and Smad4. EMBO J 16:5353–5362

Nakao A, Okumura K, Ogawa H (2002) Smad7: a new key player in TGF-β-associated disease. Trends Mol Med 8:361-363

Ott J (1991) Analysis of human genetic linkage, Revised edition. The Johns Hopkins University Press, Baltimore Maryland USA

Pesce K, Rothe MJ (1996) The premature aging syndromes. Clin Dermatol 14:161-170

QuantiFast SYBR Green PCR Handbook 01/2007, QIAGEN

Roche Applied Science LightCycler® DNA Master SYBR Green I Handbook, Roche Diagnostics, Mannheim, Germany

Ross, S, Hill CS (2008) How the Smads regulate transcription. Int J Biochem Cell Biol 40: 383–408 doi:10.1016/j.biocel.2007.09.006

Saito E, Fujimoto M, Hasegawa M, Komura K, Hamaguchi Y, Kaburagi Y, Nagaoka T, Takehara K, Tedder TF, Sato S (2002) CD19-dependent B lymphocyte signaling thresholds influence skin fibrosis and autoimmunity in the tight-skin mouse. J Clin Invest 109:1453-1462

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: A laboratory manual, Second edition. Cold Spring Harbour Laboratory Press, Cold Spring Harbour New York USA

Sasaki N, Izawa M, Watahiki M, Ozawa K, Tanaka T, Yoneda Y, Matsuura S, Carninci P, Muramatsu M, Okazaki Y, Hayashizaki Y (1998) Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc Natl Acad Sci USA 95:3455–3460

Shah R, Rahaman B, Hurley CK, Posch PE (2006) Allelic diversity in the *TGFB1* regulatory region: characterization of novel functional single nucleotide polymorphisms. Hum Genet 119:61-74 DOI 10.1007/s00439-005-0112-y

Shi S, Wang Y-F, Jayaraman L, Yang H, Massague J, Pavletich NP (1998) Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-β signaling. Cell 94:585-594

Shin A, Shu X-O, Cai C Gao Y-T, Zheng W (2005) Genetic polymorphisms of the Transforming Growth Factor-β1 gene and breast cancer risk: a possible dual role at different cancer stages. Cancer Epidemiol Biomarkers Prev 14:1567-1570

Silverman ES, Palmer LJ, Subramaniam V, Hallock A, Mathew S, Vallone J, Faffe DS, Shikanai T, Raby BA, Weiss ST, Shore SA (2004) Transforming

Growth Factor-β1 promoter polymorphism C-509T is associated with asthma. Am J Respir Crit Care Med 169: 214–219, 2004 DOI 10.1164/rccm.200307-973OC

Sommer S, Goodfield MJD (2002) Connective tissue disease and the skin. Clin Med 2:9-14

Steen VD (2005) The lung in Systemic Sclerosis. J Clin Rheumatol 11:40-46 DOI 10.1097/01.rhu.0000152147.38706.db

Stoll C, Mengsteab S, Stoll D, Riediger D, Gressner AM, Weiskirchen R (2004) Analysis of polymorphic *TGFB1* codons 10, 25, and 263 in a German patient group with non-syndromic cleft lip, alveolus, and palate compared with healthy adults. BMC Med Genet 5:15

Suthanthiran M, Li B, Song JO, Ding R, Sharma VK, Schwartz JE, August P (2000) Transforming growth factor-β1 hyperexpression in African-American hypertensives: A novel mediator of hypertension and/or target organ damage. Proc Natl Acad Sci USA 97: 3479–3484

Takagawa S, Lakos G, Mori Y, Yamamoto T, Nishioka K, Varga J (2003) Sustained activation of fibroblast Transforming Growth Factor-β/Smad signaling in a murine model of Scleroderma. J Invest Dermatol 121:41-50

Ten Dijke P, Miyazono K, Heldin C-H (2000) Signaling inputs converge on nuclear effectors in TGF-β signaling. TIBS 25:64-70

Terwilliger JD, Weiss KM (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr Opin Biotechnol 9:578-594

Tesseur I, Zou K, Berber E, Zhang H, Wyss-Coray T (2006) Highly sensitive and specific bioassay for measuring bioactive TGF-β. BMC Cell Biol 7:15 DOI 10.1186/1471-2121-7-15

Tsukazaki T, Chiang TA, Davison AF, Attisano L, Wrana JL (1998) SARA, a FYVE domain protein that recruits Smad2 to the TGFβ Receptor. Cell 95:779–791

Van Laar JM, Huizinga TWJ (2005) Tweaking microtubules to treat scleroderma. PLoS Med 2:1230-1231

Varga J (2002) Dysfunctional Smad family dynamics culminating in fibrosis. Arthritis Rheum 46: 1703–1713

Venkatesan N, Pini L, Ludwig MS (2004) Changes in Smad expression and subcellular localization in bleomycin-induced pulmonary fibrosis. Am J Physiol Lung Cell Mol Physiol 287:1342-1347

Verrecchia F, Chu M-L, Mauviel A (2001) Identification of novel TGF-β/Smad gene targets in dermal fibroblasts using a combined cDNA microarray/promoter transactivation approach. J Biol Chem 276: 17058–17062

Verrecchia F, Mauviel A (2007) Transforming growth factor-β and fibrosis. World J Gastroenterol 13: 3056-3062

Wang LL, Levy ML, Lewis RA, Chintagumpala MM, Lev D, Rogers M, Plon SE (2001) Clinical manifestations in a cohort of 41 Rothmund-Thomson syndrome patients. Am J Hum Gen 102:11-17

Weary PE, Hsu YT, Richardson DR, Caravati CM, Wood BT. (1969) Hereditary sclerosing poikiloderma. Arch Dermatol 100:413–422

White B (2003) Interstitial lung disease in scleroderma. Rheum Dis Clin N Am 29:371–390 DOI 10.1016/S0889-857X(03)00025-5

Whitley E, Ball J (2002) Statistics review 4: Sample size calculations. Crit Care 6:335-341

Xaubet A, Marin-Arguedas A, Lario S, Ancochea J, Morell F, Ruiz-Manzano J, Rodriguez-Becerra E, Rodriguez-Arias JM, Iñigo P, Sanz S, Campistol JP, Mullol J, Picado C (2003) Transforming Growth Factor-β1 gene polymorphisms are associated with disease progression in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med 168:431–435 DOI: 10.1164/rccm.200210-1165OC

Yamada Y, Miyauchi A, Goto J, Takagl Y, Okuizumi H, Kanematsu M, Hase M, Takai H, Harada A, Ikeda K (1998) Association of a polymorphism of the Transforming Growth Factor-β1 gene with genetic susceptibility to osteoporosis in postmenopausal Japanese women. J Bone Miner Res 13: 1569-1576

Zawel L, Dai J, Buckhaults P, Zhou S, Kinzler KW, Vogelstein B, Kern SE (1998) Human Smad3 and Smad4 are sequence-specific transcription activators. Mol Cell, 1:611–617

Zhang Y, Feng X-H, Wu R-Y, Derynck R (1996) Receptor-associated Mad homologues synergize as effectors of the TGF-β response. Nature 383:168-172

Zhu Z, Homer RJ, Wang Z, Chen Q, Geba GP, Wang J, Zhang Y, Elias JA (1999) Pulmonary expression of interleukin-13 causes inflammation, mucus hypersecretion, subepithelial fibrosis, physiologic abnormalities, and eotaxin production. J Clin Invest 103: 779-788

Zhu Q, Pearson-White S, Luo K (2005) Requirement for the SnoN oncoprotein in Transforming Growth Factor β-induced oncogenic transformation of fibroblast cells. Mol Cell Biol 25:10731–10744

9. Appendices

Appendix A: Buffers and Reagents for Agarose Gel Electrophoresis

10x TBE

108g Tris (USB Corporation, Cleveland, USA)

55g Boric Acid (Promega)

7.4g EDTA (Merck, Darmstadt, Germany)

Made up to final volume of 1 litre using distilled water and autoclaved.

Diluted to working stock of 1x

1% Agarose incorporating 0.5µg/ml Ethidium Bromide

1g Agarose LE (Roche Diagnostics)

100ml 1X TBE

5µl 10mg/ml Ethidium Bromide (SIGMA-ALDRICH)

Loading Buffer

0.125g (0.25% v/v) Bromophenol Blue (Merck)

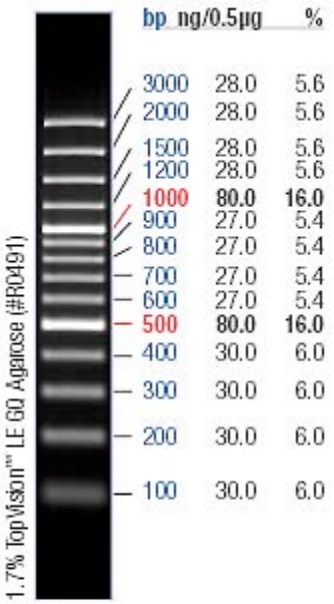
20g (40% v/v) Sucrose (BDH Laboratory Supplies, Poole, UK)

Made up to 50ml using distilled water

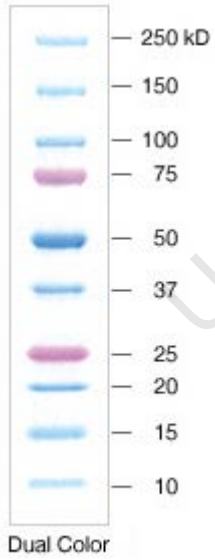
pH>8

Appendix B: Molecular Weight Markers

B.1 Gene Ruler™ 100bp DNA Ladder Plus *Improved* (Fermentas)



B.2 Precision Plus Protein Standard Dual Colour (BIO-RAD)



Appendix C: Annotated *TGFB1* mRNA Sequence Indicating Position of Sequencing Primers

```
   1 ccttcgcgcc ctgggccatc tccctcccac ctccctccgc ggagcagcca gacagcgagg
  61 gccccggccg ggggcagggg ggacgccccg tccggggcac ccccccggct ctgagccgcc
 121 cgcggggccg gcctcggccc ggagcggagg aaggagtcgc cgaggagcag cctgaggccc
 181 cagagtctga gacgagccgc cgccgccccc gccactgcgg ggaggagggg gaggaggagc
 241 gggaggaggg acgagctggt cgggagaaga ggaaaaaaac ttttgagact tttccgttgc
 301 cgctgggagc cggaggcgcg gggacctctt ggcgcgacgc tgccccgcga ggaggcagga
 361 cttggggacc ccagaccgcc tccctttgcc gccggggacg cttgctccct ccctgccccc
 421 tacacggcgt ccctcaggcg cccccattcc ggaccagccc tcgggagtcg ccgacccggc___F
 481 ctcccgcaaa gacttttccc cagacctcgg gcgcaccccc tgcacgccgc cttcatcccc
 541 ggcctgtctc ctgagccccc gcgcatccta gacccttttct cctccaggag acggatctct
 601 ctccgacctg ccacagatcc cctattcaag accacccacc ttctggtacc agatcgcgcc
 661 catctaggtt atttccgtgg gatactgaga caccccggt ccaagcctcc cctccaccac
 721 tgcgcccttc tccctgagga cctcagcttt ccctcgaggc cctcctacct tttgccggga
 781 gacccccagc ccctgcaggg gcggggcctc cccaccacac cagccctgtt cgcgctctcg___F*
 841 gcagtgccgg ggggcgccgc ctcccccatg ccgccctccg ggctgcggct gctgccgctg
 901 ctgctaccgc tgctgtggct actggtgctg acgcctggcc ggccggccgc gggactatcc
 961 acctgcaaga ctatcgacat ggagctggtg aagcggaagc gcatcgaggc catccgcggc
1021 cagatcctgt ccaagctgcg gctcgccagc cccccgagcc aggggaggt gccgcccggc
1081 ccgctgcccg aggccgtgct cgccctgtac aacagcaccc gcgaccgggt ggccggggag
1141 agtgcagaac cggagcccga gcctgaggcc gactactacg ccaaggaggt cacccgcgtg
1201 ctaatggtgg aaacccacaa cgaaatctat gacaagttca agcagagtac acacagcata
1261 tatatgttct tcaacacatc agagctccga gaagcggtac ctgaacccgt gttgctctcc
1321 cgggcagagc tgcgtctgct gaggctcaag ttaaaagtgg agcagcacgt ggagctgtac
1381 cagaaataca gcaacaattc ctggcgatac ctcagcaacc ggctgctggc acccagcgac
1441 tcgccagagt ggttatcttt tgatgtcacc ggagttgtgc ggcagtggtt gagccgtgga
1501 gggaaattg agggctttcg ccttagcgcc cactgctcct gtgacagcag ggataacaca
1561 ctgcaagtgg acatcaacgg gttcactacc ggccgccgag gtgacctggc caccattcat
1621 ggcatgaacc ggcctttcct gcttctcatg gccaccccgc tggagagggc ccagcatctg
1681 caaagctccc ggcaccgccg agccctggac accaactatt gcttcagctc cacggagaag
1741 aactgctgcg tgcggcagct gtacattgac ttccgcaagg acctcggctg gaagtggatc___R*
1801 cacgagccca agggctacca tgccaacttc tgcctcgggc cctgccccta catttggagc
1861 ctggacacgc agtacagcaa ggtcctggcc ctgtacaacc agcataaccc gggcgcctcg
1921 gcggcgccgt gctgcgtgcc gcaggcgctg gagccgctgc ccatcgtgta ctacgtgggc
1981 cgcaagccca aggtggagca gctgtccaac atgatcgtgc gctcctgcaa gtgcagctga___R
2041 ggtcccgccc cgccccgccc cgccccggca ggcccggccc caccccgccc cgcccccgct
2101 gccttgccca tgggggctgt atttaaggac acccgtgccc caagcccacc tggggcccca
2161 ttaaagatgg agagaggact gcggatctct gtgtcattgg gcgcctgcct ggggtctcca
2221 tccctgacgt tcccccactc ccactccctc tctctccctc tctgcctcct cctgcctgtc
2281 tgcactattc ctttgcccgg catcaaggca caggggacca gtggggaaca ctactgtagt
2341 tagatctatt tattgagcac cttgggcact gttgaagtgc cttacattaa tgaactcatt
     cagtcaccat agcaacactc tgagatgcag ggactctgat aacacccatt ttaaaggtga
     ggaaacaagc ccagagaggt taagggagga gttcctgccc accaggaacc tgctttagtg
     ggggatagtg aagaagacaa taaaagatag tagttcaggc caggcggggt ggctcacgcc
      tgt
```

Key: Exon 1 is indicated in pink, Exon 2 in red, Exon 3 in light green, Exon 4 in blue, Exon 5 in orange, Exon 6 in dark green and Exon 7 in grey. Primers are underlined (F, whole transcript forward primer, F*, internal fragment forward primer, R, whole transcript reverse primer and R*, internal fragment reverse primer). The translation start and stop codons are indicated in bold, while the 3' untranslated region is designated by italics. Numbers correspond to positions in the mRNA sequence obtained from the NCBI database.

Appendix D: Manufacturer's Protocols

D.1 High Pure RNA Isolation Kit (Roche)

Isolation of Total RNA from Cultured Cells (suited for 1 × 10^6 cells)

1. Resuspend cells in 200 µl PBS.
2. Add 400 µl Lysis/-Binding Buffer (green cap) and vortex for 15 s.
3. To transfer the sample to a High Pure Filter Tube:
   - Insert one High Filter Tube in one Collection Tube.
   - Pipet entire sample into the upper reservoir of the Filter Tube (max. 700µl)
4. - Insert the entire High Pure Filter Tube assembly into a standard tabletop centrifuge.
   - Centrifuge the tube assembly 15 s at 8,000 × *g*.
5. After centrifugation:
   - Remove the Filter Tube from the Collection Tube, discard the flowthrough liquid, and again combine the Filter Tube and the used Collection Tube.
6. After re-inserting the Filter Tube:
   - Pipette per sample 90 µl DNase Incubation Buffer (white cap) into a sterile reaction tube, add 10 µl DNase I, mix and pipette the solution on the glass filter fleece in the upper reservoir of the filter tube.
   - Incubate for 15 min at +15 to +25°C.
7. - Add 500 µl Wash Buffer I to the upper reservoir of the Filter Tube assembly and centrifuge 15 s at 8,000 × *g*.
   - Discard flowthrough and combine Filter Tube with the used Collection Tube.
8. - Add 500 µl Wash Buffer II (blue cap) to the upper reservoir of the Filter Tube assembly and centrifuge 15 s at 8,000 × *g*.
   - Discard flowthrough and combine Filter Tube with the used Collection Tube.
9. Add 200µl Wash Buffer II (blue cap) to the upper reservoir of the Filter Tube assembly and centrifuge for 2 min at maximum speed (approx. 13,000 × *g*) to remove any residual Wash Buffer.

   * The extra centrifugation time ensures removal of residual Wash Buffer

10. Discard the Collection Tube and insert the Filter Tube into a clean, sterile 1.5 ml microcentrifuge tube.
11. To elute the RNA:
    • Add 50 – 100 µl Elution Buffer to the upper reservoir of the Filter Tube.
    • Centrifuge the tube assembly for 1 min at 8,000 × *g.*
12. The microcentrifuge tube now contains the eluted RNA. Either use the eluted RNA directly in RT-PCR or store the eluted RNA at –80°C for later analysis

D.2 QIAquick® Gel Extraction Kit (QIAGEN)

1. Excise the DNA fragment from the agarose gel with a clean, sharp scalpel.
2. Weigh the gel slice in a colourless tube. Add 3 volumes of buffer QG to 1 volume of gel (100mg ≈ 100µl)
3. Incubate at 50ºC for 10min (or until the gel slice has completely dissolved). To help dissolve gel, mix by vortexing the tube every 2-3min during the incubation,
4. After the gel slice has dissolved completely, check the colour of the mixture is yellow (similar to Buffer QG without dissolved agarose).
5. Add 1 gel volume of isopropanol to the sample and mix.
6. Place a QIAquick spin column in a provided 2ml collection tube.
7. To bind DNA, apply the sample to the QIAquick column, and centrifuge for 1min.
8. Discard flow through and place QIAquick column back in the same collection tube.
9. Add 0.5ml of Buffer QG to QIAquick column and centrifuge for 1 min.
10. To wash, add 0.75ml of Buffer PE to QIAquick column and centrifuge for 1 min.
11. Discard the flow through and centrifuge the QIAquick column for an additional 1 min at 13000rpm.
12. Place the QIAquick column into a clean 1.5ml microcentrifuge tube.
13. To elute DNA, add 30ml dH_2O to the centre of the QIAquick membrane and centrifuge the column for 1min.

D.3 Transcriptor First Strand cDNA Synthesis Kit (Roche)

Procedure A: cDNA Synthesis with anchored oligo(dT)18 primer

1. • Thaw all frozen reagents before use.
   • Briefly centrifuge them before starting the procedure.
   • Keep all reagents on ice while setting up the reactions.
2. In a sterile, nuclease-free, thin-walled PCR tube on ice, prepare the template-primer mixture for one 20 µl reaction by adding the components in the order listed below.

   * Always wear gloves when handling RNA.

Template-Primer Mix (for 1 reaction)

Component	Vol.	Final conc.
total RNA or poly(A)$^+$ mRNA		1 µg total RNA or 10 ng poly(A)$^+$ RNA[a)]
Primer – choose either:		
Anchored-oligo(dT)$_{18}$ Primer, 50 pmol/µl (vial 5)	1 µl	2.5 µM
OR Random Hexamer Primer, 600 pmol/µl (vial 6)	2 µl	60 µM
OR Sequence-Specific Primer	variable	0.5-2.5 µM
Water, PCR-grade (vials 7 or 9)	variable	to make total volume = 13 µl
Total volume	**13 µl**	

a) These are the suggested concentrations for initial experiments. Suitable template concentrations may range from 10 ng to 5 µg total RNA and from 1 to 100 ng mRNA.

3. Optional Step:

   • Denature the template-primer mixture by heating the tube for 10 min at 65°C in a thermal block cycler with a heated lid (to minimize evaporation). This step ensures denaturation of RNA secondary structures.
   • Immediately cool the tube on ice.
4. To the tube containing the template-primer mix, add the remaining components of the RT mix in the order listed below.

Component	Vol.	Final conc.
Transcriptor Reverse Transcriptase Reaction Buffer, 5× conc. (vial 2)	4 µl	1× (8 mM $MgCl_2$)
Protector RNase Inhibitor, 40 U/µl (vial 3)	0.5 µl	20 U
Deoxynucleotide Mix, 10 mM each (vial 4)	2 µl	1 mM each
Transcriptor Reverse Transcriptase, 20 U/µl (vial 1)	0.5 µl	10 U
Final volume	**20 µl**	

5. • Mix the reagents in the tube carefully.

   * Do not vortex!

   • Centrifuge the tube briefly to collect the sample on the bottom of the tube.

   • Place the tube in a thermal block cycler with a heated lid (to minimize evaporation).

6. Depending on the primer used and the length of the target mRNA, incubate the RT reaction as described in the table below:

Depending on the primer used and the length of the target mRNA, incubate the RT reaction as described in the table below:

If you are using…	And the target mRNA is…	Incubate the RT reaction……
Anchored-oligo$(dT)_{18}$ primer, 50 pmol/µl **OR** Sequence–specific primer	Up to 4 kb	30 min at 55°C
	>4 kb	60 min at 50°C
Random hexamer primers, 600 pmol/µl	Up to 4 kb	10 min at 25°C, followed by 30 min at 55°C
	>4 kb	10 min at 25°C, followed by 60 min at 50°C

7. • Inactivate Transcriptor Reverse Transcriptase by heating to 85° for 5min.

   • Stop the reaction by placing the tube on ice.

   • At this point the reaction tube may be stored at +2 to +8°C for 1-2 hrs or at -15 to -25°C for longer periods.

Appendix E: Buffers and Reagents for SDS-PAGE and Western Blot

7.5% Resolving Gel (Laemmli Resolving Gel 29:1 or 37:5:1 Ratio)

2.5ml 1.5M Tris-HCl pH 8.8

2.5ml 30% Acrylamide/Bisacrylamide (SIGMA-ALDRICH)

4.85ml Deionised Water

0.1ml 10% SDS

50µl 10% Ammonium Persulfate

5µl TEMED (BIO-RAD)

4% Stacking Gel

0.66ml 30% Acrylamide/Bisacrylamide (SIGMA-ALDRICH)

1.26ml 0.5M Tris-HCl pH 6.8

3ml Deionised Water

50µl 10% SDS

25µl 10% Ammonium Persulfate

5µl TEMED (BIO-RAD)

10x Electrophoresis Running Buffer

30.28g Tris Base

144.13g Glycine

10g SDS

Dissolve in 1l deionised water.

For electrophoresis, dilute to 1x in deionised water before use.

10x Transfer Buffer

30.28g Tris-HCl

144.13g Glycine

Dissolve in 800ml deionised water.

Before use, dilute 80ml of stock in 720ml deionised water and add 200ml Methanol, for 1x Transfer Buffer.

1.5M Tris-HCl pH8.8

28.5g Tris-HCl

69.05g Tris-Base

Dissolve in 500ml deionised water

0.5M Tris-HCl pH 6.8

38.15g Tris-HCl

0.95g Tris-Base

Dissolve in 500ml deionised water

1M Tris-HCl pH 6.8

74.86g Tris-HCl

3.03g Tris-Base

Dissolve in 500ml deionised water

Tris Buffered Saline + 0.1% Tween 20 (TBST)

6.8g Tris-HCl

0.82g Tris-Base

8.77g NaCl

Dissolve in deionised water, then add

5ml 20% Tween 20

Make up to 1l with deionised water.

5x Laemmli Sample Buffer (Laemmli, 1970)

1.25ml 1M Tris-HCl pH 6.8

1g SDS

0.33ml 1% Bromophenol Blue

5ml Glycerol

2.5ml 2-Mercaptoethanol

1.25ml Deionised Water

Stripping Buffer

1.56ml	1M Tris-HCl pH 6.8
5ml	10% SDS
0.175µl	2-Mercaptoethanol

Make up to final volume of 25ml using Deionised Water.

Lysis Buffer

50mM	HEPES pH 7.5 (SIGMA-ALDRICH)
1%	NP-40
10%	Glycerol
50mM	NaCl
10mM	NaF
1 tablet	Complete Mini Protease Inhibitor Cocktail (Roche)
1 tablet	Phos Stop Phosphatase Inhibitor Cocktail (Roche)