A Pilot Study To Identify Links Between Genetic Variation And Shoulder Pain And Dysfunction After Breast Cancer Radiotherapy

By

Callum Mclarty

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

In fulfilment of the requirements for the degree:

MSc Physiology

DEPARTMENT OF HUMAN BIOLOGY

FACULTY OF HEALTH SCIENCE

UNIVERSITY OF CAPE TOWN

Supervisors

Professor D Shamley[1]

Professor AV September[2]

1. Clinical Research Centre, University of Cape Town, Old Main Building, L51, Groote Schuur Hospital, Observatory.

2. Division of Exercise Science and Sports Medicine, Department of Human Biology, Faculty of Health Sciences, University of Cape Town.

Declaration

I, Callum Mclarty, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this, or any other university.

I authorise the university to reproduce, for the purpose of research, either the whole or any portion of the contents of this dissertation in any manner whatsoever.

Signature:……Signed by candidate…….

Date:……Thursday, 19 December 2019……

Table of Contents

Acknowledgements

I would like to thank the University of Cape Town for providing the facilities and environment this project was undertaken it. I would like to thank the city of Cape Town, and South Africa for hosting me.

The South African Medical Research Council (MRC) is also acknowledged for their financial contribution to the UCT/MRC Research Unit of Exercise Science and Sports Medicine (ESSM).

Thank you also to the Groote Schuur Hospital, its staff and all the participants for donating their time and DNA, which ultimately is the only reason this study was possible.

I would offer my strongest appreciation to my co-supervisors, in alphabetical order; Professor Alison September and Professor Delva Shamley. Throughout the project they have been incredibly supportive and incredibly patient, even when I'm sure that it wasn't easy. The time spent on this programme has been an exceptional learning experience, and I feel like a completely different person than I was at the start, for which I owe them both a great deal.

Within ESSM I would particularly like to thank the research team and fellow students, in no particular order: Shameemah Abrahams, Marilize Burger, Masouda Rahim, Andrea Gibbon, David Leith, Daneil Feldmann, Kathryn Van Boom, Sharon Mkumbuzi, Tamzyn Murphy, Kyle Willard, Nicole Ellman, Siddiqah George, Wendy Holidsay, Samantha Beckley, Matthew Leppan, and everyone else. Despite such a huge range of personalities, backgrounds and nationalities, it's the people that make ESSM such a wonderful team endeavour. Thank you for tips and guidance from everything from lab protocol and statistics; to where you go to pay speeding tickets.

My teammates, Firzana and Trevor: Thanks for being extremely supportive role models within the team. I wish you all the best of luck with your future endeavours.

Trevino and Neezaam for their limitless help in the lab, even when I’m sure some of the questions weren’t the smartest.

I also want to thank all the staff and students at the ESSM Research Unit and Sports Science Institute of South Africa. It was humbling to be sharing a building with such an inspiring and dedicated set of people and athletes. A side thank you to the staff, trainers and biokineticists, for teaching me the basics of nutrition, banting, and the worthwhile pursuit of at least some level of fitness.

And last but not least, my family and friends, for the unyielding support regardless of the circumstances. I owe them a tremendous amount, both financially and otherwise.

Abbreviations

ANOVA	two-way analysis of variance test
AT	ataxia telangiectasia
ATM	ataxia telangiectasia mutated
AWS	axillary web syndrome
AGVP	African Genome Variability Project
CT	chemotherapy
DSB	double strand breaks
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
EMT	endothilial mesenchymal transition ·
ERα	estrogen receptor alpha
ERβ	estrogen receptor beta
FS	frozen shoulder
GWAS	genome wide association studies
HT	hormone therapy
HWE	Hardy-Weinburg equillibrium
H3A	Human Heritability and Health project
IL	interleukin
IL6	interleukin 6
MAF	minor allele frequency
MRM	modified radical mastectomy
MSK	musculoskeletal
mRNA	messenger rna
NSAIDS	non-steroidal anti-inflammatory drugs
PCR	polymerase chain reaction
PF	pectoralis major fascia
PROM	patient recorded outcome measure
RIF	radiation induced fibrosis
RIPF	radiation induced pulmonary fibrosis
ROM	range of motion
ROS	reactive oxygen species
RP	radiation pneumonitis
RT	radiotherapy

SMAD4	mothers against decapentaplegic homolog 4
SMI	small molecule inhibitor
SNP	single nucleotide polymorphism
SOD2	superoxide dismutase 2
SPAD/SPADI	Shoulder Pain and Disability (Index)
STROBE	STrengthening the Reporting of OBservational Studies in Epidemiology
TGFß	transforming growth factor (TGF-ß)
TGFß1	transforming growth factor beta isoform 1
TGFßR	transforming growth factor beta receptor
TNF	tumour necrosis factor
VEGF	vascular endothilial growth factor
XRCC	x-ray repair cross-complementing protein
XRCC1	x-ray repair cross-complementing protein 1

List of Figures

List of Tables

Abstract

Introduction – Treatment for breast cancer is associated with a risk of chronic shoulder and upper limb morbidity in up to 30% of patients. There is currently no consensus for the possible reason for this often repeated finding in the literature. Previous research has suggested that development of fibrotic tissue in response to cancer treatments such as surgery and radiotherapy could be an underlying cause of musculoskeletal dysfunction and pain. This study investigated if any genetic variants in several key fibrosis-modulating genes could be shown to be associated with risk of upper limb musculoskeletal dysfunction and pain in breast cancer survivors.

Participants and Methods – A cross sectional study design was employed, using a candidate gene approach. A total of 326 South African breast cancer survivors were recruited from a tertiary hospital in the Western Cape (343 total, minus 17 samples with insufficient data collected). Each participant was scored for symptom severity using the shoulder pain and disability index (SPADI) questionnaire. Participants were then grouped for symptom severity using low, med or high SPADI scores. The low SPADI group served as controls (controls n=273, cases n=70). Participants were invited to donate a blood sample from which DNA was extracted. Each DNA sample was genotyped at seven polymorphic sites; three in TGF-ß, two in ATM, one in SOD2 and one in XRCC1, using PCR technologies and TaqMan allelic-discrimination probes. The resultant genotypes were analysed using multivariate analysis, including inferred haplotype analysis to search for association to shoulder pain and morbidity after treatment. A logistic regression analysis was also performed to investigate the association between SPADI score and age of participant.

Results – When participant age was compared with symptom severity, it was found that younger participants were more likely to have moderate-to-severe symptoms than older participants. There was a significant difference in the minor allele frequencies between case and control groups for the rs4880 (C>T, SOD2) polymorphism. The T allele was present more in the case group than in controls, with minor allele frequencies of 0.67 vs 0.55 respectively. No other independent associations were noted for any of the remainder variants tested. When haplotypes were inferred for genes SOD2 and ATM, combinations between the rare alleles at rs4880 and rs1800058 (C>T, ATM) were associated ($F=4.35$, $p<0.01$) with a low SPADI score.

Conclusions – Based on these results, the allele combination between SOD2 rs4880 C>T and ATM rs1800058 is recommended for further study, in addition to the rs4880 polymorphism in SOD2. These novel results are suggesting that there may be an association between fibrotic genes and the development of upper limb sequelae after treatment for breast cancer. A larger case-control study would be required to validate and explore these findings.

Keywords – fibrosis, breast cancer, radiotherapy, shoulder dysfunction, shoulder pain, polymorphism

Introduction and Scope of Dissertation

This project was intended to explore a facet of a central research area from a larger study (HREC 2012-312).

It has been observed that some women, after treatment for breast cancer, suffer from long lasting and debilitating symptoms, affecting the upper limbs and shoulder joints. This area of research is largely underexplored, especially in a South African context [1]. Due to this paucity in research, healthcare professionals have limited resources to diagnose and treat this morbidity. Consequently, patients are not offered satisfactory information nor treatment, which adversely affects their quality of life. For these reasons, it was decided to investigate the self-reported symptoms experienced by patients after treatment for breast cancer. The aim of the research was to explore a link between genetic sequence variation and prevalence of symptoms, to better inform further research and potential treatment options.

The aim of this research was to test the hypothesis that an association could be found between variation in a gene (or multiple) and the SPADI score (as a measure of tissue damage). This was achieved by the following objectives: genotyping the participants at several key SNP loci and statistically analysing the results against their SPADI scores to identify any associations; comparing participant age against SPADI score to determine any association between age and SPADI score.

Chapter 1 reviews the current literature body of previously conducted research into the topic of late morbidity in tissue after treatment for breast cancer, using the published research from Shamley *et al.* (2007, 2014) as a basis [1–3]. The citations within these papers were then investigated, used to develop a bibliography, and used to compile the list of search terms. At the conclusion of the literature review, a list of suitable candidate SNPs was compiled for testing, which is shown in Table 1.1.1

Table 1.1.1 – A list of the genes selected for inclusion within the study, and the specific genetic polymorphism under investigation. Samples were genotyped at the; rs1800469 c.1347 C>T TGF-ß, rs1800472 c.788 C>TGF-ß , rs4880 c.47 C>T SOD2, rs25487 c.1196 A>G XRCC1, rs1800058 c.4258 C>T ATM and rs1801516 c.5557 G>A ATM polymorphisms in Mixed Ancestry South African population.

Gene	SNP
TGF-ß	rs1800469
TGF-ß	rs1800471
TGF-ß	rs1800472
ATM	rs1800058
ATM	rs1801516
SOD2	rs4880
XRCC1	rs25487

Abbreviations: SNP, single nucleotide polymorphism.

The literature review, Chapter 1, starts with an overview of breast cancer and its epidemiology within the targeted population of female South African, breast cancer survivors. The various treatment modalities for breast cancer were then presented including known complications. This section sought to find links in the literature between the treatment of breast cancer and musculoskeletal disease. The next section investigated fibrosis in general, and then presented evidence linking fibrosis to the pathology observed after breast cancer treatment. Finally, a list of candidate single nucleotide polymorphisms (SNPs) presented, which was compiled using previous literature and clinical research into key genes linked to fibrosis.

Chapter 2 - presents the materials and methods of the genetic association study conducted investigating links between genetic sequence variation and the clinically observed outcomes. In Chapter 3 - the results are provided, along with statistical analyses used to achieve the research goals. Finally, Chapter 4 - discusses the novel findings of this paper, and the possible biological implications of these findings. It discusses the potential clinical implications and potential future treatment modalities for fibrosis. In addition, future research directions are presented including the areas of strengths and limitations of the current study.

Chapter 1 - Review of Literature

1.1 Literature search

The databases PubMed, Scopus, Medline and the Cochrane Foundation were first searched for published work. The date of publication range was set from the years of 1986 - 2018 (unless otherwise noted). This ensured that research collated is based on modern research, with minimal exclusion of informative material.

The literature search commenced using a hand search of records provided at the outset of the project from the research team. This began with papers published by Shamley *et al.* A list of chosen keywords was then developed to conduct the following database search: (“breast cancer” OR “breast neoplasm” OR “breast carcinoma”) AND (“surgery” OR “chemotherapy” or “radiotherapy” OR “treatment”) AND (“fibrosis” OR “fibrotic” OR “scarring” OR “shoulder dysfunction” OR “shoulder pain”). The second main search conducted used the same keywords, with the exclusion of breast cancer specific terms (“cancer” rather than “breast cancer” etc), and the results were collated. This was done in order to collect relevant articles concerning similar morbidity in other cancers, not included in the breast cancer specific literature search. Records selected from this search process were first reviewed by title, if they matched the search criteria, the abstract, followed by full body, was reviewed. This process resulted in 78 records, with the study, outlined in Figure 1.1.1. Further articles were accessed using citations from this literature body. Finally, gene/SNP specific keywords (“*Gene name*” OR “*accession number*”) and outcome keywords were searched for, in order to locate any potential unexplored related pathologies. Results from all of these searches were included in this thesis.

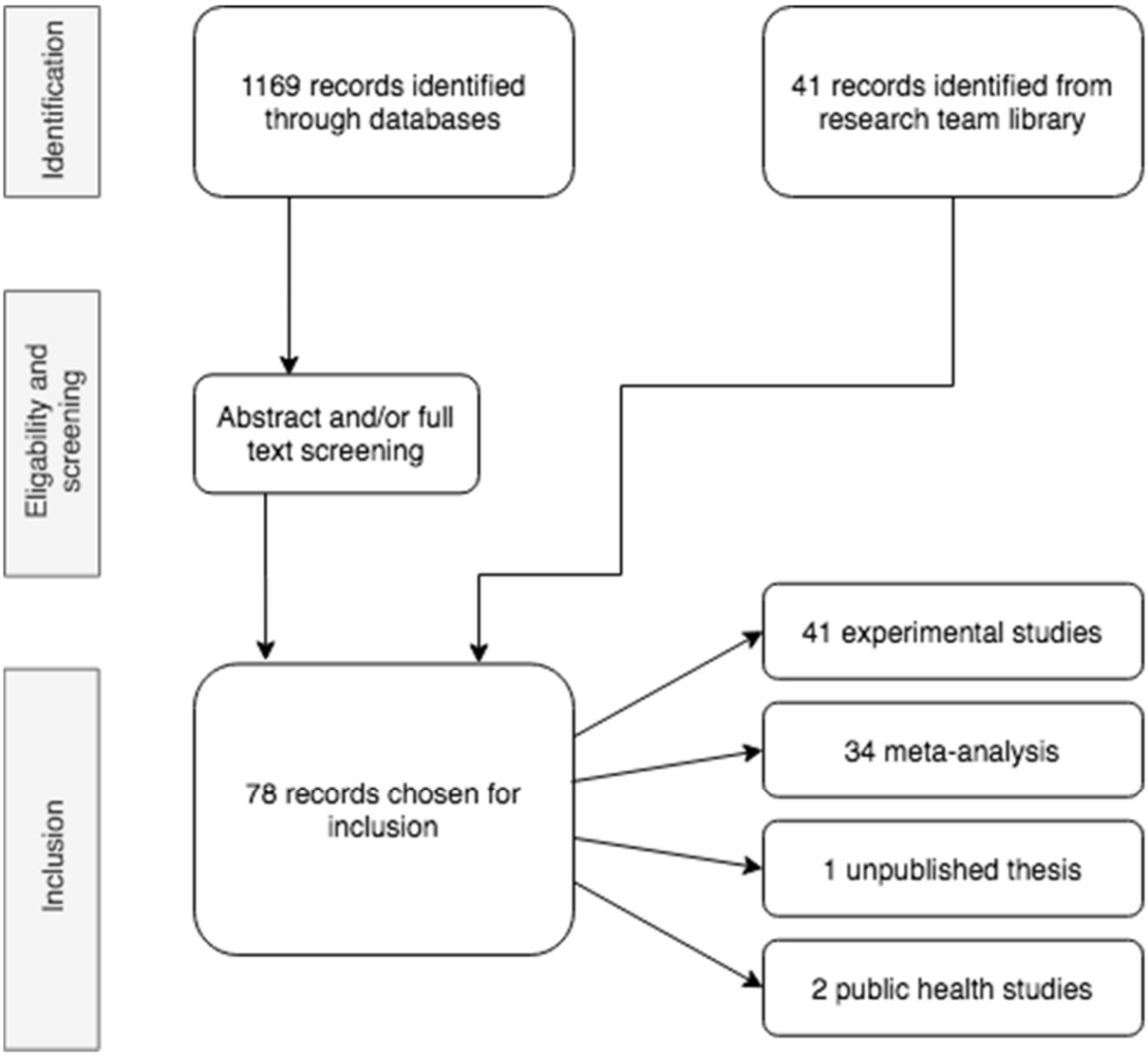

Figure 1.1.1 - Flow chart describing the initial literature review, the steps taken to screen and identify records used for inclusion, with the number of records included at each step, and the breakdown of the final records included.

1.2 Breast cancer in South Africa

Globally, breast cancer is one of the most common forms of cancer, it is also one of the most treatable forms, having the highest survival rates for long term prognosis of any cancer [4]. This success is partly due to the ability to remove breast tissue without catastrophic consequences for the individual [5–7]. Despite these factors, this form of cancer contributed to 12.7% of cancer mortality worldwide in 2008, with a 16.6% share of all cancer diagnoses in women [4,7–9]. Breast cancer is, like in many countries, the most common form of cancer in South Africa, with 14,097 new cases of breast cancer in estimated 2018. This is an incidence rate of 24.55 (new diagnoses per 100,000 people per year), which comprises 13.1% of all South African cancer diagnoses [10,11]. However, South Africans have a lower reported incidence of breast cancer cases than most western countries; for example, the same source estimates the incidence rate for the United Kingdom at 83.27 in 2018 [12]. This could be due to a number of biological factors such as a later menarche and a relatively early maternal age at birth of the first child, both factors that are associated with decreased breast cancer risk due to the later onset of hormonal cycle/fluctuation [13–15]. Lower reported incidence rates could also be due to socio-economic factors and standards of healthcare within South Africa. These factors can include poor access to diagnostic facilities among isolated rural/impoverished areas, poor education, poor awareness of early cancer self-detection practices, and inadequate reporting standards for prevalence and morbidity of breast cancer [7,16].

In the subsequent years since the end of apartheid, the South African national healthcare system has transformed into an integrative, comprehensive national service [17]. Many factors have hindered this transformation however, including economic disparity, poor national leadership and the HIV crisis. South Africans remain an underserved population, South Africa is technically considered a middle-income country, whilst health outcomes underperform many lower income countries. This is to be expected to be a contributing factor to the underperformance in cancer diagnosis and treatment. For example, a study conducted in the Western Cape reported an average time to diagnosis (of breast cancer) of eight and a half months after first self-identification of physical symptoms, which was primarily attributed to lack of education and knowledge, with access to healthcare being another major contributing factor [8].

Another example is the disparity in healthcare spending; less than 15% of the population have private medical cover, yet 46% of total expenditure is private [18].

1.3 Treatment of breast cancer

At the stage of diagnosis, an individual's breast cancer is rated according to stage of progression. In the early stages, breast cancer is confined to a single tumour colony, either in breast tissue, or the surrounding lymphoid tissue. When identified early as a single tumour, surgical removal of the tumour can be enough to halt the disease entirely, with or without the requirement for adjuvant therapy. However, once able to colonise other remote tissues and metastasize, the cancer is now malignant, which increases complexity of treatment and worsens prognosis [5]. Survival rates are much lower in cases where diagnosis is made later. Breast cancer is not female-specific, but it is much rarer in males, who make up less than 1% of cases, and differ in cancer biology [6,12,19].

Surgery physically removes solid tumours, while chemotherapy (CT), hormone therapy (HT) and radiotherapy (RT) aim to impede tumour cell reproduction through various biological signalling pathways and/or chemical damage [5]. These treatments are explored in more detail below. [20]

1.3.1 Surgery

The earliest recorded surgery for breast cancer was by cauterising drill in ~3000BC. More recently, the Romans excised the entire breast and/or pectoralis muscle in ~100BC. Unfortunately for patients, little progress was made for nearly two millennia [21]. The next significant improvement was the recording of the breast anatomy, specifically vascular and lymphatic structures in the 1500's. The first recognisable surgical procedure was the Halsted Radical Mastectomy, published by William Halsted in 1894 [21]. Although undeniably a disfiguring operation, a remarkable 40% of patients survived to five years post-surgery, a significant accomplishment for the time [21]. The current standard in surgery, the modified radical mastectomy (MRM), was published in 1972 by John Madden, which preserves both pectoral muscles [20,22].

Surgical techniques have since progressed, which has led to development of procedures where healthy breast tissue as well as underlying muscle and connective tissue can be conserved [23,24]. In the South African population, the majority of cancers

are treated with a radical or MRM and/or traditional external beam radiotherapy. This is partly due to late diagnoses, when a mastectomy is the only pragmatic option. This surgical practice may lead to complications in the tissues of the upper limbs [25].

1.3.2 Radiotherapy

RT is the standard treatment protocol in cases where local surgery is considered inadequate to reliably remove cancerous tissue and is predominantly delivered via an external beam [6,25,26]. This external beam uses ionising photon radiation to selectively destroy the reproductive ability of cells, by acting directly on DNA and depositing energy into the molecular structure, and indirectly by forming reactive oxygen species [5]. If delivered in small doses, the cellular DNA repair machinery is able to repair the damaged DNA with acceptable speed and accuracy. However with enough damage through larger doses of RT, the repair function stalls at the G2/M checkpoint with the biological purpose of stalling the reproduction of damaged cells into daughter cells with similar damage. These cells which are unable to undergo mitosis ultimately die by either apoptosis or necrosis. Cancerous cells that undergo mitosis regardless of this DNA damage (due to non-function of the stalling mechanism) usually do so in a haphazard and incomplete way which results in apoptosis or necrosis within a few divisions [27,28].

1.3.3 Radiotherapy morbidity

The high-energy beams used to deliver radiotherapy cannot discriminate between healthy and pathological tissue. As a result, it is inevitable that healthy tissue local to the target area will also be irradiated, and that side effects of this irradiation will sometimes present after treatment, either in the irradiated tissue, or in areas directly adjacent to the irradiated zone in this local stroma. This is in contrast to pharmacological therapies, where side effects are typically systemic, affecting all parts of the body to which the drug molecules are distributed by the circulatory systems.

As a result of this radiation damage, healing pathways are initiated which include a number of inflammatory and repair cytokines, the effects of which can be far ranging and damaging [29,30]. The early effects of RT become apparent within several weeks of

the completion of the course. Typical symptoms can include erythema of the skin, mucositis, nausea and diarrhoea. Late effects are more often present in the range of months to years after treatment and include radiation-induced fibrosis (RIF), atrophy of tissue, vascular degeneration, neural damage and deregulation of endocrine system function [31,32]. Progress in molecular radiobiology over recent years has vastly improved the understanding of the mechanisms of radiation pathogenesis, which is vital considering the increased population of cancer survivors who have undergone radiotherapy [33,34].

1.3.4 Radiation-induced pulmonary fibrosis

One of the first documented examples of pathology as a result of radiotherapy is that of radiation pneumonitis (RP). RP is the acute clinical manifestation of radiation-induced pulmonary fibrosis (RIPF). The later manifestation of this is severe radiation fibrosis of the lungs [35,36]. RIPF was first observed in patients undergoing RT and surgery for lung cancer in 1922, while the risk of developing RP was first quantified in a large cohort study by Lingos *et al.* (1990) [37]. Research performed into RIPF and similar conditions provided potential insights into possible causes of the clinical signs exhibited within the population of this study [38].

As stated, RIF was initially investigated in lung tissue, and has also been studied in other organ systems after cancer treatment. These include prostate cancer, bowel cancer, liver etc.[29,34,39,40]. A lot of useful data can be gleaned from these studies, but few studies involve musculoskeletal tissue, such as in the shoulder complex. As such there may be very different pathological factors in play.

In the final stage of late manifestation, RIPF is a serious disease, characterised by excessive deposition of extracellular matrix (ECM), which reduces lung function and leads to difficulty breathing [41,42]. In addition to this, the ECM destroys function of alveolar-capillary membranes, leading to hypoxemia and hypertension. This can be fatal without lung transplant. During investigation into the pathophysiology, there appear to be competent anti-fibrotic mechanisms in the soma, which may be malfunctioning [42]. An individual's response to the ionising effects of RT and other ionising agents is a complex and multifactorial trait, which carries an intrinsic polygenic

component as well as environmental factors such as diet, smoking status and age [43–45].

The prevailing view for much of the previous century was that dosage received by a cell/tissue was the predominant factor for the increased risk of developing RIPF. The theory states that tissue is damaged by cellular death from receiving RT, which leads to functional deficiency of that tissue [46]. However, increasingly the shift in paradigm is to a model by which radiation instigates a biochemical reaction in tissue mediated by signalling molecules, resulting in ECM deposition and mesenchymal cell proliferation which characterise fibrosis in connective tissue [47]. Attempts have been made to develop a predictive assay for individual response to RT, but as of yet no universal clinically applicable assay has been developed [47].

1.3.5 Direct beam and intraoperative radiotherapy

In response to this increased knowledge of radiotherapy-induced pathology, forms of treatment that aim to minimise this risk have been developed. Modern treatment facilities use radio-emitting devices that modulate the shape and orientation of the beam in three dimensions. This allows practitioners to deliver precise quantities of radiation and minimise erroneous irradiation [48]. This is performed under the hypothesis that the total dose of radiation received by healthy tissue is dosimetrically proportional to the damage done to the tissue, which is likely to be an oversimplification [49].

Within the field of radiation therapy, it quickly became necessary to strike a healthy balance between the therapeutic dose of any radiation received and the associated toxicity that RT exhibits towards healthy tissue. The study of this interaction has become a greater area of study as the burden of cancer survivorship post-radiotherapy has grown, initially in more economically developed countries and increasingly in the rest of the world with the dissemination of medical technology.

1.3.6 Radiogenomics

Andreassen *et al.* formed the term radiogenomics to broadly encompass the response to radiotherapy by the genetic component of the cell [50]. Named as a reference to

pharmacogenomics, radiogenomics broadly deals with the response to radiotherapy by the genetic component of the cell – “radiobiology meets molecular pathology” in a paper by Andreassen *et al.* (2002, p.1) [47]. Specifically, it refers to the role of genetic variation in response to exposure to radiation received as part of medical treatment. This information, along with improvements in radiobiology, allow greater advances in treating and preventing RT induced morbidity. Research into radiogenomics has initially focused on candidate SNP sites, but whole genome technologies are increasingly becoming more feasible for research [43,51].

The study of certain genetic diseases, such as ataxia telangiectasia, Fanconi’s anaemia and Mijmegen breakage syndrome [51–54], has provided evidence that tissue radio-sensitivity can be determined by genetic factors. All of these conditions derive from genetic abnormality in the cellular DNA damage repair systems. Patients suffering from these conditions have been shown to be more clinically radiosensitive than the general population [53]. This is also supported by animal studies, in which mammals such as the *scid mouse* have mutations in similar DNA damage/repair genes and suffer from extreme radiosensitivity [54,55].

1.3.7 Chemotherapy

As RT refers to treatment in the form of radiation, CT refers to the use of chemical agents to treat cancer [56]. Chemotherapy is the use of non-specific intracellular poisons, rather than molecules that have specific categories of targets, such as in hormonal therapy that target receptors.

One of the most commonly used chemotherapeutic agents is cisplatin, a metallic compound of platinum complexed with water, ammonia and chloride ions [57,58]. Cisplatin and other drugs with similar mechanisms of action are in the class of ‘cytotoxic’ drugs, and form an essential component in the treatment of malignant breast cancer [57,59]. When introduced to cellular conditions the chlorine ion is displaced by water to form cis-[PtCl(NH3)2(H2O)]. This complex is then able to bind to DNA bases, which stalls cellular mitosis, and initiates DNA repair mechanisms. This process ultimately leads to cell death via apoptotic pathways [57]. Therefore this

destroys cancerous cells due to the ability of cisplatin to preferentially kill cells which replicate quickly.

Although a highly effective treatment, cisplatin treatment is associated with a range of side effects. The most severe of these being a risk of developing nephrotoxicity, due to the cisplatin being cleared by the kidneys and accumulating within the kidney tissue [58,60]. This occurs via membrane-bound copper transporters which are highly expressed in kidney cells. Once within the kidney tissue, cisplatin is metabolically processed into more toxic forms, which damage both nuclear and mitochondrial DNA, leading to cell death [58–60] and nephrotoxicity.

Chemotherapy-induced peripheral neuropathy (CIPN) is a common complication for cancer survivors who have been treated with certain CT agents (cisplatin, thalidomide, taxol) [29,61]. CIPN is caused by the direct anti-mitotic and anti-angiogenic action of the drugs [61,62]. This loss of vascular tissue and other forms of toxic damage leads to axonal degradation, causing a loss of sensation and pain alongside other associated symptoms. The risk of these symptoms developing is a major limiting factor in chemotherapy. Once the neuropathies present symptomatically they are generally lifelong and incurable. As the population of cancer survivors grows, so does the burden of CT related disease [61,63].

As the burden of CIPN, and other cancer treatment-caused diseases has risen, so has the research conducted into potential treatments. One example of this is vascular endothelial growth factor (*VEGF*) gene therapy, which has been shown to inhibit the endothelial vascular tissue damage associated with CT, although no accepted form of treatment exists for use in human treatment [62].

1.3.8 Hormonal and targeted therapy

Hormonal therapies are anti-cancer agents that specifically target pathways of growth management. This is achieved through affecting chemical signallers of growth and impacting the receptors, and the intracellular transducers of these signals.

The role of growth hormones in breast tissue growth and differentiation has been known since the end of the 19th century. The primary hormonal regulators of growth are progesterone and oestrogen, which are both produced in the ovaries [64].There are two major oestrogen receptors; oestrogen receptor alpha (*ERα*) and beta (*ERβ*). Once the growth hormone molecules have been transported into the nucleus of the cell, the molecules bind to and activate these oestrogen and progesterone receptors. This effects cellular growth by altering levels of transcription. The alpha oestrogen receptor is expressed in 70% of cancer cases increasing the rate of cellular growth [64]. *ERα* ablation as a form of adjuvant therapies were investigated, leading to the discovery of tamoxifen by Harper & Wolpole in 1967 [64]. Tamoxifen is one of the most frequently used hormonal therapies for breast cancer treatment, belonging to the class of selective oestrogen-receptor modulating drugs [64–66].

Another form of HT is aromatase inhibitor (AI) drugs, which inhibit the aromatase catalysed synthesis of oestrogen [64]. One side effect of this type of HT drug is that it reduces levels of oestrogen amongst post-menopausal women. A study found, that amongst 437 Chinese women undergoing AI therapy, 11% had to prematurely discontinue AI therapy. The primary cause of discontinuation was joint pain amongst 57% of these women [67]. Another study reported that 10% of patients required a change of AI after development of severe sequela, of which joint pain (arthralgia) was the most commonly reported side effect[68].

As with many other treatments, HT is also associated with side effects. One of the most significant sequelae is arthralgia. One study reported that after two years post-treatment, 71% of women still reported arthralgia [69]. This joint pain has been shown to have significant impact on quality of life.

1.3.9 Therapy related cancers

Many of the treatment modalities for cancer create an associated increased risk of developing secondary cancers, termed therapy-related cancers. These secondary cancers can be due to the ionising radiation of radiotherapy destabilising the genomes of otherwise healthy tissue. Another cause could be from cytotoxic therapies promoting survival of cancerous cells with unstable genomics [56,70]. These therapy-

related cancers are unusual in that they have specific and quantifiable aetiological factors. The study of these cancers has led to several advances in understanding the role of the loss of DNA repair pathways in leading to genomic instability, including DNA double strand break gene translocation [56].

Finally, RT and CT can interact with other pre-existing factors, such as diet, cigarette smoking and genetic predisposition to cancer [56]. As shown, the treatment of cancer is often associated with treatment sequelae. In breast cancer, these side effects often manifest in the upper torso, shoulder and upper limb regions. The risk of these therapy related cancers, along with the risk of RT/CT sequelae, are two of the major dose limiting factors when planning a regime of therapy.

1.4 Shoulder joint complex and treatment

While drug therapies exert their effect systemically, surgery and RT exert a local effect initially. In breast cancer these local tissues primarily comprise the upper limb and shoulder. The tissue immediately surrounding the breast tissue are musculoskeletal and connective tissue components of the shoulder. The shoulder joint complex is the most mobile joint in the body and healthy movement of the joint requires unhindered movement of the gleno-humeral joint and the scapula over the surface of the thorax [3]. These motions, when combined, are known as scapula-humeral rhythm and are required when lifting and mobilising the arm. Further movement is achieved by coordination of the muscles acting across the joints [3]. There are a number of points during movement where pathology can affect this motion. Loss of function or stiffness in one muscle or muscle group will also affect the action of the antagonist muscle groups shown in **Error! Reference source not found.** leading to possible compensation injuries and dyskinesia in the opposing musculature. An example of this is shoulder impingement syndrome, where inflammation of the sub-acromial tendon can lead to impingement of the tendon against the acromion process, causing pain when lifting the limb laterally.

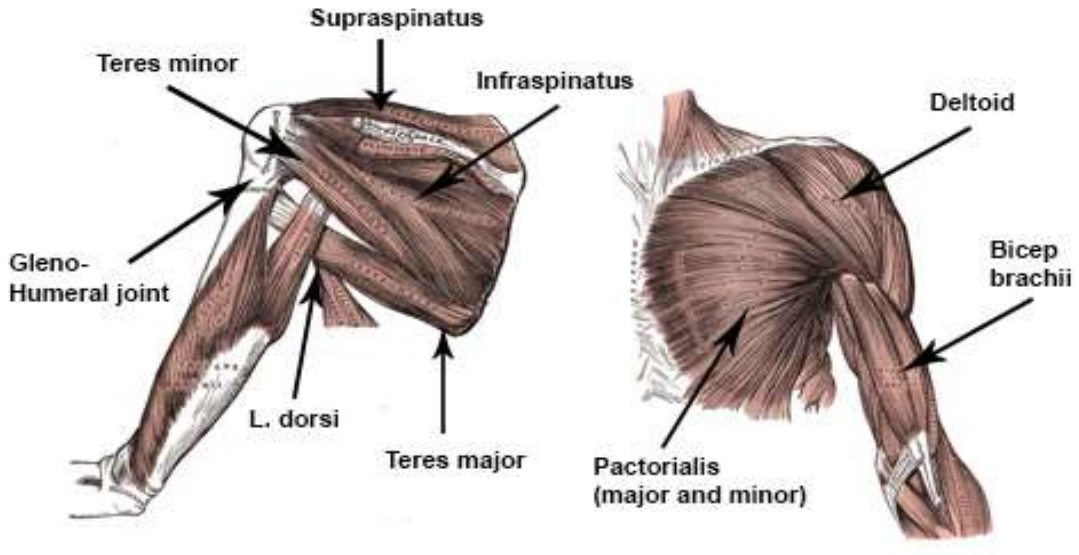

Figure 1.4.1 – Posterior (left) and anterior (right) view of the shoulder joint, with labelling of several key anatomical structures for shoulder stability and movement. Reproduced from Gray's Anatomy (1918) (Retrieved 04/06/2017), creative commons attribution, permitted use.

1.4.1 Clinical evidence

The previous section discussed the complexity of the shoulder joint, and the effects that musculoskeletal dysfunction will exert on this fragile joint. This section will explore the clinical evidence in the literature showing the disease burden on the population and the justification for conducting research into possible causative factors for a shoulder joint injury in breast cancer patients.

Many women who are successfully treated for breast cancer, later complain of a range of symptoms that cause pain and affect the function of the shoulder joint and upper arm. These symptoms have been noticed immediately and up to six years post-surgery [3]. This limits the ability of women to return to normal lifestyle after convalescing [3]. Upon investigation, Shamley *et al.* (2014) reported that 10-55% of women in a European cohort of breast cancer survivors showed restricted range of motion (ROM), 22-38% reported shoulder pain, and 42-56% of patients reported difficulty with lifting the arm above head height (Figure 1.4.2) [1,3].

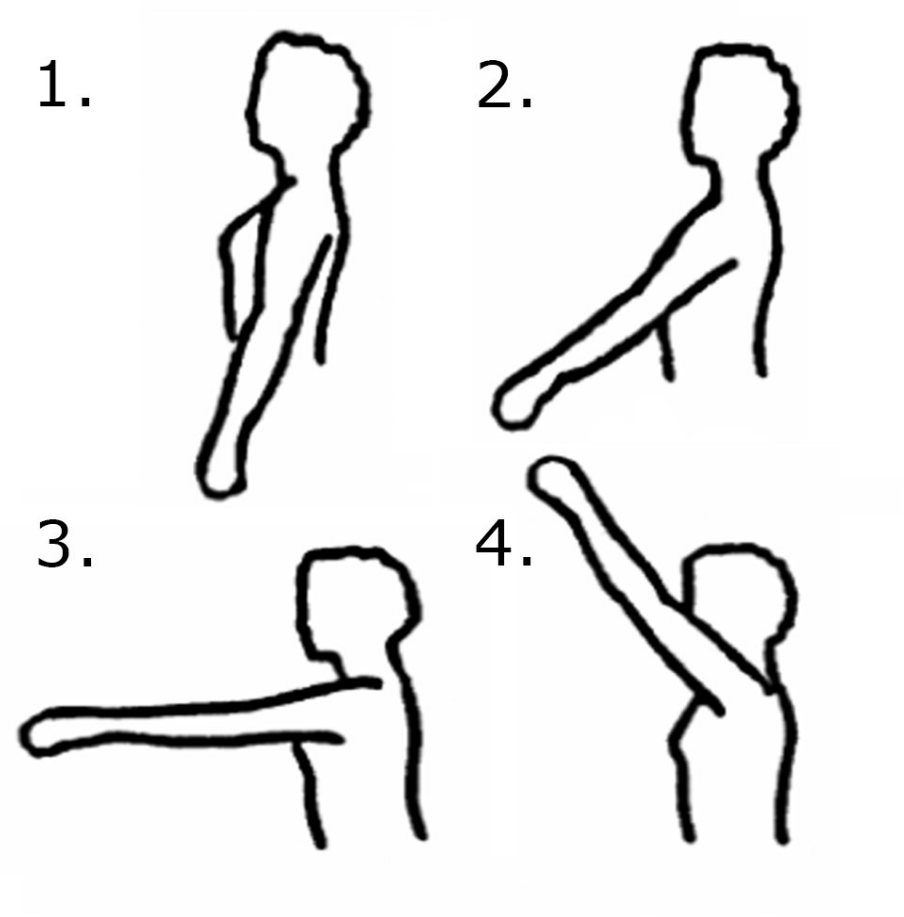

Figure 1.4.2 – A method of assessing range of motion of the shoulder. A patient is asked to move their aim from resting to vertical (1-4), and the limit of comfortable motion is used as an assessment. Adapted from Caban *et al.* (2006) [71].

The most comprehensive meta-analysis into this shoulder and upper limb phenomenon found was conducted by Rietman *et al.* (2003) into late morbidity of the arm after surgery [72]. Studies were assessed on patient population, study design, confounding factors and correct documentation of cohort. Studies were also assessed overall for reliability and validity. Within this meta-analysis six studies were found to fit their inclusion criteria. The main reason for exclusion was inconsistency between reports including lack of standardised shoulder assessment. Within the European cohort studies analysed, the prevalence of reported pain 12 months post-surgery ranged from 12-51%, which increased to 23-39% in the follow up between 14 to 38 months [72]. The prevalence of restricted range of motion of the treated side arm varied from 2-51%, with radical mastectomy showing most significant loss of ROM (Figure 1.4.2). Prevalence of edema was 6-43%, suggesting issues related to incorrect function of lymphatic/blood vessels. Although this meta-analysis identified a

relationship between breast cancer treatment and late upper limb mobility, Rietman *et al.* (2003) stressed that the strength of the relationship was low, and more work was needed [72]. A more recent meta-analysis indicated that the vast majority of studies showed evidence of shoulder impairment as a late effect from breast cancer treatments, with effects varying from small to substantial [73]. A similar conclusion was reached by Hidding *et al.* (2014) [74]. Some reviews indicated a stronger correlation between variants and late effect severity in patients who received adjuvant RT, as compared to patients treated only surgically [75]. Similar results have been reported by other groups such as Tengrup *et al.* (2000), Johansen *et al.* (2014) and Karki *et al.* (2004) [76–78].

After treatment for breast cancer kinematic changes to scapula-humeral rhythm and altered muscular control were observed [1,3]. In a study of 155 women treated for unilateral breast carcinoma, patients demonstrated a different movement dysfunction depending on whether the right or the left shoulder had been affected [1,3]. Muscle dysfunction was observed in both the cancer affected and unaffected side of patients, with use of adjuvant treatment [1,3]. These patterns of movement observed were similar to that seen in known shoulder conditions such as frozen shoulder syndrome/adhesive capsulitis, suggesting possible underlying trends that link the aetiology of several shoulder conditions with similar characteristics [1,3]. Shamley *et al.* (2014) used electromyography within a sample of 74 women (post breast cancer treatment) and results indicated that there was a decrease in muscular activity in several of the key muscles acting on the shoulder complex during upper arm movement [1]. This decrease in muscular activity was seen the upper trapezius, pectoralis major, pectoralis minor, serratus anterior and rhomboid muscles. The pectoralis muscles lie within the field of surgery and radiotherapy for breast cancer treatment, this damage can therefore be expected. However neither of the trapezius or rhomboid muscles fall within the field of surgery or RT for breast cancer, which indicates ‘secondary effects’ from surgery or RT. These effects have been shown to persist up to five years after breast cancer treatment [1,3].

A recent review serves to show that modern advances in surgical technique and RT have not adequately addressed the disease burden post treatment [30]. Hille-Betz *et al.*

(2015) studied 159 women following their outcome after breast cancer surgery & three dimensional conformal radiotherapy over a period of seven years. At a follow-up three months' post-surgery, participants were asked to rate their pain or discomfort in their arms or shoulders; 37.6% reported slight to severe pain in the arm or shoulder. This indicates that even with advanced and more modern conformal RT techniques, symptoms are still widely present [30]. However, at three months post-surgery, the time-frame is still too recent to indicate a chronic condition, as the local/systemic response to surgical trauma and adjuvant treatment may not have resolved.

Caban *et al.* (2006) after following up arm or shoulder pain 12 months post-surgery, suggested that the full impact of radiation induced sequelae might take longer to present, and suggested a five year follow up period to adequately show late effects [71]. Supporting this, Bentzen at al. (2001) found that 90% of radiotherapy damage was present within 3.7-4.2 years post-surgery, depending on clinical covariates [31].

1.4.2 Axillary web syndrome

One form of specific pathology observed in this upper limb morbidity is axillary web syndrome (AWS). AWS manifests as a visible web of axillary skin overlying palpable cords of tissue which become taut during shoulder abduction. This extends from the axilla towards the thumb which makes it painful to touch and creating tightness and a pulling sensation on the surrounding tissue as reported by Moskovitz *et al.* (2001). AWS has also been shown to affect ROM of the shoulder joint, which may be due to the cords causing the shoulder to be taut during abduction [79].

Despite small sample sizes, differences in classification and testing modalities, AWS prevalence after breast cancer surgery has been reported at 6-72% [2,79–82]. Kohler *et al.* (2013) noted that more aggressive forms of surgery, including removal of all connective tissue, adipose tissue and lymph nodes seem to predispose AWS development [81]. Upon investigation, immunohistological findings in the literature suggest the cord may be comprised of fibrosed and stased vessels, of either lymphatic or vascular origin [82,83]. A fibrosed cord could adhere to the surrounding fascial plane and contribute to the loss of ROM and shoulder dysfunction [2,84].

1.4.3 Frozen shoulder

The focus of this literature review is to explore outcomes after specific anti-cancer treatments. Other non-cancer reported conditions affecting the shoulder joint share similar symptoms, often with idiopathic onset, which could include underlying similar pathological mechanisms. An example of this is frozen shoulder syndrome (FS), a disease loosely defined in both aetiology and pathophysiology (also frequently referred to as frozen shoulder or adhesive capsulitis). First identified as a distinct condition in 1934 [85], frozen shoulder shares similar characteristics to the shoulder pain and disability described in this paper. Acquired FS is due to a causative factor such as surgery or trauma, whilst FS is idiopathic if there is an absence of any known causative factor. Bentzen *et al.* (2001) studied the loss of ROM after radical mastectomy and RT, and compared it to FS [31]. One study found that the average age for developing FS was ~50 years old, this observation correlates with a common age for breast cancer, although no causation is implied by this observation [86]. A monozygotic twins study by Hakim *et al.* (2003) showed that the risk of developing FS is strongly heritable [87]. This suggests a strong genetic link to FS, and possibly genetic links to similar pathologies.

In attempting to treat FS, Hand *et al.* (2007) took arthroscopic biopsies of rotator interval tissue and subjected them to histological and immunocytochemical analysis. This revealed the presence of excess, proliferating fibroblasts and an infiltrate of inflammatory cells, which are hallmarks of chronic fibrosis [88]. This pathogenic progression of inflammation to fibrosis is frequently referenced in other papers [88–93]. It is possible that FS and shoulder pain and disability (SPAD or SPADI) share similar pathologies.

1.4.4 Connective tissue and fascia

More advanced medical procedures are endeavouring to preserve as much tissue as possible during the mastectomy. This is considered with the intention of minimising the risk of post-treatment secondary disease. The fascia, is a series of connective tissue sheaths covering the musculoskeletal system, which are comprised mainly of loose collagens [94]. Fascial tissue interconnect with denser forms of connective tissue

(i.e tendons) and also contains adipocytes and other non-connective tissue aspects [94].

Removal of the pectoralis fascia is considered an essential part of a standard radical mastectomy, for the purpose of minimising the risk of any residual cancerous tissue remaining. However, this assertion is not supported by all evidence. Several studies have shown that there is no significant risk of remittance when fascia-sparing procedures are used [83]. A retrospective study on 1031 patients who underwent a MRM, but not RT was performed by Katz *et al.* (2001) [95]. It was found that patients with pectoralis major fascia (PF) invasion had a 33% risk of relapse in ten years compared to 18% risk of relapse without PF invasion (p=<0.001). However, this was only found in univariate analysis, and the difference was not observed in multivariate analysis. This lead to the conclusion that the relapses could have been due to PF invasion or to increased invasiveness of more aggressive tumours. Senelein *et al.* (2004) performed 235 MRM's without removal of the PF, and local relapse rate was 6.4% at five years post-surgery, this led to the conclusion that the procedure is oncologically safe [96]. It is yet unclear as to whether the removal of the fascia is necessary or not. An increasing body of research indicates that connective tissue such as pectoralis major fascia has more bio-kinetic importance than previously imagined [94,96,97], which consequently impacts the function of the shoulder after treatment.

1.4.5 Fibrosis

As discussed, RT is highly damaging to living tissue and is capable of causing genetic damage which leads to cellular destruction long after the initial treatment morbidity has abated. Many links have been made between RT (for breast cancer), SPAD and fibrotic tissue found within the tissue [43,98]. Tissue toxicity has shown a range of asymptomatic changes in tissue structure and function. Early reactions occur within 90 days of treatments, with late effects presenting after 90 days [33]. Early reactions affect tissues that have a high metabolic turnover, such as the skin, gastrointestinal tract and bone marrow, due to the rapid rate at which healthy, pre-damage cells are replaced with post-radiation, damaged cells [99]. In contrast to this, tissues which turnover at a lower rate generally present later, such as tendons, muscles and bone [99]. This radiation damage has also been shown to affect cells and tissues that were

not directly radiated. This is via a proposed pathway whereby cells damaged by ionising radiation release cytokine signalling molecules that can diffuse through tissue, or be transmitted through cell to cell gap junctions, which cause a pro-reaction in cells not immediately irradiated [100–102]. It is believed that this cell damage can result in fibrotic plaques within soft tissue after RT [103]. Fibrotic plaques, comprising of inert connective tissue will not serve the contractile function required of muscle tissue. This will alter function and place additional compensatory stress on surrounding tissue, exacerbating issues and potentially having huge effects on quality of life.

One of the most comprehensive sources of evidence linking fibrosis to cancer treatment is a meta-analysis conducted by Westbury *et al.* (2012) [98]. This meta-analysis consisted of a comprehensive literature review into the topic of morbidity and side effects after treatment for cancer (RT/CT/HT), and was the broadest meta-analysis found in terms of size and duration. In addition to this, Shamley *et al.* (2007) reported gross areas of connective tissue in the shoulders of women upon autopsy after radiotherapy. As described in Figure 1.4.3, fibrotic areas of connective tissue can be observed in the affected shoulder of a patient at autopsy. The area of intense blue response to a connective tissue stain can be seen in contrast to the more moderate connective tissue stain response on the unaffected shoulder of the patient, in which more vascular tissues in red [1–3].

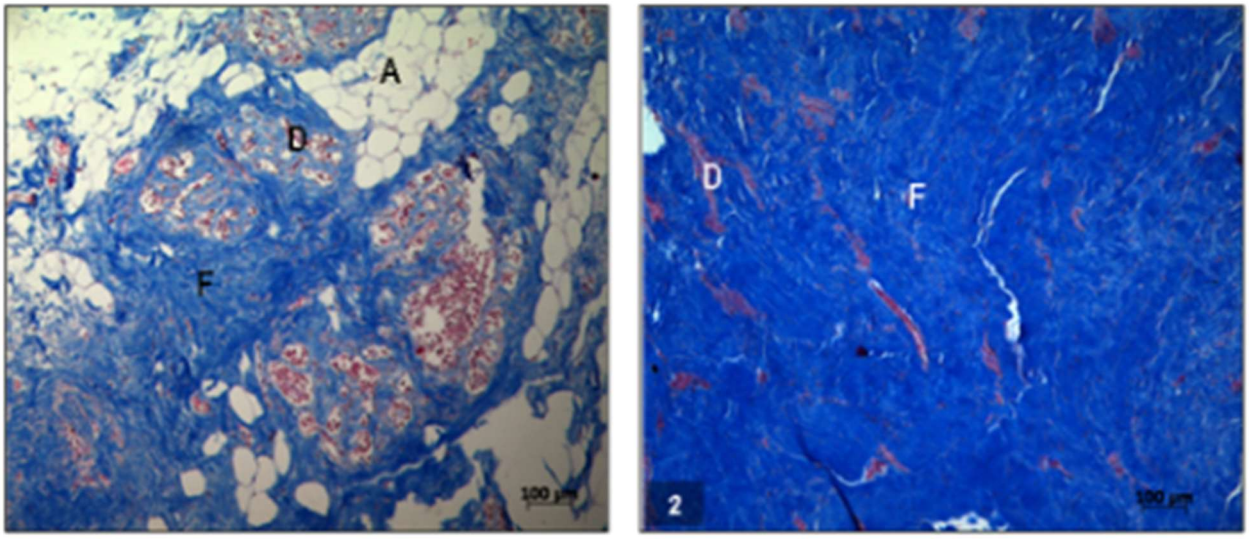

Figure 1.4.3 – Histological changes between unaffected side (left), and affected side (right) of upper limb at autopsy, showing pathological changes in structure to tissue. Patient treated with radiotherapy. Azure blue stain was used to indicate presence of connective tissue in blue. A – unstained adipose tissue. F – collagen fibres. D – blood vessels. Reproduced with permission from Shamley *et al.* (personal communication).

A description frequently used in the literature for the progression of fibrotic tissue is 'a wound that doesn't heal' [103,104]. This reflects the evidence that the molecular mechanism of radiation fibrosis appears to share mechanisms involved in normal wound healing, as described in Figure 1.4.3. In healthy healing, the process involves distinct phases as described in Figure 1.4.4 [36]. The significance is that in the final stage of healthy healing, active cells involved in the healing response (myofibroblasts, inflammatory cells, reactive oxygen species defence, immune cells) either undergo apoptosis, migrate away or mature and decrease in activity. This feature is typically missing in chronic, progressive fibrosis [34,54,105–107]. Research suggests that genetic variation in genes coordinating these wound healing biological pathways could impact the degree to which individuals are susceptible to this pathology [47,53,54]. This genetic association, if shown to be present, could indicate that post breast cancer treatment, it would be possible to identify individuals at a high risk for fibrotic progression. This could allow preventative or management treatment plans to be developed to mitigate this wherever possible.

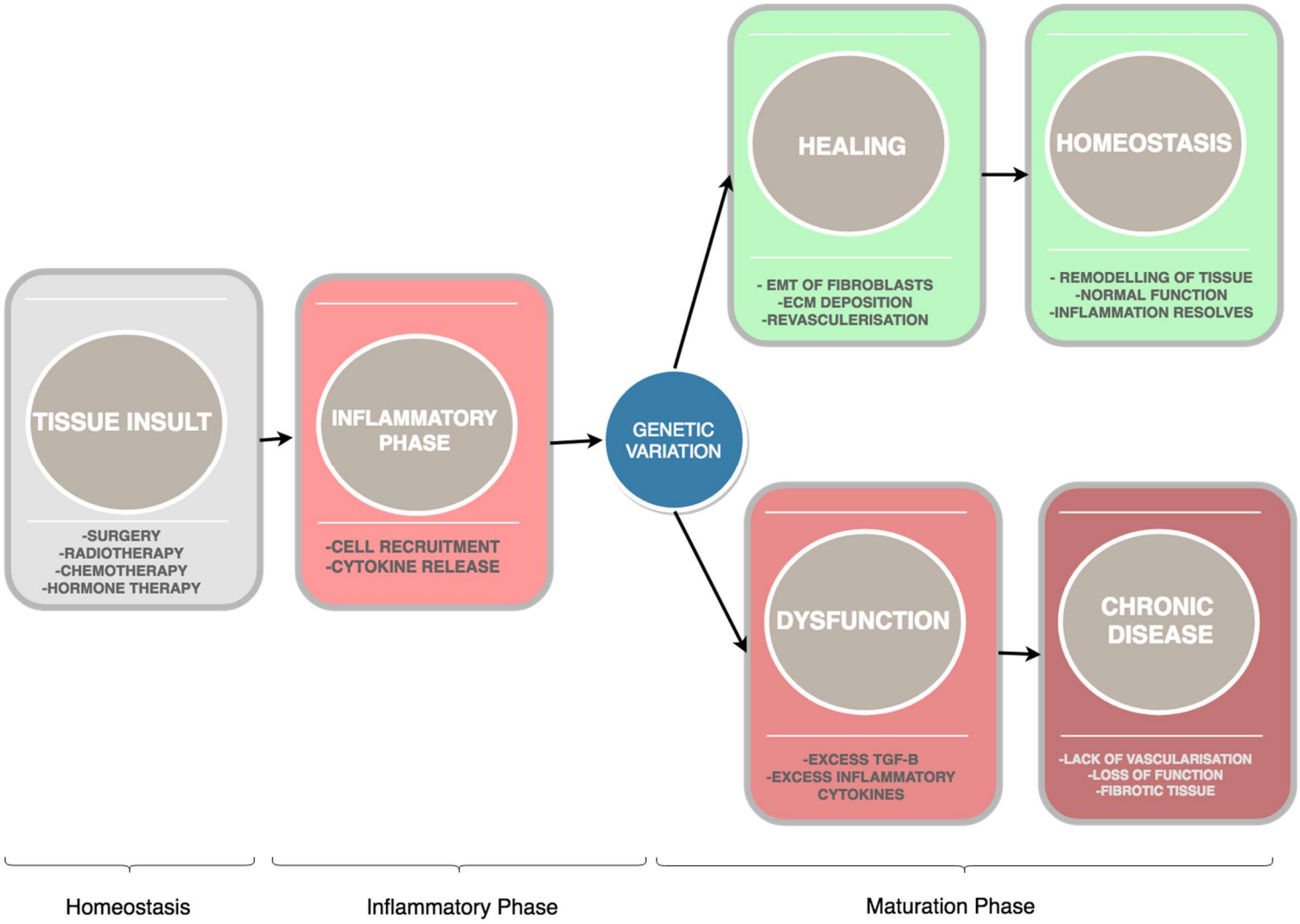

Figure 1.4.4 – The proposed model for development of fibrosis following cancer treatment. Genetic variation controls the actions of the cellular healing pathways following cellular insult, either promoting healthy healing and a return to normal tissue function, or degradation into a dysfunctional state of chronic disease, due to proposed but unproven factors. Abbreviations; EMT, endothilial-mesenchymal transition; ECM, extra cellular matrix; TGF-ß, transforming growth factor beta.

1.4.6 Shared pathogenic pathways in cancer and fibrosis

The pathophysiology of both cancer and fibrosis are associated with regions of hyper-proliferative stromal cells, particularly myofibroblasts [104,108]. Myofibroblasts are fibroblasts that have undergone endothelial-mesenchymal-transition (EMT), and are associated with areas of mass deposition of extracellular matrix components as well as instigating fibrotic processes through paracrine effects on surrounding stromal cells. The ECM deposited in this form of chronic fibrosis has an altered composition, comprising of higher levels of collagens I and III compared to normal collagen composition. Myofibroblasts also act physically, by contracting on the surrounding ECM [104,105]. Some authors have suggested that cancerous tissue in a region can predispose the region to fibrosis, and vice-versa [105,109]. Other forms of potential interplay between cancer and fibrosis exist, for example in the processes of inflammation. The inflammation present surrounding a tumour site (for example in breast cancer) has inflammatory factors which can additionally drive EMT and contribute to development of myofibroblasts, possibly driving fibrosis [104,110].

1.5 Biomolecular pathways in fibrosis

Development of fibrotic tissue within the shoulder complex following either RT, CT or HT has been hypothesised to be a major cause of the shoulder pain and dysfunction suffered post treatment. The inter-person variability of symptom severity is high. It is possible that some of this inter-patient variation is due to differing levels of activity in key fibrotic pathways. This has been shown to be the case in similar pathologies in different tissues. The most frequently reported pathways are those involved in DNA damage repair, inflammatory response and oxidative stress [44].

1.5.1 Reactive oxygen species

Several pathways are involved in mitigating the intra-cellular damage caused by reactive oxygen species (ROS), which are formed by the high-energy radiotherapy beam. ROS is a molecule capable of oxidising cellular components, including proteins, DNA and lipids [44,45]. These molecules are regarded as a form of intracellular signalling [53] and are naturally generated as part of normal cellular reduction/oxidation reactions and there are pathways within the cell that mitigate and repair the damage, for example via the action of anti-oxidant peptides such as glutathione and thioredoxin [44,45]. In this capacity they are believed to activate cellular stress pathways [107,111]. ROS generation forms an aspect of the cytotoxic effectiveness of RT. The massive oxidative damage destroys the structure of cells, by disrupting cellular macromolecules and through dysregulating cellular processes [45]. This complements the direct ionising effect of RT, as such it is believed that genetic variation within genes coding for ROS defence proteins is believed to alter an individual's reaction to ROS damage [43,44,112,113].

1.5.2 Genetic variation

Research presented in the previous sections has shown that inflammatory and damage repair pathways are heavily involved with the response to cancer treatment. Differences however in population, disease and treatment modality provide confounding factors that can complicate the process of identifying viable gene targets [114,115]. In addition, studies often do not reliably produce positive results and often produce results with low confidence [99,116–119]. It is a general consensus within the literature that individual gene variations have only a mild effect individually, and so far

no single master genetic ‘switch’ towards fibrosis has been identified [43]. Due to this lack of knowledge, the following sections will explore genetic variation as a potential modulator of fibrosis, which in turn influences upper limb morbidity.

1.6 Candidate gene approach

In order to investigate links between SPADI symptoms and genetic variation a candidate gene genetic association study was performed. The aim of genetic association studies is to identify an association between one or more genetic variants and a trait [120,121]. Genetic association studies can be used in a case-control design, for example, a group of individuals with a severe tissue response to RT can be compared to a group of individuals who did not have this abnormal response to RT. Statistical analysis can then determine whether a candidate genetic variant or group of variants are over-represented in one of these groups compared to the other. If this is found to be the case, then one, or a group of variations can be said to be associated with this outcome. This association is not proof that a variant is responsible for an outcome, but it can provide putative targets for future investigation. The candidate gene approach is used when candidate genetic variants are identified prior to experimentation for having a potential biological role in the studies end-point and is a hypothesis driven approach. Candidate variants are usually identified by reviewing the literature into similar outcomes, normally disease states or phenotypes, and are similarly limited by the degree of limitation of the knowledge of underpinning biology [120,121].

Polymorphisms within a gene can affect the polypeptide product in different ways. Either by variation in the amount of protein produced, or change in the structure and function of the protein. This change in protein structure and function can have amplified net effects in downstream biological pathways. In this way even a minor change in a protein coding sequences can have a significant result on the final biological outcome.

1.6.1 Transforming growth factor-beta

Transforming growth factor beta (*TGF-ß*) (19q13.2) is a widely expressed critical gene family with a large variety of roles. These roles are primarily related to modulating cellular differentiation, growth, proliferation and apoptosis [103,122,123]. In relation to cancer and fibrosis, TGF-ß protein activity has been shown to drive myofibroblast

differentiation during tumour progression. Due to this, it is frequently studied in relation to cancer biology, chronic fibrosis, musculoskeletal diseases and more [44].

TGF-ß is a hugely influential factor in modulating cellular events. *TGF-ß* activity regulates the proliferation and differentiation of cells, wound healing, and angiogenesis and receptors for it are found in almost all tissues [124,125]. **Figure 1.6.1** shows the activation of inactive TGF-ß dimers into active TGF-ß, recognition by the TGF-ß(1,2,3) receptors, and transduction of the intracellular signal, primarily through via mothers against decapentaplegic homolog (*SMAD*) complexes, which translocate to the nucleus and act as transcription factors [106].

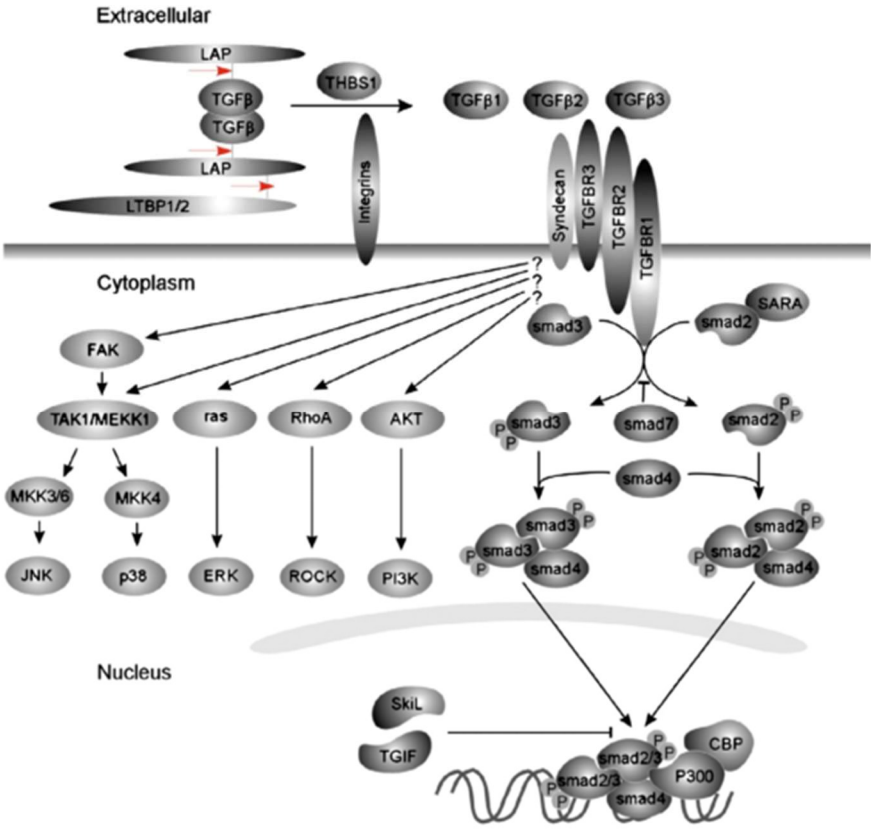

Figure 1.6.1 – A functional model showing inactive TGF-ß complexes, recognition by the receptor, and internal SMAD complexes that facilitate gene transcription, along with integration into other signal transduction pathways. Adapted from Pohlers et al. (2009) with permission [34]. Key abbreviations; TGF-ß, transforming growth factor beta; SMAD, mothers against decapentaplegic homolog.

Using the model of *TGF-ß* activity presented in **Figure 1.6.1**, an increased amount of TGF-ß1 protein produced will increase the amount of molecule/receptor binding. This increase would be mediated through the SMAD complex pathway and lead to

transduction of signalling pathways increasing ECM deposition and decreasing degradation of ECM.

TGF-ß activity has been linked to many fibrotic conditions, including atherosclerosis and fibrotic disease of the kidney, liver, and lung. In one study, breast cancer patients with significantly higher than normal *TGF-ß* serum concentrations were 90% more likely to suffer from idiopathic interstitial fibrosis after chemotherapy for breast cancer [126]. Another observational study of 91 women showed that higher levels of *TGF-ß* levels pre-RT was prognostic for the severity of the response to RT [127]. TGF-ß protein levels are included on lists of 'frailty' biomarkers, referring to risk of loss of muscle tone and coordination associated with aging [128–131].

One of the key manifestations of *TGF-β* activity in wound healing is activation of quiescent fibroblasts into myofibroblasts, which have a high capacity for ECM synthesis and form plaques of connective tissue. This fibrotic tissue is not capable of performing the biological function of the muscle and connective tissue that it is replacing and will affect the function of the musculoskeletal system as a whole. Activation of *TGF-ß* and the subsequent myofibroblast differentiation has been shown to be triggered by the physical flow of interstitial fluid into an area [70,132]. Additionally, it has been accepted that irradiation of somatic cells induced increased activity of *TGF-ß* and other pro-fibrotic genes [133–135]. However, results are not universally corroborated, and often correlation based rather than causally linked and the exact mechanism is not confirmed [27,135–137]. *TGF-ß* has also been studied in relation to fibrosis in conditions such as pulmonary fibrosis and gastrointestinal fibrosis conditions [38,138–142].

Skin and connective tissue have been shown to have increases in *TGF-ß* gene activity in a biphasic manner after injury, with twin peaks of increased activity after irradiation indicating activation of latent TGF-ß protein and *de novo* synthesis apparent two-four weeks after injury [34]. This is exacerbated by an associated increase in the TGF-ßR2 receptor productions due to positive feedback mechanisms [34]. It has also been shown that levels of TGF-ß can be a reliable indicator for risk of fibrosis after radiotherapy for breast cancer [127]. In attempting to develop more effective treatments for fibrosis,

studies have shown that inhibition of SMAD knockdown can inhibit the progression of fibrosis and relieve the restricted range of motion in murine models [92].

Figure 1.6.2 shows a summary of this fibrosis process and the believed influence of TGF-ß activity. Once fully progressed, fibrosis is extremely slow to self-correct, if not permanent. One potential prophylactic treatment could be attempting to intervene and modulate the early stages of fibrosis, which could be feasible if at-risk patients could be identified early enough in the process.

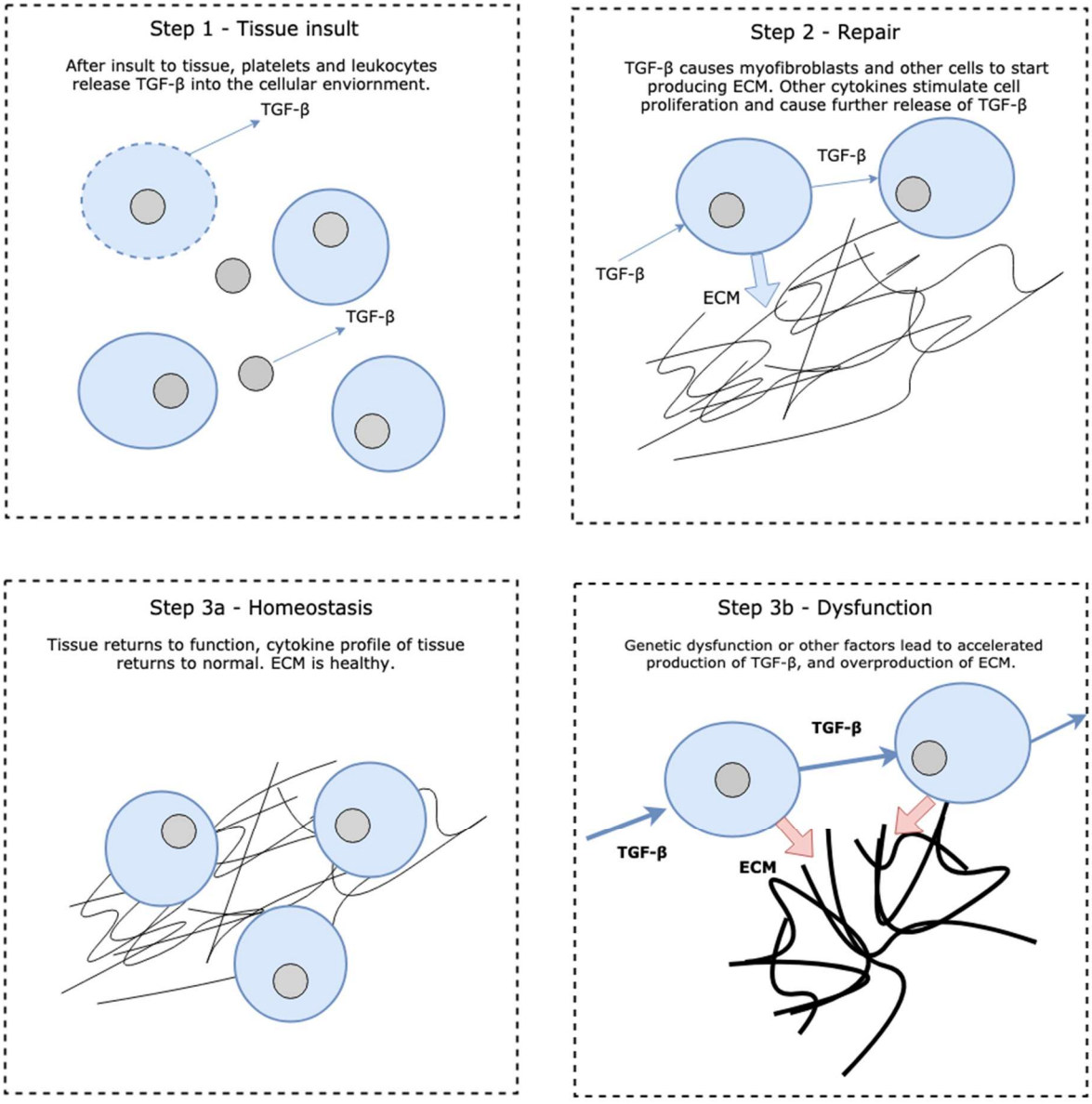

Figure 1.6.2 – A cellular model for the production and activation of TGF-ß after tissue injury; with the function of the TGF-ß protein either leading to normal healing and self-shutdown (3a) or a failure to - self-shutdown leading to a disease state (3b). Abbreviations; ECM, extra cellular matrix; TGF-ß, transforming growth factor beta.

Intra-gene variance in TGF-ß has been shown to be influential in driving individuals' pre-disposition to fibrosis [43,143,144]. Rs1800469 C>T is the most frequently reported site within the literature and has a clear proven increase on amount of TGF-ß protein produced [124]. Minor alleles at this particular polymorphism site can lead to increased amount of *TGF-ß* activity in cells, with reports by Shah *et al.* (2006) of up to twice as much protein in serum [124]. Rs1800469 is located within the promotor region of *TGF-ß*, and the T allele variant at this polymorphism prevents binding of a regulatory protein, thereby increasing the amount of TGF-ß produced. *TGF-ß* gene activity also drives a positive feedback loop that drives additional TGF-ß production, amplifying the effect of the variant allele [125,136]. In another study by Grainger *et al.* (1999), T homozygous individuals at rs1800469 produced up to twice as much TGF-ß1 compared to the C homozygous individuals [146].

Three candidate polymorphism sites within TGF-ß were selected, rs1800469 C>T, rs1800471 C>G and rs1800472 *G>A*. These were all considered to have sufficient evidence to be viable candidates for inclusion in this study. Rs1800470 G>A was initially also considered, but was discounted for further investigation as it was found in high LD with rs1800469 and thus the presence of one indicates the presence of the other [147].

1.6.2 Interleukin-6

The *interleukin*-6 (*IL-6*) (7p15.3) cytokine, along with other members of the *IL* family, is an important mediator in the acute phase of injury response. *IL-6* has a pro-inflammatory role, and also serves as a growth factor for fibroblasts. In mouse models, *IL-6* null mice showed a delayed wound healing response due to diminished myofibroblast differentiation [99]. Myofibroblasts are crucial to fibrosis, by releasing signalling factors that maintain an inflammatory response to injury. In addition to this, myofibroblasts in some tissues have been shown to produce IL-6 protein, providing a feedback mechanism which could drive fibrosis [106,148,149]. In observational studies of patients after RT, serum levels of IL-6 rise, which is not true for other interleukin family members, and this has been linked to an increased risk of radiation pneumonitis in patients [150].

1.6.3 X-ray repair cross-complementing protein 1

Another studied protein is X-ray repair cross-complementing protein 1 (XRCC1) (19q13.31), one of the core proteins involved in base excision repair [151]. It is believed to be a scaffold structure which coordinates the correct sequence of repair by other members of the base excision repair pathway [152]. *XRCC1* has also been investigated for its role in fibrosis. In a study of telangiectasia, fibrosis and atrophy in breast cancer patients, *XRCC1* variant genotypes were found to be positively correlated with increasing symptom severity (along with *TGF-ß*) [99]. This is consistent with results found by other groups [152–155]. Polymorphisms present in other members of the X-ray repair cross-complementing protein (*XRCC)* protein family have been linked to morbidity after treatment, but *XRCC1* has been most consistently linked and is believed to be the most promising gene for investigation [156,157]. The strongest candidate SNP site was *rs25487 T>C* with numerous reports finding links between variation and late effects after RT and other forms of treatment [158].

1.6.4 Superoxide dismutase

Neutralising the harmful effects of intracellular ROS is a crucial aspect of the response to ionising damage. This is necessary to minimise oxidation damage which if unchecked, can rapidly lead to necrosis. The gene superoxide dismutase 2 (*SOD2*) (6q25.3) is crucial to this process. The protein SOD2 acts by binding to superoxide, the by-product of oxidative phosphorylation, using manganese ions to catalyse the conversion into hydrogen peroxide (H_20_2) and diatomic oxygen (0_2). This maintenance of internal ROS levels is crucial and highly conserved. In this way, *SOD2* activity is highly cytoprotective, and has links to promoting tumour growth in cancer. Evidence in the literature links a polymorphic variant in the *SOD2* gene (*rs4880 A>G)* tissue response to RT. This variant is believed to affect the structure of the protein, and alter the ability of the protein to migrate into the mitochondria [36,47,159]. This variant (*rs4880*) leads to a reduction in the amount of protein produced and has been linked to increased survival in breast cancer patients treated with surgery and tamoxifen [113]. It is known that part of the effectiveness of RT is indirectly through generation of ROS [45], therefore a reduction in cytoprotective effects of SOD2 within cancer tissue would

enhance the effectiveness of RT to induce ROS damage. It is not completely understood how changes in the levels of intracellular ROS affects the ability of pre-cancerous tissue to progress into cancer. As such, it is not possible to predict accurately the effect of structural change or expression level of SOD2 on the risk of developing cancer at this time.

Andreassen *et al.* (2003) reported that amongst mastectomy patients genotypic variation at the rs4880 polymorphism in *SOD2* was associated with an increased risk of subcutaneous fibrosis leading to fibrotic disease [43]. Tentative correlations of this variant to symptoms has been reported in other studies [63-72]. rs4880 was considered a viable candidate for inclusion in this study [45,113,152,160].

1.6.5 Tumour necrosis factor-alpha

Inflammatory pathways are activated in response to threats such as pathogens, infection and injury and serve various functions. Tumour necrosis factor-alpha (TNFα) (6p21.33) signalling controls expression levels and influences TGF-ß signalling as part of inflammation [161–163].

Polymorphisms within this gene have been linked to a disease phenotype. Helmig *et al.* (2011) studied individuals possessing homozygous wild type alleles for the rs1800629 G>A polymorphism which was shown to be linked to the highest *TNFα* messenger RNA (mRNA) expression. In contrast, the homozygous rare allele (A/A) showed the lowest mRNA expression [161]. These results also correlated with lowered levels of *TGF-ß* mRNA expression [161]. This could indicate that variants in *TNFα* can act as a pro or antifibrotic factor in an individual. This was supported in the measured values of 178 study participants in pulmonary idiopathic fibrosis [164].

TNFα activity in fibrosis is indirect, and is as a result of its effect on *TGF-ß* activity. If *TNFα* polymorphisms were to be included in this study, it would mean that the end result (SPADI symptoms) are a result of an unknown (*TNFα* association) acting on another unknown (*TGF-ß* association). This indirect effect may well be beyond the power of this study to detect. Due to this, no variants in *TNFα* were included for further study.

1.6.6 Ataxia telangiectasia mutated

The gene product of ataxia telangiectasia mutated (*ATM*) (11q22.3) is a Serine/Threonine kinase enzyme, which is activated following the detection of DNA double strand breaks in response to ionizing radiation. The *ATM* was first identified in studies of patients suffering from ataxia telangiectasia syndrome (AT) [157,165]. In addition to the ataxia telangiectasia symptoms, it was also observed that these patients developed severe fibrosis after receiving RT [67,68]. Because of the link to a known gene mutation, the *ATM* gene was one of the first to be investigated for an association with normal tissue radiosensitivity [44,157,165]. However, individual reports of association is often unreproducible in other studies [157,159,165,166].

A loss of function of the ATM protein would impede-cycle arrest after DNA repair, as is seen in ataxia–telangiectasia syndrome. In addition, non-functional *ATM* would be unable to phosphorylate and active *p53*, which is also an essential component cell-cycle arrest [40,167,168]. Finally, *ATM* and *TGF-ß* activity are linked via the *SMAD* transcription factors (two and seven), and it has been shown that functional *TGF-ß* signalling is required for *ATM* activation [169]. This cross talk between *ATM* and *TGF-ß* has been shown to promote epithelial-mesenchymal transition and proliferation of fibroblasts in a renal model [167]. The altered levels of *ATM/p53* protein mediated cessation of cell cycle progression and associated growth could lead to cell death and atrophy in affected tissues. This uncontrolled cell necrosis and tissue atrophy could impede healing after cellular trauma and lead to tissue disruption/fibrosis and altered gross shoulder function [167,170]. Two polymorphism sites, rs1800058 and rs1801516 were included in the list of candidate sites for this study.

An overview of the pathways featured in this section are presented in Figure 1.6.3. Several of the downstream genes within a given pathway have been shown, along with the action of each gene to either positively or negatively influence downstream activity levels.

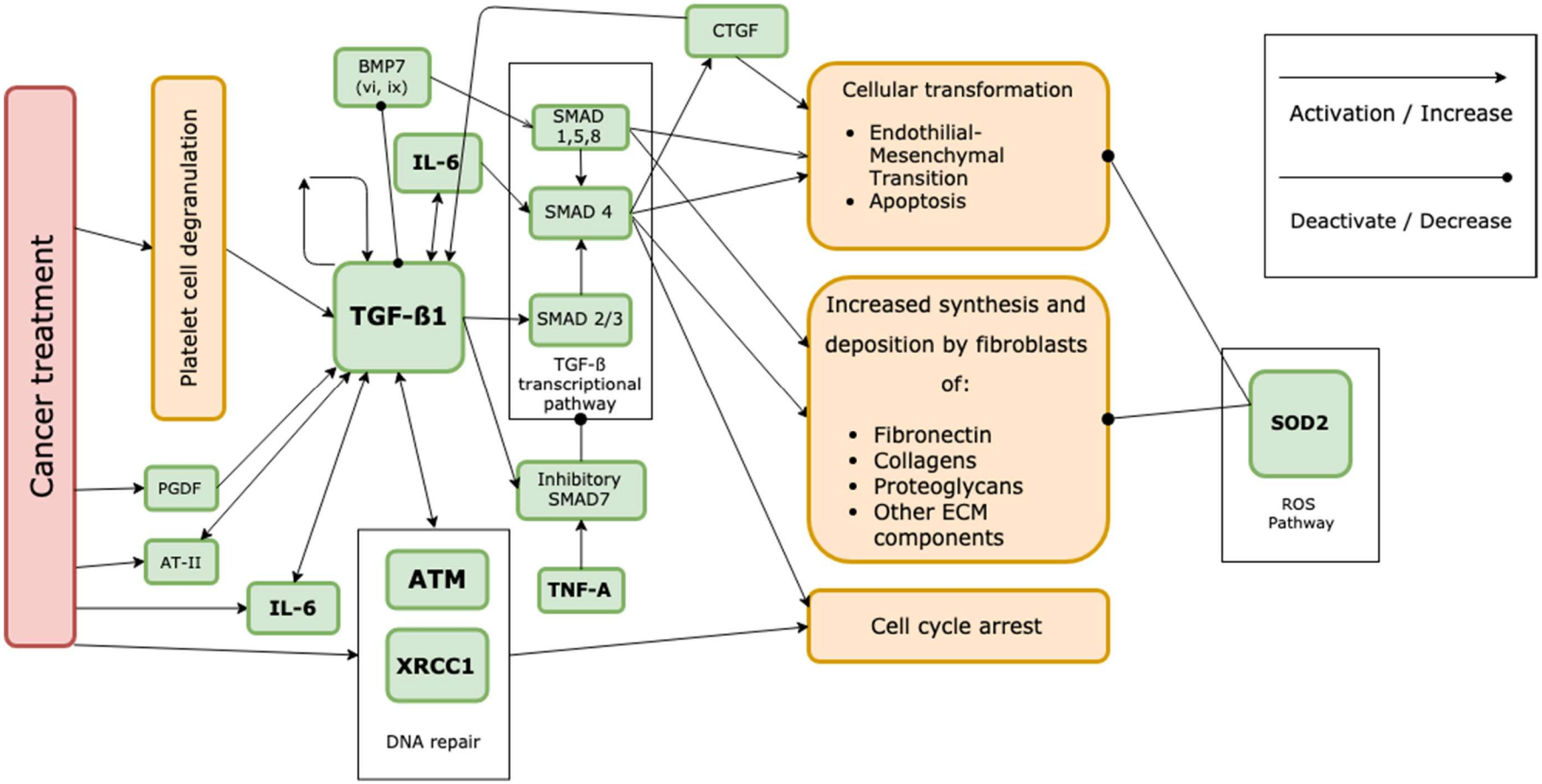

Figure 1.6.3 – A diagram of interplay between several of the key cytokines and processes believed to influence the progression of fibrosis. Orange indicates a biological process, green is a gene name, red is an external stimulus. Dotted tips indicate that an interaction is negatively modulating expression, arrow tips indicate that the interaction is increasing overall expression. Abbreviations: SOD2, superoxide dismutase; XRCC1, x-ray repair cross-complementing protein; ATM, ataxia telangiastica mutated; TGF-ß, transforming growth factor beta; PDGF, platelet derived growth factor; TNFα, tumour necrosis factor alpha; SMAD, mothers against decapentaplegic homolog; BMP-7, bone morphogenetic protein 7; CTGF, connective tissue growth factor; AT-II, angiotensin 2; IL-6, interleukin 6.

Table 1.6.1 – References for Figure 1.6.3.

Reference	Paper
154	Rapid induction of fibrosis and angiogenesis in vivo and stimulation of collagen formation in vitro
155	TGF-β: the master regulator of fibrosis
156	A portrait of Transforming Growth Factor superfamily signalling: Background matters.
157	TGF-β1 accelerates the DNA damage response in epithelial cells via SMAD signalling.
158	New tricks for an old fox: impact of TGFβ on the DNA damage response and genomic stability.
159	TGF-β/SMAD Pathway and Its Regulation in Hepatic Fibrosis.
160	TGFβ1-Induced Activation of ATM and p53 Mediates Apoptosis in a SMAD 7-Dependent Manner
160	ATM And Related Protein Kinases: Safeguarding Genome Integrity
161	A mechanism of suppression of TGF–b/SMAD signalling by NF-kB/RelA
162	MnSOD Overexpression Reduces Fibrosis and Pro-Apoptotic Signalling in the Aging Mouse Heart
163	The structural basis of XRCC1-mediated DNA repair.
164	Role of crosstalk between interleukin-6 and transforming growth factor-beta 1 in epithelial–mesenchymal transition and chemoresistance in biliary tract cancer
165	Interleukin-6 regulation of transforming growth factor (TGF)-beta receptor compartmentalization and turnover enhances TGF-β1 signalling.
105	Signalling in Fibrosis: TGF-β, WNT, and YAP/TAZ Converge
166	TGF-β signalling in fibrosis.
167	TGF-β signalling and the fibrotic response

Abbreviations: SOD2, superoxide dismutase; XRCC1, x-ray repair cross-complementing protein; ATM, ataxia telangiastica mutated; TGF-ß, transforming growth factor beta; ROS, reactive oxygen species; BER, base excision repair.
C - codon, * - references, † - Disputed or unclear.

1.7 Summary

To summarise, breast cancer is the most common cancer in the general female population but is one of the most effectively treated cancers [7]. Treatment usually consists of surgical excision of the cancerous tissue, radiotherapy targeting any cancerous tissue believed to be left remaining, and depending on the characteristics of the cancers, forms of chemotherapy, or targeted therapies are used [73]. Local therapies affect tissue in the shoulder complex, which is a highly mobile but fragile joint. Movement of blood, lymph and nerve signals from the trunk to the arm limb also traverse the shoulder joint and upper arm.

Despite increasing success rates in treating breast cancer, a subset of patients are left with debilitating side effects, often affecting the shoulder and upper limb function. Previously published literature into similar symptoms have suggested that a pathological condition involved could be fibrosis, driven by the cellular insult of cancer treatment, leading to pain and discomfort within the shoulder joints. Further research leads to the possibility of a link between genetic variation at certain genes and an individual's response to radiotherapy and other cancer treatments, either to return to healthy function or degenerate into dysfunctional healing. The size of effect of any polymorphisms is likely to be minor, and may be unnoticeable in the general population, only having a noticeable health affect in those who have undergone radiotherapy or other similar cancer therapies.

The experimental design of a candidate gene association study has been explored, and candidate polymorphisms have been presented, with an overview of the research findings that justify their inclusion in this study. These polymorphisms alter the activity of genes within pathways that instigate and modulate fibrosis. Using this previously published literature, the list of candidate SNP sites was developed to include the genes; *TGF-ß, SOD2, ATM and XRCC1*. Table 1.7.1 describes the candidate gene variants. Genotyping a cohort of participants at these sites was performed to determine if an association could be found statistically to the mentioned disease. This could indicate these SNPs as targets for future research, as well as further validating the experimental candidate gene approach for this form of observational study.

Table 1.7.1 – Overview of the candidate genes and single nucleotide polymorphisms.

Gene	Function of Gene Product	NCBI dbSNP id	SNP	Minor allele	Minor allele freq (1000genome)	Functional change	Consequence of variant
TGF-ß	Pro-fibrotic cytokine	rs1800469	c.1347 C>T, -509 C>T	T	0.36	Promoter region change	↑ in amount of protein produced (113,172,173,178–180)
		rs1800471	Promotor, c.74 G>C	C	0.048	Arg > Pro	↓ in amount of protein produced (113,172,173,181)
		rs1800472	c.788 C>T	T	0.01	Thr > Ile	Altered function (†) (173,179)
ATM	DNA repair (Ser/Thr protein kinase)	rs1800058	c.4258 C>T	T	0.01	Leu > Phe	Altered function (174,173)
		rs1801516	c.5557 G>A	A	0.07	Asp > Asn	Altered function (40,140,182)
XRCC1	DNA repair (BER)	rs25487	c.1196 A>G	G	0.26	Gln > Arg	Altered function (140,145,176)
SOD2	ROS scavenger	rs4880	c.47 C>T	C	0.41	Ala > Val	Decrease in activity (†) (146,147,183)

Abbreviations: SOD2, superoxide dismutase; XRCC1, x-ray repair cross-complementing protein; ATM, ataxia telangiastica mutated; TGF-ß, transforming growth factor beta; BER, base excision repair.
† - consequence of variant is highly debated or unknown.

Chapter 2 - Materials and Methods

2.1 Introduction

Breast cancer is typically treated with a combination of surgery, chemotherapy, hormonal therapy and radiotherapy [72]. While often successful in treating cancer, all of these therapies are associated with side effects, and a range of late sequelae that can last indefinitely [3]. In breast cancer, one such treatment complication is a musculoskeletal dysfunction, in the form of upper limb morbidity and shoulder pain after treatment for breast cancer, which researchers have quantified into the shoulder pain and disability index (SPAD or SPADI) [1,3]. This index uses a two part questionnaire to quantify the degree of the upper limb and shoulder morbidity as self-reported by the patient.

In the literature review, much of the previous literature available studying this shoulder pain and disability was explored. One proposed aetiology was that the cytotoxic effects of cancer treatments have led to improper formation of fibrotic tissue within the shoulder, which is responsible for causing observed symptoms [2,36,123,171,172]. Although the aetiology of this uncontrolled fibrosis is speculative, several genetic pathways are known to be crucial in modulating the fibrotic process . These include inflammatory pathways, DNA repair, oxidative repair, cell differentiation, cell cycle arrest and ECM management [99,115,139,153,173–176]. Research into similar diseases has shown that several key sites of genetic variation within genes in these pathways can be crucial in modulating fibrosis. This is due to the functional variation altering the activity levels of the protein produced. Variations in *TGF-ß, ATM, SOD2* and *XRCC1* were investigated [44,99,178–180,115,118,119,153,155,156,176,177].

Each of these genes encode proteins that function enzymatically within pathways that potentially influence fibrosis. *TGF-ß* is a crucial and pluripotent signalling gene family, that encode proteins implicated in numerous cellular processes including cell differentiation, maturation, apoptosis and damage repair processes [173]. Several reports have independently identified variants within *TGF-ß* as increasing the amount of protein produced, which can have amplified effects downstream on fibrotic pathways [36,99,183,119,139,153,155,157,173,181,182].

XRCC1 protein functions as a key scaffold protein in the repair of single-strand DNA breaks from ionising radiation during breast cancer treatment. Giotopoulos *et al.* (2007) showed an association between polymorphisms in *XRCC1* and severity of late effects after RT, as have other researchers [99,115,119,155,183,184].

The ATM protein is a serine/threonine kinase enzyme that activates DNA damage checkpoint in response to double stranded DNA breaks. This is achieved by ATM phosphorylating proteins that initiate a mediating response by halting cell cycle progression, and either commencing DNA repair or apoptosis [40,115,119,168,177,185].

The final candidate gene, *SOD2*, codes for multiple protein isoforms depending on splicing. SOD2 protein works to convert cytotoxic superoxide ions formed during radiotherapy into less harmful oxygen containing compounds [44,155,157,177,179,183,186].

The aim of this research was to test whether an association could be found between variation in a gene (or multiple) and the SPADI score (as a measure of tissue damage). This was achieved by genotyping the participant cohort for genetic variation at SNP sites within these genes, and then statistically analysing the cohort genotyping results against their SPADI scores, to identify any associations.

2.2 Materials and methods

This study was an observational pilot study, utilising a cross-sectional design, with candidate gene genetic association analysis as the primary outcome. The study followed the working genetic research paradigm, whereby a causal and quantifiable link was presumed to exist in the genotype-to-phenotype disease pathway at specific single loci.

Reporting of findings was done in alignment with STrengthening the Reporting of Genetic Association (STREGA) guidelines as described by Little *et al.* (2009) [187], which is an extension of the STrengthening the Reporting of OBservational Studies in Epidemiology (STROBE) statement [187].

2.2.1 Participants

All participants were previously identified and recruited from the parent study (HREC 312-2012). All biological samples and questionnaire results had been collected in the main study prior and have been used to form this sub-study. During the parent study, 352 patients were recruited and these were suitable for inclusion in this study.

The inclusion criteria consisted of females who had undergone one, or a combination of surgery, RT, CT and other forms of treatment for breast cancer in the range of the last one to six years at the Oncology Clinic of the Groote Schuur Hospital (GSH, Cape Town). Participants were recruited at their annual follow-up clinic visit. The participants were recruited from the Mixed Ancestry population, as this is the largest population group within the catchment area of the Western Cape, South Africa. This population is genetically diverse, with a genetic background suggested to widely contain four major distinct groups, Europeans, South Asians, Indonesians, Xhosa, and many more in smaller proportions. This group comprises 8.9% of South Africans as of the 2011 census [188,189]. All participants in the study were provided with a participant information form (Appendix A) and written informed consent (Appendix A) was obtained from each participant once it was acknowledged that they understood what was required of them. This informed consent also provided permission for their data to be used in further studies.

In order to minimise the risk of confounding factors participants were excluded from this study if they were younger than 18 years of age, had a history of shoulder or neck pathology prior to diagnosis of breast cancer, have had a diagnosis of any connective tissue disorders or any other systemic diseases believed to be associated with shoulder pathology such as rheumatoid arthritis, systemic lupus erythematosus, hyperparathyroidism, renal insufficiency, diabetes mellitus and familial hypercholesterolemia. This information was self-reported or available from medical records.

2.2.2 Ethics

This study was approved by the Research Ethics Committee of the Faculty of Health Sciences within the University of Cape Town, South Africa. All research was performed in adherence to the principle of the World Medical Associations Helsinki Declaration (1964) [190]. This included ensuring security of data, informed consent for participants and ensuring participant anonymity (Appendix B).

2.2.3 Clinical variables and patient reported outcome measures

Participants were asked to complete a validated SPADI questionnaire, a patient recorded outcome measure (PROM) with two domains, pain, disability and a lymphoedema questionnaire [75]. The SPADI questionnaire contained two sections, pain and disability. The pain section had five questions, with a range of 1-10. One was described as no noticeable pain completing everyday tasks, whilst ten was described as the worst pain imaginable. The section on disability had eight questions, range 1-10, with an answer of one indicating no difficulty at all completing everyday tasks, whilst 10 is impossible without assistance. The answers from these questions were combined into an overall SPADI score.

SPADI scores were used to group patients into the case/control groups for association testing. A score below a threshold SPADI score, 30 out of 130, was considered to be the controls, with little to no abnormal post treatment sequelae. This is described in Table 2.2.1. Case group scores were then divided into low, medium and high SPADI

bands which were used in the statistical analysis. The SPADI questionnaire, and other data collection documents are included in Appendix B.

All clinical variables were extracted from medical records, and transferred onto a source document before being captured onto a REDCap database, a secure data recording software platform hosted by the University of Cape Town (Research Electronic Data Capture, [191]).

Table 2.2.1 – Participants categorised into bands of shoulder pain and disability index score

Groups	Groups	Total SPADI score, 0-130	SPADI category
Control	Low	0-29	Little or no pain/disability
Case	Middle	30-49	Mild to moderate pain/disability
Case	High	50-130	Moderate to severe pain/disability

Patients were stratified into cases and controls, and into low, medium and high groups. Testing was performed on both sets of category.
Abbreviations: SPADI, shoulder pain and disability index.

2.2.4 DNA extraction

Approximately 4.5 ml of venous blood was taken from each participant by venepuncture in a forearm vein and collected into an ethylenediaminetetraacetic acid (EDTA) vacuum container tube. The tubes were kept on crushed ice and transferred into appropriately labelled cryo-storage boxes to the University of Cape Town Department of Human Biology (University of Cape Town, Anzio Road Observatory, Cape Town, South Africa) by the research assistant.

The whole blood samples were then immediately stored at -20 °C until DNA extraction would be performed. DNA extraction was performed by a standard salt-precipitation method as described by Lahiri *et al.* (1991) and modified by Mokone *et al.* (2005) (Appendix C) [192,193].

2.2.5 Single nucleotide polymorphism selection and genotyping

Within the literature review, seven SNPs within four candidate genes of were identified, and details were confirmed using the genetic databases of the National Centre for Biotechnology Information (NCBI, ncbi.nlm.nih.gov) and the European Bioinformatic Institute (EBI, www.ebi.ac.uk). Preference was given to SNPs with a MAF of >0.05. SNPs were either functional or regulatory variants. SNP nomenclature is in accordance with NCBI guidelines and full details of each SNP under investigation are included in

Table 2.2.2.

Table 2.2.2 – A list of candidate genes, gene products, and the single nucleotide polymorphism sites chosen for genotyping

Gene Symbol	Protein function	dbSNP id	SNP	Assay ID	Reporter dyes	Context sequence	*
TGF-ß	Pro-fibrotic cytokine	rs1800469	c.509 C>T,	C___8708473_10	VIC/FAM	GAGGAGGGGGCAA CAGGACACCTGA **[A/ G]**GGATGGAAGGGT CAGGAGGCAGACA	*
		rs1800471	c.74 G>C	C__11464118_30	VIC/FAM	GCAGGTGGATAGTC CCGCGGCCGGC **[G/ C]**GGCCAGGCGTCA GCACCAGTAGCCA	
		rs1800472	c.788 C>T	C___8708464_20	VIC/FAM	CAGATGCTGGGCCC TCTCCAGCGGG **[A/ G]**TGGCCATGAGAA GCAGGAAAGGCCG	*
ATM	DNA repair (Ser/Thr kinase)	rs1800058	c.4258 C>T	C__45273752_10	VIC/FAM	CTAGGATTCCTATC AGAAAATTCTT **[C/T]** TTGCCATATGTGAG CAAGCAGCTGA	
		rs1801516	c.5557 G>A	C__26487857_10	VIC/FAM	CTTGATTCATGATAT TTTACTCCAA **[A/G]** A TACAAATGAATCAT GGAGAAATCT	
XRCC1	DNA repair (BER)	rs25487	c.1196 A>G	C____622564_10	VIC/FAM	GGGTTGGCGTGTG AGGCCTTACCTC **[C/ T]**GGGAGGGCAGCC GCCGACGCATGCG	*
SOD2	ROS scavenger	rs4880	c.47 T>C	C___8709053_10	VIC/FAM	CTGCCTGGAGCCCA GATACCCCAAA **[A/G]**CCGGAGCCAGCTG CCTGCTGGTGCT	*

* - Some SNPs have flipped chromosomal locations on older/newer genome assembly constructs (older - RefSeqGene, newer - GRCh38.p12). As such, the assayed nucleotide is the complement of the reported SNP nucleotide. For the purposes of this study, RefSeqGene nomenclature is used to remain consistent with the majority of literature reports, which used the RefSeqGene.

Abbreviations: SPADI, shoulder pain and disability index; SOD2, superoxide dismutase; XRCC1, x-ray repair cross-complementing protein; ATM, ataxia telangiastica mutated; TGF-ß, transforming growth factor beta; BER, base excision repair; SNP, single nucleotide polymorphism; AKA, also known as.

Three SNP sites within *TGF-ß* were chosen; the first is rs1800469, c.509 C>T, which prevents regulating factors binding to the UTR, and increases amount of TGF-ß protein produced [153]. The second SNP site was rs1800471, c.74 G>C a functional polymorphism which predisposes fibrosis in some organs and has been linked to RT

tissue injury [194]. The final SNP was rs1800472, c.788 C>T, a SNP which is frequently investigated alongside rs1800469 [195].

Within the ATM gene, two SNP sites were chosen; firstly rs1800058, c.4258 C>T, linked to severity of RT effect post breast cancer treatment [178,196]. Then rs1801516 c.5557 G>A, similarly linked to tissue response [50]. One SNP site within XRCC1 was chosen, rs25487 c.1196 A>G and finally one SNP site within SOD2 rs4880 c.47 C>T, both linked to tissue response after breast cancer treatment [158,159,166,197,198].

Genotyping performed using TaqMan (Applied Biosystems, Foster City, California, USA) allelic-discrimination assays. The platform used was the Thermo Fisher QuantStudio Five real-time Polymerase Chain Reaction (PCR) system (Thermo Fisher, Waltham, Massachusetts, USA), using 96-well PCR places with Thermo Fisher reagents. These reagents had allele specific probes (sequences available from manufacturer on request) targeting the candidate SNPs. The PCR master mix was also provided from Thermo Fisher PCR, containing *Taq* DNA polymerase, dNTPs and all reaction component except for primers and template DNA. These were used according to manufacturer guidelines in a final reaction volume of 8 µl. Final reaction mixture was as follows: TaqMan PCR mastermix – 4 µl (containing ampliTaq DNA polymerase), H_20 – 2.85 µl, TaqMan probe and primer mix – 0.15 µl, DNA template – 1 µl.

The standard PCR cycling procedure was; 30 second hold at 60 ˚C, followed by 40 cycles at 95 ˚C for 15 seconds and 60 ˚C for 1 minute, with a 30 second hold step at 60 ˚C at the end. Genotypes were determined by endpoint fluorescence and alleles called automatically by the QuantStudio software provided as part of the previously inventoried QuantStudio system. Where the software was not able to automatically call a result, alleles were called manually or flagged as erroneous results manually in the software. For standard quality control and to minimise risk of erroneous results every 96-welled PCR plate setup, included negative control (no DNA present) and duplicates of samples to test for reproducibility.

All preparatory laboratory work and genotyping assays took place within the Division of Exercise Science & Sports Medicine facility, Department of Human Biology, Faculty of Health Sciences, University of Cape Town.

2.2.6 Data management

Patient data were managed using the REDCap software suite to ensure adequate data security [191]. This software suite also allowed for quality control of input, using a redundant call/input system for data entry. Databases were stored in a password protected form, with patient information protected using a Study ID no. system to ensure anonymity. captured data input was checked for validity by a separate independent data capturer.

2.2.7 Sample size

A power calculation was used to determine the appropriate sample size required to find effects of a given odds ratio at acceptable statistical power. This was performed using the Genetic Association Study Power Calculator (J. Johnson, University of Michigan, USA [199]).

Determining an accurate baseline prevalence for use in the power calculation was difficult to determine, as the majority of studies are relatively small, use different populations and different methodologies. Based on a rough estimation on a number of the previous studies, a 30% risk of significant shoulder pain and/or disability was used. [1,3,72–78].

Using an accepted significance level of 5% , an additive genetic model, and assuming minor allele frequencies of ≥0.1 it was shown that a sample size of n=343 is sufficient to detect an OR (genotype relative risk) of ≥1.7 at 80% power.

2.2.8 Statistical analysis

Statistical analysis was performed using R-Project v 3.4 (R Foundation for Statistical Computing, www.r-project.org). Seven SNPs in five genes were included in the current analysis. Genotypes were summarised as counts and percentages. To mediate the

risk of confounding error in candidate genetic studies, the genotyping Hardy-Weinburg equilibrium (HWE) testing was performed to determine if any of the measured SNP's deviated significantly from HWE. HWE testing via chi squared test requires the presence of all genotype combinations present in the sample, where this was not possible and where noted, Fishers exact test was used.

SPADI scores were corrected to a range of 0-100 for testing purposes, and were then summarised using mean and standard deviation, with median and interquartile ranges used when appropriate. Plots were constructed to show the results. The Mann-Whitney test was used to assess the difference in the SPADI score between cases and controls. Statistical analysis was performed using chi-square test of independence for categorical variables and unpaired t-test for continuous variables. A logistic regression analysis was conducted in order to adjust for age and determine if age was associated with the development of mid-to-high SPADI score. The R-Project package "dlpyr" was used (https://dplyr.tidyverse.org/)

Two-way analysis of variance (ANOVA) was performed to assess whether any of the genotypes for various SNPs was significantly associated with SPADI score. The study population (n=326, from total study of 343, minus 17 samples with no data collected) is above the minimum required number of samples as determined by the ANOVA calculator (The Australia and New Zealand Melanoma Trials Group [200]), so an ANOVA test was used. ANOVA was used as it is highly robust to skewness, especially when large sample sizes are used. Additionally, ANOVA was used because the independent variable (genotype) had more than three levels while the continuous variable (SPADI score) was continuous. Following the study power calculation, normality was assumed through the group. Two-way ANOVA was chosen to take into account the association between group (case *vs*. control) with the SPADI score. Haplotypes were inferred and used to test for possible genotype interaction effects.

Pairwise post-hoc comparisons were performed following the two-way ANOVA to assess the pairs that are significantly different from each other (Bonferroni correction was used to adjust for multiple comparisons). Significance was accepted at $p<0.05$ throughout the analysis.

Chapter 3 - Results

3.1 Participant demographic and clinical characteristics

Table 3.1.1 describes the clinical information and participant characteristics recorded during the interview. Race was self-reported, and age was as at the time of consent, not at time of questionnaire completion. Age was significantly higher in the control group compared to the case group.

Table 3.1.1 – Descriptive statistics for the study participants, stratified by case or control group.

Characteristic	Level	Case group	Control group	p value	Test
Group	Control	273 (100)	0 (0)		Independent sample t-test
	Case	0 (0)	70 (100)		
Age at surgery, (years)		53.43 ± 0.3	50.66 ± 0.2	0.14	
SPADI score		7.02 ± 8	51.06 ± 17	**<0.001**	
Race	Black	37 (14)	13 (19)		
	Mixed Ancestry	215 (79)	54 (77)		
	White	21 (8)	3 (4)		
Side of primary affect	Missing data	9 (3)	0 (0)	0.195	Fishers exact test
	Left	138 (51)	41 (59)		
	Right	126 (46)	29 (41)		
Tumour nuclear grade at surgery	Missing Data	47 (17)	9 (13)	0.788	Chi^2
	I	52 (19)	14 (20)		
	II	119 (44)	34 (49)		
	III	55 (20)	13 (19)		
Did participant receive chemotherapy?	Missing data	45 (17)	7 (10.0)	0.087	Fishers exact test
	No	33 (12)	4 (6)		
	Yes	195 (71)	59 (84)		
Does participant have problem opposite shoulder	Missing data	26 (10)	0 (0)	**<0.001**	Chi^2
	No	221 (81)	52 (74)		
	Yes	26 (10)	18 (26)		

Data reported as frequency (%) or mean ± SD. Age and race were self-reported at time of interview. Significant p values are highlighted in bold. Tumour grade assessed by nuclear grading system [201]. Missing or incomplete data could be due to patient consenting, but not completing data collection, patient leaving study centre prior to data collection, or due to incomplete medical records.
Abbreviations: SPADI, shoulder pain and disability index; SD, standard deviation.

3.2 Shoulder pain and disability index scores

Figure 3.2.1 shows that the SPADI score distribution for both controls and cases were skewed to the right or positively skewed. A total of 79 participants had a SPADI score of zero, which indicates that no pain or disability symptoms were reported by these participants (SPADI=0, *n*=79). This is expected in clinical research, partly due to the subjective nature of pain reporting. SPADI score data was missing for 23 participants. The median SPADI scores for the controls were heavily skewed towards lower SPADI score, as expected (median = 3.8).

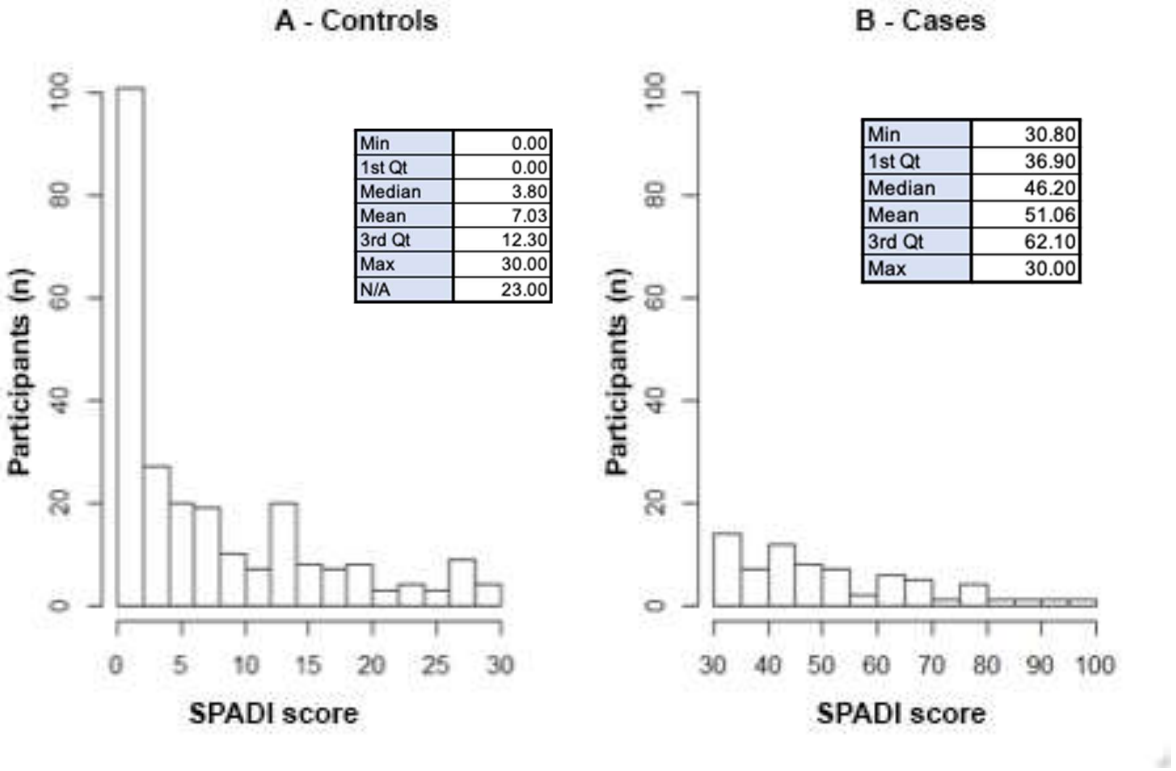

Figure 3.2.1 – The distribution of total shoulder pain and disability index scores amongst testing population, with control on the left side and case in the right side. Histogram breaks of 12 were used. Participants given as count. A – control group, B – case group. Interquartile ranges, minimum and maximum scores given in boxes. There was missing data for 23 participants in the control group.

3.3 Age as a predictor of SPADI score

Table 3.3.1 displays the results of an initial logistic regression model using age at consent as a potential risk factor for mid-to-high SPADI score. **Error! Reference source not found.** displays the age data from **Table 3.3.1** in a box-whisker plot to indicate the arithmetic values for the participant groups. When these data was used in the logistic regression model, it was found that for every additional year of age, the log odds of having low SPADI score increases by 0.034. The p-value indicates that age is statistically significant ($p<0.01$) in predicting whether the participant will have a low SPADI score.

Table 3.3.1 – Logistic regression analysis comparing participants age at consent to SPADI score.

Group	Minimum	1st quarter	Median	Mean	3rd quarter	Maximum	No data, n
All participants, yrs	25.4	46.6	53.9	53.6	61.4	74.5	30
Control group, yrs	25.4	48.1	55.6	54.4	61.4	74.5	26
Case group, yrs	32.8	42.3	49.6	50.7	58.6	71.3	4

Predictor	Odds ratio	Standard error	z value	p value
Age, yrs*	0.0342	0.0128	2.68	**<0.01**

Participant age was used as independent variable, control and case group as conditions, with SPADI score as dependant variable. Values for age are given in number of years. No data given as a count. Significant values are highlighted in bold.
*Statistical analysis performed by a logistic regression test.

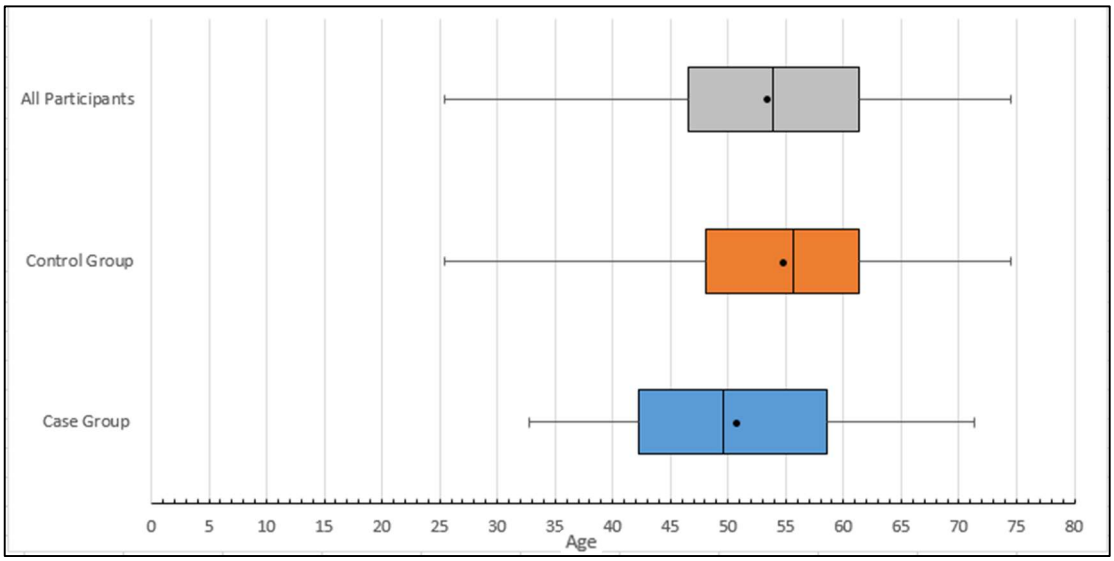

Figure 3.3.1 - Box-whisker plot showing the distribution of age in all study participants (top), and separated into the case and control groups (middle and bottom bar). The total bar length is range, the boxes are the interquartile range, the black line bisecting the box is the median, and the black dots indicate the arithmetic means age of the group. Age is given in years.

3.4 Genotyping results and allele frequencies

Table 3.4.1 reports the data for the distribution of genotypes throughout the cohort, as reported from the PCR procedure.

Table 3.4.1 - Genotype and minor allele frequency distributions, and Hardy-Weinburg equilibrium tests.

Polymorphism	Genotype	Controls (*n*=273)	Cases (*n*=70)	p value
Transforming growth factor ß	No data	110 (40)	25 (36)	
	CC	57 (21)	19 (27)	0.924
rs1800469	CT	99 (36)	24 (34)	
c.1347 C>T	TT	7 (3)	2 (3)	
	HWE	**<0.001**		
	T Allele	0.35	0.31	
Transforming growth factor ß	No data	55 (20)	12 (17)	
	CC	214 (78)	56 (80)	0.718
rs1800472	CT	4 (2)	2 (3)	
	TT	0 (0)	0 (0)	
c.788 C>T	HWE Fishers exact test	0.609		
	T Allele	0.02	0.02	
Superoxide dismutase 2	No data	75 (28)	17 (24)	
	CC	74 (27)	20 (29)	**0.045**
rs4880	CT	98 (36)	19 (27)	
c.47 C>T	TT	26 (10)	14 (20)	
	HWE	0.721		
	T Allele	0.38	0.44	

Polymorphism	Genotype	Controls (*n*=273)	Cases (*n*=70)	p value
Xray repair cross complementing protein 1	No data	73 (27)	17 (24)	
	AA	116 (43)	29 (41)	0.650
rs25487	AG	68 (25)	20 (29)	
c.1196 A>G	GG	16 (6)	4 (6)	
	HWE	0.205		
	G Allele	0.25	0.26	
Ataxia-telangiectasia mutated	No data	68 (25)	15 (21)	
	CC	200 (73)	53 (76)	0.319
rs1800058	CT	5 (2)	2 (3)	
c.4258 C>T	TT	0 (0)	0 (0)	
	HWE Fishers exact test	0.642		
	T Allele	0.01	0.02	
Ataxia-telangiectasia mutated	No data	53 (19)	13 (19)	
	GG	196 (72)	52 (74)	0.393
rs1801516	AG	23 (8)	5 (7)	
c.5557 G>A	AA	1 (0)	0 (0)	
	HWE	0.826		
	A Allele	0.06	0.04	

Values are the number of participants in each group (n) with percentages of total in parentheses. Samples were genotyped at the following sites: rs1800469 c.1347 C>T TGF-ß, rs1800472 c.788 C>TGF-ß, rs4880 c.47 C>T SOD2, rs25487 c.1196 A>G XRCC1, rs1800058 c.4258 C>T ATM and rs1801516 c.5557 G>A AT, in a Mixed Ancestry South African population. No data was used where either genotyping failed to produce a response, or DNA samples were not available from participant. Hardy-Weinburg Equilibrium via chi squared test requires the presence of all genotype combinations to calculate in significant numbers. Where noted, Fishers exact test was used. Significant values are highlighted in bold.

Abbreviations: HWE, Hardy-Weinburg Equilibrium.

3.5 Genotyping results

For 48 patient samples, genotyping were unsuccessful for all SNPs. This complete unresponsiveness to testing was consistent across repeated troubleshooting procedure results. In addition, the genes investigated are at different loci. In total 23 participants had missing clinical data for the SPADI scores. A total of 17 participants had missing data for both the questionnaire data and genotyping data and were therefore excluded from the analysis, which led to a final number of *n*=326 participants in the study. Table 3.5.1 shows the numbers of successful genotyped population for each SNP. Experimentally derived minor allele frequencies were generally in line with globally reported minor allele frequencies.

When genotyping was unsuccessful, steps were taken to resolve the issue. The procedure was repeated to account for operator error, and different batches of reagents were used. If this failed then the assay mixture was altered to increase the relative amounts of the key reagents in the reaction vessel. This involved increasing the total reaction volume to 10 µl, and doubling the volume of the probe and DNA samples within the reaction vessel. Finally, to maximise the PCR amplification, the number of PCR cycles were increased to increase the chance of a successful reaction.

Table 3.5.1 – Major and minor allele frequencies compared to the reported genotypes.

Gene symbol	SNP	Samples genotyped, n	Samples failed to amplify, n	Total genotyping yield, %	Reported MAF, %	Experimental MAF, %
TGF-ß	rs1800469	208	135	**61%**	33.9	36.0
TGF-ß	rs1800471	N/A	N/A	N/A	N/A	N/A
TGF-ß	rs1800472	276	67	**80%**	0.0	1.0
ATM	rs1800058	260	83	**76%**	1.3	1.0
ATM	rs1801516	277	66	**81%**	5.4	7.0
SOD2	rs4880	251	92	**73%**	39.2	41.0
XRCC1	rs25487	253	90	**74%**	25.3	26.0

Reported allele frequencies were taken from 1000 genome database, consisting of 2,504 individuals across 26 populations and were not specific to the studied population. Samples were genotyped at the following sites: rs1800469 c.1347 C>T TGF-ß, rs1800472 c.788 C>TGF-ß, rs4880 c.47 C>T SOD2, rs25487 c.1196 A>G XRCC1, rs1800058 c.4258 C>T ATM and rs1801516 c.5557 G>A AT, in a Mixed Ancestry South African population.

Abbreviations: MAF, minor allele frequency; SNP, single nucleotide polymorphism.

3.5.1 Rs1800471 - TGF-ß

The real-time PCR procedure for this polymorphism produced an unacceptably low assay success rate. Attempts were made to modify and repeat using appropriate troubleshooting procedures, which were unsuccessful. Based on this failure, the generated data were excluded from analysis as the data was not of high enough quality to ensure accuracy or specificity.

3.6 Hypothesis testing – case *vs* control

Genotyping data, shown in **Table 3.5.1**, were analysed with a two-way analysis of variance test to indicate if any of the results show a statistically significant association to severity of symptoms.

When the total cohort was tested for association between genotype and SPADI score, there was no evidence found within this study to support an association in any polymorphism, as shown in **Table 3.6.1**.

Table 3.6.1 – Inferred haplotype analysis as a proxy for gene-gene interactions was performed on; rs1800469 c.1347 C>T TGF-ß, rs1800472 c.788 C>TGF-ß, rs4880 c.47 C>T SOD2, rs25487 c.1196 A>G XRCC1, rs1800058 c.4258 C>T ATM and rs1801516 c.5557 G>A ATM polymorphisms in a South African population.

Haplotype										
Gene 1	**SNP 1**	**Gene 2**	**SNP 2**	**Gene 3**	**SNP 3**	**Df**	**Sum Sq**	**Mean square**	**F value**	**Pr, >F**
TGF-ß	rs1800469	TGF-ß	rs1800472			4	1786	447	0.904	0.463
TGF-ß	rs1800469	SOD2	rs4880			8	3344	418	0.846	0.563
TGF-ß	rs1800472	SOD2	rs4880			5	1449	290	0.586	0.711
TGF-ß	rs1800469	XRCC1	rs25487			8	2592	324	0.656	0.73
TGF-ß	rs1800472	XRCC1	rs25487			1	12	12	0.024	0.877
SOD2	rs4880	XRCC1	rs25487			7	2919	417	0.844	0.552
TGF-ß	rs1800469	ATM	rs1800058			4	880	220	0.445	0.776
TGF-ß	rs1800472	ATM	rs1800058			1	126	126	0.254	0.615
SOD2	rs4880	ATM	rs1800058			4	2338	585	1.18	0.319
XRCC1	rs25487	ATM	rs1800058			4	295	73.6	0.149	0.963
TGF-ß	rs1800469	ATM	rs1801516			4	1424	356	0.72	0.579
SOD2	rs4880	ATM	rs1801516			3	1761	587	1.19	0.315
XRCC1	rs25487	ATM	rs1801516			3	1404	469	0.947	0.419
ATM	rs1800058	ATM	rs1800058			1	461	461	0.932	0.335
TGF-ß	rs1800469	SOD2	rs4880	XRCC1	rs25487	8	2279	285	0.576	0.797
TGF-ß	rs1800469	SOD2	rs4880	ATM	rs1800058	2	101	50.7	0.103	0.902
SOD2	rs4880	XRCC1	rs25487	ATM	rs1800058	1	0	0.1	0	0.992
TGF-ß	rs1800469	SOD2	rs4880	ATM	rs1801516	3	401	134	0.27	0.847
TGF-ß	rs1800469	XRCC1	rs25487	ATM	rs1801516	2	1652	826	1.67	0.19
SOD2	rs4880	XRCC1	rs25487	ATM	rs1801516	1	28	28.2	0.057	0.811
					Residuals		12	470	39.2	

Abbreviations: SNP, single nucleotide polymorphism; DF, degrees of freedom; SPADI, shoulder pain and disability index; Sum Sq, sum of squares.

3.7 Hypothesis testing – low, medium, high case testing

In addition to the ANOVA testing of the cohort case *vs* control, testing was also performed on the case group, subdivided into low case (SPADI 0-29), medium case (SPADI 30-49) and high case (SPADI 50-100). This represents a low SPADI phenotype through to a stronger SPADI phenotype.

Within the low SPADI phenotype group, the haplotype of rare alleles at rs4880 and rs1800058 was strongly associated ($p<0.01$) with being in the low SPADI score group. None of the tested allele combinations were significantly associated with risk of medium or high SPADI phenotype, as shown in **Table 3.7.1**.

Table 3.7.1 - Inferred haplotype analysis as a proxy for gene-gene interactions was performed on; rs1800469 c.1347 C>T TGF-ß, rs1800472 c.788 C>TGF-ß , rs4880 c.47 C>T SOD2, rs25487 c.1196 A>G XRCC1, rs1800058 c.4258 C>T ATM and rs1801516 c.5557 G>A ATM polymorphisms in a South African population.

Low SPADI										
Haplotype										
Gene 1	**SNP 1**	**Gene 2**	**SNP 2**	**Gene 3**	**SNP 3**	**Df**	**Sum Sq**	**Mean Square**	**F Value**	**Pr (>F)**
TGF-ß	rs1800469	TGF-ß	rs1800472			3	267	89.1	1.53	0.208
TGF-ß	rs1800469	SOD2	rs4880			8	150	18.7	0.322	0.957
TGF-ß	rs1800472	SOD2	rs4880			4	136	34.1	0.587	0.673
TGF-ß	rs1800469	XRCC1	rs25487			7	534	76.3	1.31	0.247
TGF-ß	rs1800472	XRCC1	rs25487			1	48	48.5	0.83	0.362
SOD2	rs4880	XRCC1	rs25487			7	628	89.8	1.54	0.156
TGF-ß	rs1800469	ATM	rs1800058			3	331	110	1.9	0.132
TGF-ß	rs1800472	ATM	rs1800058			1	0	0.04	0.001	0.978
SOD2	rs4880	ATM	rs1800058			4	1011	252	4.35	**>0.01**
XRCC1	rs25487	ATM	rs1800058			2	103	51.7	0.889	0.413
TGF-ß	rs1800469	ATM	rs1801516			3	66	21.8	0.376	0.771
SOD2	rs4880	ATM	rs1801516			3	313	104	1.79	0.150
XRCC1	rs25487	ATM	rs1801516			3	286	95.2	1.64	0.182
ATM	rs1800058	ATM	rs1801516			1	219	219	3.78	**0.054**
TGF-ß	rs1800469	SOD2	rs4880	XRCC1	rs25487	8	770	96.2	1.66	0.113
TGF-ß	rs1800469	SOD2	rs4880	ATM	rs1800058	1	11	10.8	0.187	0.666
SOD2	rs4880	XRCC1	rs25487	ATM	rs1800058	1	2	1.67	0.029	0.866
TGF-ß	rs1800469	SOD2	rs4880	ATM	rs1801516	2	169	84.5	1.45	0.237
TGF-ß	rs1800469	XRCC1	rs25487	ATM	rs1801516	1	0	0.34	0.006	0.939
SOD2	rs4880	XRCC1	rs25487	ATM	rs1801516	1	56	55.5	0.955	0.330
					Residuals		**167**	**9706**	**58.1**	

Medium SPADI									
Haplotype									
Gene 1	**SNP 1**	**Gene 2**	**SNP 2**	**Gene 3**	**SNP 3**	**Df**	**Sum Sq**	**Mean Square**	**F Value**	**Pr (>F)**
TGF-ß	rs1800469	TGF-ß	rs1800472			1	0.6	0.59	0.015	0.904
TGF-ß	rs1800469	SOD2	rs4880			5	111	22.2	0.565	0.725
TGF-ß	rs1800469	XRCC1	rs25487			1	0.2	0.2	0.005	0.944
SOD2	rs4880	ATM	rs1800058			2	33.7	16.9	0.43	0.660
TGF-ß	rs1800469	ATM	rs1800058			2	60.5	30.2	0.772	0.484
TGF-ß	rs1800469	ATM	rs1801516			1	15.7	15.7	0.401	0.538
TGF-ß	rs1800469	SOD2	rs4880	XRCC1	rs25487	1	97.6	97.61	2.49	0.141
					Residuals	**2371**	**2371**	**2371**		

High SPADI									
Haplotype									
Gene 1	**SNP 1**	**Gene 2**	**SNP 2**			**Df**	**Sum Sq**	**Mean Square**	**F Value**	**Pr (>F)**
TGF-ß	rs1800469	SOD2	rs4880			3	163	54.3	0.252	0.859
TGF-ß	rs1800469	XRCC1	rs25487			3	356	119	0.55	0.659
SOD2	rs4880	XRCC1	rs25487			3	62.2	20.7	0.096	0.961
					Residuals	**2371**	**2371**	**2371**		

Abbreviations: SNP, single nucleotide polymorphism; DF, degrees of freedom; SPADI, shoulder pain and disability index; Sum Sq, sum of squares.

3.8 Summary

This study investigated the hypothesis that genetic variation in TGFß, ATM, SOD2 and XRCC1 genes may be associated with the inter-patient variation of SPADI symptoms observed after cancer treatment [36,155,183]. The underlying causative factors of these symptoms are not understood, and as the population of breast cancer survivors is likely to increase with improved medical care, it is likely that the prevalence of these symptoms will increase in the near future without intervention [1,3,16,79,81,82].

A literature review was conducted to compile and assess previously published data into incidence and prevalence of musculoskeletal and connective tissue dysfunction in cancer survivors [1,3,168]. From this, a list of potential candidate risk alleles were chosen at specific SNP sites within genes that function in the pathways of inflammation and wound healing, which lead to fibrosis [44,99,178–180,115,118,119,153,155,156,176,177]. Some of these candidate alleles might have only been present in a low frequency in the population, this represented an unknown factor due to the lack of genetic data on the studied populations. In addition the genetic component of one gene is unlikely to have a large affect [183]. The hypothesis was that within the population of breast cancer survivors, the genotype of these candidate alleles collectively would be associated with SPADI score. It is now possible to interpret the results to see if this hypothesis can be accepted, although proving causality is beyond the scope of this study.

In the next section, novel findings of this study will be linked to the wider biological and clinical environment. Possible future pathways for research and treatment will be explored, to build upon the results of this study, and other studies in the literature, and the strengths and limitations of this study will be reported, in order to ensure that results are taken with an appropriate level of confidence.

Chapter 4 - Discussion

4.1 Novel findings

4.1.1 Demographics

When participant age was compared with symptom severity, it was found that the arithmetic age mean for the cases and controls were different, with an average age of 54.4 years in the control group vs an average age of 50.7 years in the case group. This indicates that younger participants were more likely to have moderate-to-severe symptoms than older participants, this finding is in agreement with many other reports of the effects age in similar conditions [202–204]. There is currently no consensus for the possible reasons. It is possible that there is a biological cause, but it is also possible that there are confounding factors. There may be a difference between age groups in the propensity to either report, or to not report, the same level of pain and disability. There may also be socio-economic factors, individuals with high levels of access to care may be diagnosed and treated at an earlier stage, requiring less severe treatment, and therefore be more likely to survive with minor symptoms to an older age.

4.1.2 Single nucleotide polymorphisms

When the allele frequencies were compared for cases *vs* controls in **Table 3.4.1**, the frequency of minor alleles for *SOD2* rs4880 (C>T) was found to be significantly deviated from the normal distribution. When allele frequencies in the case and control groups were calculated, it was found that the T allele is over-represented in the case group compared to the control group, with a frequency of 0.44 *vs* 0.38 respectively.

The *SOD2* gene, and the rs4880 genetic polymorphism site, were chosen for investigation in this study following associations with pathology reported in several papers [159,184,186,198]. The genetic dimorphism at rs4880 is a C>T change, which confers a codon shift from alanine to valine. There is debate in the literature as to the impact of the change of alanine to valine at codon 16, which functions in the mitochondrial targeting sequence (MTS), which facilitates the transport of the precursor enzyme across the mitochondrial membranes [160,205,206]. In one paper, that alanine form of the SOD2 MTS has been shown to have a 40% greater enzymatic activity than the valine

form [186,206,207]. Increased *SOD2* activity is associated with an anti-oxidative, anti-fibrotic effect in most papers, and this would be supported by the finding of an increased minor allele frequency of T alleles in the case group. There are conflicting reports however, as one study has reported 30% higher activity in CT or TT SOD2 *vs* CC at codon 16 [208]. In addition, a study of patients undergoing RT for prostate cancer found that CT, but not CC or TT genotypes, was associated with 8% increased risk of severe late rectal bleeding [184]. Whilst a general trend towards an increase of function of SOD2 seems to be associated with a beneficial increased anti-oxidant functionality, more investigation will clearly be needed to elucidate the effect of variation at rs4880.

4.1.3 Inferred haplotypes

No intragenic inferred haplotypes were found, however, when inferred haplotypes were constructed as a proxy for gene-gene interactions, an allele combination was identified between *SOD2* rs4880 (C>T) and *ATM* rs1800058 (C>T) and the low SPADI group, indicating that the interaction between the variant forms of these gene is associated with a low SPADI score in this sample population.

These two specific polymorphisms were selected for genotyping in this study because both of the gene products interact in pathways that can modulate the formation of fibrotic tissue. Because of this, it was anticipated that in an individual with multiple rare genotypes, we would possibly see an amplified effect of the individual polymorphisms. As previously mentioned, rs4880 C>T is understood to contribute to a decrease of function of SOD2, so with this model of function, we would possibly not have expected the rare allele to be statistically inferred in relation to low SPADI score [44,160,205]. When interpreting these results, it is important to remember the limits of this study, the conflicted nature of reports, and the clear need for more investigation of the function of *SOD2* and genetic variation within.

The genetic variation rs1800058 in *ATM* has been investigated in several radiosensitivity studies, including in other cancers such as prostate cancer [178,209,210]. A large proportion of genetic variants in *ATM* (70%) are reported to be truncating or otherwise deleterious, and therefore linked with lower enzymatic activity [178,211]. Certain variants have also been linked to an increase of breast cancer risk in AT allele carriers

[178,211]. Individuals with major loss of function of *ATM* have catastrophic vulnerability to ionising radiation, leading to extreme radiosensitivity [178,196,212]. In the general European population, 0.5-1% of the population are heterozygous carriers of at least one *ATM* variant allele [211,213,214]. This haplo-insufficiency is associated with increased sensitivity to radiation [185,211].

In contrast to many of these findings in *ATM*, rs1800058 has been identified as having a protective effect from fibrosis. In a paper by Edvardsen *et al.* (2007) studying *ATM* allele frequencies in a population of breast cancer patients treated with RT, a decrease in the minor allele frequency at rs1800058 was found to be associated with an increasing level of adverse side effects for late radiation tissue injury, thus suggesting a protective effect of the rare allele. These findings are in concordance, and suggest that there may be a protective effect of this gene-gene interaction [178]. We therefore recommend further investigation of these combinations in a larger case control study.

Finally, there was another observation which fell below the allowed significant level, between rs1800058 and rs1801516 (G>A) both in the ATM gene, again with a low SPADI score. In the literature, rs1801516 has been investigating in a larger meta-analysis (2000 patients in nine studies) for risk of RIF, in which presence of the minor allele was associated with a significantly increased risk of RIF, with high-RIF patients having an OR for minor allele presence of 3.19 *vs* 1.09 [50,176,211,213]. This finding contradicts this, as the two SNPs linked have reports of contradictory effects (rs1800058 variant alleles are believed to be protective, while rs1801516 variants are harmful). This, along with a significance level below the accepted limit, indicates that this finding may not be reliable, and is not suggested for further study at this stage.

4.1.4 Novel findings summary

Within this section, the observed results were explored in the context of the previously existing literature, and biological context. Polymorphisms within two genes, *SOD2* and *ATM,* have been identified as recommended for future study, based on associations identified within the results. *SOD2* and *ATM* genes both function within the biological processes that allow tissue to repair and return to homeostasis after cellular insult. The possible significance of these variants in relation to their biological function has

been explored. These results must be taken within the overall limited strength of this study. Suggesting a causal link between these genetic variants and large scale effects, is beyond the power of this study. More research is needed, but these findings can suggest these variants for future study.

No significant results were noted for the polymorphic sites within *TGF-ß.* Given that *TGF-ß* gene activity is proposed to be one of the major factors in the pathophysiology of radiotherapy-induced fibrosis, this was unexpected. This is a pilot study of limited power, and it is extremely possible that any possible associations may become apparent in studies of greater power. It is also possible that the many reports of *TGF-ß* associations are spurious, or merely highly contextual. It is impossible to conclude further due to the many unknown factors at play.

4.2 Future research

As discussed in the literature review, there are significant gaps in both understanding and treatment options of RIF and other cancer treatment related morbidities. These knowledge gaps exist in many areas, including the underlying biology, epidemiology and possible treatment options. Future research in all fields will endeavour to increase our ability to treat these conditions, and prevent occurrence in the first place.

4.2.1 Population genetic data

One finding of this paper is that genetic research is underrepresented in economically poorer populations. When attempts were made to find large-scale, comprehensive genetic data for the Mixed Ancestry South African population, none was found [215–217]. The vast majority of genome-wide studies are based in European populations, and even within the few African population based studies, there is an unacceptable amount of heterogeneity in recording of descriptions, with large geographic regions being combined into one result [215]. This issue is compounded by the genetic heterogeneity within the Mixed Ancestry South African population itself. This group is genetically diverse, with a genetic background widely containing four major distinct groups, Europeans, South Asians, Indonesians, Xhosa, and many more in smaller proportions [188,215]. When considering that this population accounts for up to five million South Africans, it is important that this imbalance is resolved, and the genomic data of this, and other understudied groups, especially in Africa, is compiled. Without this, these population groups will not be able to benefit equally from improvements in personalised and precision medicines. Some projects have started to improve the situation, such as the African Genome Variation Project (AGVP), and the Human Hereditary and Health (H3A) project, the latter of which is currently processing the genome of 70,000 participants across the continents, accompanied by detailed clinical information banks [215,216].

4.2.2 Genomics and precision medicine

In this study, a candidate gene approach was determined to be the most suitable approach. While this is currently a standard research approach, it is expected that in future genome-wide approaches will become more suitable as costs decrease and

access to data analysis tools increase. Currently, the cost and complexity of running genome wide analysis (GWAS) makes this tool unavailable or unfeasible for many research teams. This is changing as technology progresses and costs are decreasing. As this happens, and as the impact of projects such as H3A start to produce results and due to the huge amount of data produced by GWAS there is also an increasing need for reliable analysis tools and comprehensive review publications in genomics [119,218–220]. One example is the H3A's recently developed SNP genotyping array, based on the African data from the 1000 genome project, the AGVP project, and several smaller projects, with 1.7 million SNP's chosen as base content to cover the African genome, and ~700,000 SNPs of specific interest, requested by project collaborators in H3A [221]. This project is backed through funding to upgrade testing sites across the continent to include high throughout sample processing [221]. Consisting of 500 members, in 30 out of the 55 African countries, H3A has the potential to contribute to a generational shift in genomic research in South Africa and the wider continent.

4.2.3 Treatments

When the genetic and biological causes of SPAD are better understood, treatment regimens can be tested and eventually incorporated into the treatment regimens. Alongside treatment, preventative measures will also be introduced, such as modulation of breast cancer treatment protocols to minimise the potential risk in those who would be more susceptible to fibrosis [222].

Within the literature review, fibrosis has been heavily suggested as being one of the key pathological processes in post cancer treatment sequelae. Fibrosis is heavily driven by inflammatory factors, and therefore it is likely that targeting the inflammatory process will feature in future treatment plans, anti-inflammatory therapies could prevent large scale accumulation of inflammatory cells in the area, halting the fibrotic process before it begins. As previously noted, the lack of findings within the TGF-β gene was disappointing, due to the many reported associations in the literature [41,47,115,142,161,181,197,223]. Despite this, targeting of the TGF-β axis directly has been one of the first studied as a potential treatment for fibrotic disease and RIF [143]. Due to the highly contextual and environmentally sensitive nature of TGF-ß protein activity, this

has proven difficult to achieve [128–131]. It is important to remember that any large scale loss of function of the *TGF-β* pathway is catastrophically damaging [224].

4.3 Strengths and limitations

As with all research, this dissertation has areas of strength and areas of limitation.

4.3.1 Demographics and clinical characteristics

The demographic and clinical information for the cohort was limited in several ways. Patients were recruited from the Mixed Ancestry population of South Africa (self-reported). Due to this, bioinformatic databases frequently lacking information on the expected allele frequencies of this group. As a result of this, there was no reliable estimation of expected minor allele frequencies available.

In addition, participants were not subdivided into groups based on treatment modality. All participants had some combination of one, or multiple treatments, with the main modalities of RT, CT, HT and surgery. Ideally this study would have isolated participants by treatment modality group, in order to exclude any effect that different treatments might have upon the results. Unfortunately, as most patients had at least two modalities this was not possible, and there was already a consideration that had to be taken as to the reduction in study power by reducing the cohort size by that degree.

4.3.2 Data collection

Accurate and precise data collection is essential to any study. Much of the data collected from participants was self-reported. Self-reported data is not inherently a weakness in appropriate uses, for example when assessing subjective pain levels. However, use of self-reported data such as medical history is generally undesirable, and prone to inaccuracies due to cognitive flaws and memory loss. This was unfortunately unavoidable in some cases, as detailed, accurate and accessible medical history is not a guarantee for many individuals.

The amount of missing data for collected data was significant. This can be attributed to a number of factors, including the large number of recorded variables, the socio-economic factors of the population studied, and shortcomings within the South African medical system. Due to the under-served nature of this disease area, all participants

were included where possible, therefore a complete case analysis was not used. As this data was collected prior to this study, this shortcoming could not be addressed within this study. Despite this, the statistical tools and significance levels required for this study ensure that results are reliable despite this.

Within the sample, 79 participants reported a SPADI score of zero, indicating that they felt no pain or disability, at the time of recording (one-year post surgery). The questionnaire used to provide this score was the SPADI questionnaire, a self-reported questionnaire completed by all participants. As this question-based testing is highly subjective, there is a possibility that an individual's inclination towards reporting disability or changes in quality of life as a matter of personal preference could be skewing the results. This introduces a degree of subjectivism and therefore reduces the accuracy and precision of the measured variables within the questionnaire results. The questionnaire is partially retrospective, requiring recall and memory of pain and disability from the year prior to the testing interview, which is somewhat unreliable, as memory of pain can be unreliable. There is also the possibility of cultural differences between the researchers and population, with respect to what an individual might describe as 'hardship' or 'disability. Finally, while there was provision for non-English speakers, there may have been inadvertent ambiguity in the language used by both researchers and participants, including those for whom English is not the first language. This subjective recording of symptom severity from the questionnaires was used to designate a control group which either suffered no or little SPADI symptoms. This is a noted difficulty within case/control studies.

One method of improving future research would be to develop a more objective method of measurement of symptom severity for future research into post-cancer treatment late tissue morbidity. This could be in the form of a battery of impartial measurement of symptoms against objective measurement of pain and disability severity. Another possibility could be staggered screenings over the time period of recovery (e.g every three months up to two years post treatment) in order to provide accurate (at time of recording) measurements as well as multiple data-points longitudinally.

4.3.3 Methodology

During the laboratory testing, the genotyping success rates for several of the assays were low. This was evident during the analysis as a significant number of samples failed to be genotyped at all polymorphisms studied. Low error rates are claimed by all manufacturers and probes have high specificity with functionally proven testing for the DNA sequences targeted. Despite this, the reasons for a sample failing to be amplified are numerous and can be difficult to determine. Molecules present within reagents, or in the samples themselves could hinder amplification, the problem could also be within the DNA samples; the DNA is either no longer intact or no longer responding to testing due to degradation. Regardless of the cause, repeated genotyping of failed samples can be known to produce false results. Because of this risk, samples that failed after troubleshooting (described in section 3.5) were discarded from further analyses.

It has been acknowledged that only a relatively small number of genetic loci were investigated within this study, this is partly due to constraints on the project in terms of time and other resources. Another constraint is that this research is still a relatively new area of research, with fairly little published literature, which limits the number of candidate genetic testing targets identified within the literature. In addition, using a candidate gene approach limited the research to exploring only gene loci that had been previously investigated. These constraints are however to be expected within the constraints of an MSc thesis, and are intended to be taken within the context of a larger research body. In addition, variant loci were often identified in a drastically different population, commonly Caucasian, to those being tested, and as such it is always possible that significant genetic loci may not be relevant in a different population due to genotypic variation.

Finally, HWE quality control was significantly breached for the rs1800469 polymorphism. Deviation from HWE can indicate problems such as selection bias within the group, population stratification and genotyping errors.

One area of strength is that the hypothesis focussed approach was followed. This project is a sub-study of a larger ongoing project (HREC 312-2012), and the literature

review section presented research from this main study group. This body of research was used as basis of the review, which produced a hypothesis and aims for the experimental section.

This provides guidelines for genetic association studies, including guidelines on ethical approval, patient recruitment and carrying out laboratory work. The statement also informs how to conduct unbiased and accurate reporting of the results, with transparency. Following these guidelines, multiple samples were repeated within the study to serve as positive and negative controls whilst blanking samples, containing no DNA were also included per 96 well plate run. The laboratory protocol and results were checked by senior team members.

One of the major areas of strength of the study is the case control design. The case-control candidate genetic association study design was conducted in accordance with the guidelines provided by the STREGA statement. This provides guides for genetic association studies, including guidelines on ethical approval, patient recruitments (when applicable) and carrying out laboratory work. This also includes unbiased and accurate reporting of the results, with transparency. Multiple samples were repeated within the study to serve as positive controls, and negative controls containing no DNA samples were also includes per 96-well plate run. All attempts were made to follow the principles and guidelines of the STREGA guidelines for quality control purposes. The laboratory protocol and results were checked by senior team members.

All attempts were made to follow the principles and guidelines of the STREGA guidelines for quality control purposes. However, as with all novel findings, the associations highlighted in this study need to be investigated in greater detail, in larger studies, in order to confirm all the observed results.

4.4 Summary and concluding remarks

As discussed in the first chapter, breast cancer is a common and well treated form of cancer. A consequence of this increased post-cancer survivor population is increased prevalence of cancer-treatment related conditions. These include a subset of cancer survivors, who develop musculoskeletal dysfunction in the shoulder and upper limb after having received radiotherapy (and other treatments) for breast cancer. There is a high degree of inter-group difference, both in the range, and severity of symptoms exhibited. This range in symptoms is not explained by currently available literature.

In the literature review a hypothesis was developed that genetic signalling pathway dysfunction leading to fibrosis in these tissues could be an instigating biochemical cause of these symptoms. It was hypothesised that genetic variation between these individuals could explain why some people seem more susceptible to this fibrosis than others. A list of candidate influential genetic variant SNP sites was developed from past research, and a candidate gene genetic association study was decided upon as the most suitable study protocol.

Table 4.4.1 – A list of the genes selected for inclusion within the study, and the specific genetic polymorphism under investigation. Samples were genotyped at the; rs1800469 c.1347 C>T TGF-ß, rs1800472 c.788 C>TGF-ß , rs4880 c.47 C>T SOD2, rs25487 c.1196 A>G XRCC1, rs1800058 c.4258 C>T ATM and rs1801516 c.5557 G>A ATM polymorphisms in Mixed Ancestry South African population.

Gene	SNP
TGF-ß	rs1800469
TGF-ß	rs1800471
TGF-ß	rs1800472
ATM	rs1800058
ATM	rs1801516
SOD2	rs4880
XRCC1	rs25487

Abbreviations: SNP, single nucleotide polymorphism.

After testing, the experimental protocol for one SNP variant presented a low yield and data was discarded, leaving valid data for six SNPs. Statistical analysis was then performed on the results to determine if any results were significant.

When the age at surgery was compared to self-reported SPADI scores, it was found that younger participants were more likely to have moderate-to-severe symptoms than

older participants, which is concurrent with many similar reports [203,204]. This study provides evidence reinforcing the previous reports.

When the genotype data of the study population was directly compared to severity of symptoms, no single SNP was shown to be associated with symptom severity. However, when the allele frequencies were compared for case *vs* control, the distribution of frequency of minor alleles of rs4880 was found to be significantly deviated from the expected distribution. This is novel evidence suggesting that variation in the MTS section of *ATM* may be associated with the severity of an individual's response to cancer treatment.

When inferred haplotypes were constructed as a proxy for gene-gene interactions, a significant difference in haplotype distribution was found between *SOD2* rs4880 and *ATM* rs1800058, while a borderline significant difference was found between *ATM* rs1800058 and *ATM* rs1801516. Further to the previous finding, this is additional novel evidence that the ATM and SOD2 genes may be linked to modulating the effects of cancer treatment cyto-toxicity.

To summarise, this study provides novel evidence suggesting that the *SOD2* and *ATM* genes, and the cyto-protective and free-radical scavenging and DNA repair pathways, should be targets for future studies into breast cancer treatment morbidity.

References

1. Shamley, D. R. *et al.* Changes in shoulder muscle size and activity following treatmentfor breast cancer. *Breast Cancer Res. Treat.* **106**, 19–27 (2007).
2. K Paulssen, D. S. Axillary Web Syndrome (AWS) after treatment for breast cancer: (University of Cape Town, 2015).
3. Shamley, D., Lascurain-Aguirrebeña, I. & Oskrochi, R. Clinical anatomy of the shoulder after treatment for breast cancer. *Clin. Anat.* **27**, 467–477 (2014).
4. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN2008. *Int. J. Cancer* **127**, 2893–2917 (2010).
5. Harrington, K. J. Biology of cancer. in *Medicine* **39**, 797–876 (Garland Science; 2nd edition, 2011).
6. Koroltchouk, V., Stanley, K. & Stjernswärd, J. The control of breast cancer a World Health Organization perspective. *Cancer* **65**, 2803–2810 (1990).
7. Vorobiof, D. A., Sitas, F. & Vorobiof, G. Breast cancer incidence in South Africa. *J. Clin. Oncol.* **19**, 125S–127S (2001).
8. Moodley, J., Cairncross, L., Naiker, T. & Momberg, M. Understanding pathways to breast cancer diagnosis among women in the Western Cape Province, South Africa: a qualitative study. *BMJ Open* **6**, e009905 (2016).
9. National Institute for Occupational Health. National Cancer Registry. *NIOH NCR - 2010 factsheet* (2010). Available at: https://www.cansa.org.za/files/2015/10/NCR_Final_2010_tables1.pdf.
10. WHO. GLOBOCAN 2018 fact sheet, South Africa. *GLOBOCAN* (2018). Available at: https://gco.iarc.fr/today/data/factsheets/populations/710-south-africa-fact-sheets.pdf. (Accessed: 1st May 2020)
11. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
12. PDQ Adult Treatment Editorial Board. Breast Cancer Treatment (Adult) (PDQ®): Patient Version. *PDQ Cancer Information Summaries* (2002). Available at: http://www.ncbi.nlm.nih.gov/pubmed/26389406. (Accessed: 12th May 2020)
13. Sighoko, D. *et al.* Breast cancer in pre-menopausal women in West Africa:

Analysis of temporal trends and evaluation of risk factors associated with reproductive life. *Breast* **22**, 828–835 (2013).

14. Joko-Fru, W. Y. *et al.* The evolving epidemic of breast cancer in <scp>sub-Saharan</scp> Africa: Results from the African Cancer Registry Network. *Int. J. Cancer* ijc.33014 (2020). doi:10.1002/ijc.33014
15. Kelsey, J. L., Gammon, M. D. & John, E. M. Reproductive Factors and Breast Cancer. *Epidemiol. Rev.* **15**, 36–47 (1993).
16. Nojilana, B. *et al.* Emerging trends in non-communicable disease mortality in South Africa, 1997 - 2010. *S. Afr. Med. J.* **106**, 477–484 (2016).
17. Coovadia, H., Jewkes, R., Barron, P., Sanders, D. & McIntyre, D. The health and health system of South Africa: historical roots of current public health challenges. *The Lancet* **374**, 817–834 (2009).
18. McIntyre, D. A critical analysis of the current South African health system. (2007).
19. Sineshaw, H. M. *et al.* Black/White Disparities in Receipt of Treatment and Survival Among Men With Early-Stage Breast Cancer. *J. Clin. Oncol.* **33**, 2337–2344 (2015).
20. Plesca, M., Bordea, C., El Houcheimi, B., Ichim, E. & Blidaru, A. Evolution of radical mastectomy for breast cancer. *Journal of medicine and life* **9**, 183–186 (2016).
21. Cotlar, A. M., Dubose, J. J. & Rose, D. M. M. History of surgery for breast cancer: Radical to the sublime. *Curr. Surg.* **60**, 329–337 (2003).
22. Wright, J. R. The radicalization of breast cancer surgery: Joseph Colt Bloodgood's role in William Stewart Halsted's legacy. *Bull. Hist. Med.* **92**, 141–171 (2018).
23. Franceschini, G. *et al.* New trends in breast cancer surgery: a therapeutic approach, increasingly efficacy and respectful of the patient. *G. Chir.* **36**, 145–152 (2015).
24. Zurrida, S. *et al.* The Changing Face of Mastectomy (from Mutilation to Aid toBreast Reconstruction). *Int. J. Surg. Oncol.* **2011**, 980158 (2011).
25. Skowronek, J. & Chicheł, A. Brachytherapy in breast cancer: an effective alternative. *Menopausal Rev.* **1**, 48–55 (2014).
26. Lehman, M. *et al.* Partial breast irradiation for early breast cancer. in *Cochrane*

Database of Systematic Reviews (John Wiley & Sons, Ltd, 1996).

27. Mukherjee, D., Coates, P. J., Lorimore, S. A. & Wright, E. G. Responses to ionizing radiation mediated by inflammatory mechanisms. *J. Pathol.* **232**, 289–299 (2014).
28. Céline Bourgier Antonin Levy Marie-Catherine Vozenin & Deutsch, E. Pharmacological strategies to spare normal tissues from radiation damage: useless or overlooked therapeutics? - Springer. *Cancer Metastasis Rev.* **31**, 699–712 (2012).
29. Delanian, S., Lefaix, J.-L. & Pradat, P.-F. Radiation-induced neuropathy in cancer survivors. *Radiother. Oncol.* **105**, 273–282 (2012).
30. Hille-Betz, U. *et al.* Late radiation side effects, cosmetic outcomes and pain in breast cancer patients after breast-conserving surgery and three-dimensional conformal radiotherapy: Risk-modifying factors. *Strahlenther. Onkol.* **192**, 8–16 (2016).
31. Bentzen, S. M., Saunders, M. I., Dische, S. & Bond, S. J. Radiotherapy-related early morbidity in head and neck cancer: quantitative clinical radiobiology as deduced from the CHART trial. *Radiother. Oncol.* **60**, 123–35 (2001).
32. Bentzen, S. M., Thames, H. D. & Overgaard, M. Latent-time estimation for late cutaneous and subcutaneous radiation reactions in a single-follow-up clinical study. *Radiother. Oncol.* **15**, 267–74 (1989).
33. Bentzen, S. M. *et al.* Normal tissue effects: reporting and analysis. *Semin. Radiat. Oncol.* **13**, 189–202 (2003).
34. Pohlers, D. *et al.* TGF-β and fibrosis in different organs — molecular pathway imprints. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1792**, 746–756 (2009).
35. Rubin, P., Johnston, C. J., Williams, J. P., McDonald, S. & Finkelstein, J. N. A perpetual cascade of cytokines post irradiation leads to pulmonary fibrosis. *Int. J. Radiat. Oncol.* **33**, 99–109 (1995).
36. Vujaskovic, Z. *et al.* Radiation-induced hypoxia may perpetuate late normal tissueinjury. *Int. J. Radiat. Oncol. Biol. Phys.* **50**, 851–855 (2001).
37. Lingos, T. I. *et al.* Radiation pneumonitis in breast cancer patients treated with conservative surgery and radiation therapy. *Int. J. Radiat. Oncol.* **21**, 355–360 (1991).
38. Dadrich, M. *et al.* Combined inhibition of TGFβ and PDGF signaling attenuates

radiation-induced pulmonary fibrosis. *Oncoimmunology* **5**, (2016).

39. Maturen, K. E. *et al.* Imaging Effects of Radiation Therapy in the Abdomen and Pelvis:Evaluating ``Innocent Bystander'' Tissues. *Radiographics* **33**, 599–619 (2013).
40. Andreassen, C. N. *et al.* ATM sequence variants and risk of radiation-induced subcutaneous fibrosis after postmastectomy radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.* **64**, 776–783 (2006).
41. Yarnold, J. & Brotons, M.-C. V. Pathogenetic mechanisms in radiation fibrosis. *Radiother. Oncol.* **97**, 149–161 (2010).
42. Tsoutsou, P. G. & Koukourakis, M. I. Radiation pneumonitis and fibrosis: mechanisms underlying its pathogenesis and implications for future research. *Int. J. Radiat. Oncol. Biol. Phys.* **66**, 1281–1293 (2006).
43. Andreassen, C. N., Alsner, J., Overgaard, M. & Overgaard, J. Prediction of normal tissue radiosensitivity from poly morphisms in candidate genes. *Radiother. Oncol.* **69**, 127–135 (2003).
44. Córdoba, E. E., Abba, M. C., Lacunza, E., Fernánde, E. & Güerci, A. M. Polymorphic Variants in Oxidative Stress Genes and AcuteToxicity in Breast Cancer Patients Receiving Radiotherapy. *Cancer Res. Treat.* **48**, 948–954 (2016).
45. Funke, S. *et al.* Genetic Polymorphisms in Genes Related to Oxidative Stress (*GSTP1, GSTM1, GSTT1, CAT, MnSOD, MPO, eNOS*) and Survival of Rectal Cancer Patients after Radiotherapy. *J. Cancer Epidemiol.* **2009**, 1–6 (2009).
46. Bentzen, S. M. Preventing or reducing late side effects of radiation therapy: radiobiology meets molecular pathology. *Nat. Rev. Cancer* **6**, 702–713 (2006).
47. Andreassen, C. N., Alsner, J. & Overgaard, J. Does variability in normal tissue reactions after radiotherapy have a genetic basis – where and how to look for it? *Radiother. Oncol.* **64**, 131–140 (2002).
48. Bakiu, E., Telhaj, E., Kozma, E., Ru\cci, F. & Malkaj, P. Comparison of 3D CRT and IMRT Tratment Plans. *Acta Inform. Med.* **21**, 211–212 (2013).
49. Kerns, S. L., Ostrer, H. & Rosenstein, B. S. Radiogenomics: using genetics to identify cancer patients at risk for development of adverse effects following radiotherapy. *Cancer Discov.* **4**, 155–65 (2014).
50. Andreassen, C. N. *et al.* Individual patient data meta-analysis shows a

significantassociation between the ATM rs1801516 SNP and toxicityafter radiotherapy in 5456 breast and prostate cancer patients. *Radiother. Oncol.* (2016).

51. Rogers, P. B., Plowman, P. N., Harris, S. J. & Arlett, C. F. Four radiation hypersensitivity cases and their implications for clinical radiotherapy. *Radiother. Oncol.* **57**, 143–154 (2000).
52. Alter, B. P. Radiosensitivity in Fanconi's anemia patients. *Radiother. Oncol.* **62**, 345–347 (2002).
53. Gatti, R. A. The Inherited Basis of Human Radiosensitivity. *Acta Oncol. (Madr).* **40**, 702–711 (2001).
54. Hill, R. ., Rodemann, H.-P., Hendry, J. ., Roberts, S. . & Anscher, M. . Normal tissue radiobiology: from the laboratory to the clinic. *Int. J. Radiat. Oncol.* **49**, 353–365 (2001).
55. Chang-Claude, J. *et al.* Genetic polymorphisms in DNA repair and damage response genes and late normal tissue complications of radiotherapy for breast cancer. *Br. J. Cancer* **100**, 1680–1686 (2009).
56. Allan, J. M. & Travis, L. B. Mechanisms of Therapy-Related Carcinogenesis. *Nat. Rev. Cancer* **5**, 943–955 (2005).
57. Dasari, S. & Tchounwou, P. B. Cisplatin in cancer therapy: molecular mechanisms of action. *Eur. J. Pharmacol.* **740**, 364–78 (2014).
58. Yao, X., Panichpisal, K., Kurtzman, N. & Nugent, K. Cisplatin Nephrotoxicity: A Review. *Am. J. Med. Sci.* **334**, 115–124 (2007).
59. Hayati, F., Hossainzadeh, M., Shayanpour, S., Abedi-Gheshlaghi, Z. & Beladi Mousavi, S. S. Prevention of cisplatin nephrotoxicity. *J. nephropharmacology* **5**, 57–60 (2016).
60. Miller, R. P., Tadagavadi, R. K., Ramesh, G. & Reeves, W. B. Mechanisms of Cisplatin nephrotoxicity. *Toxins (Basel).* **2**, 2490–518 (2010).
61. Kim, P. Y. & Johnson, C. E. Chemotherapy-induced peripheral neuropathy. *Curr. Opin. Anaesthesiol.* **30**, 570–576 (2017).
62. Kirchmair, R. *et al.* Antiangiogenesis Mediates Cisplatin-Induced Peripheral Neuropathy: Attenuation or Reversal by Local Vascular Endothelial Growth Factor Gene Therapy Without Augmenting Tumor Growth. *Circulation* **111**, 2662–2670 (2005).

63. Farquhar-Smith, P. Chemotherapy-induced neuropathic pain. *Curr Opin Support Palliat. Care* **5**, 1–7 (2011).
64. Tremont, A., Lu, J. & Cole, J. T. Endocrine Therapy for Early Breast Cancer: Updated Review. *Ochsner J.* **17**, 405–411 (2017).
65. Harper, M. J. & Walpole, A. L. A new derivative of triphenylethylene: effect on implantation and mode of action in rats. *J. Reprod. Fertil.* **13**, 101–19 (1967).
66. Robert, N. J. Clinical efficacy of tamoxifen. *Oncology (Williston Park).* **11**, 15–20 (1997).
67. Chim, K. *et al.* Joint pain severity predicts premature discontinuation of aromatase inhibitors in breast cancer survivors. *BMC Cancer* **13**, 401 (2013).
68. Moscetti, L. *et al.* Adjuvant aromatase inhibitor therapy in early breast cancer: what factors lead patients to discontinue treatment? *Tumori J.* **101**, 469–473 (2015).
69. Nestoriuc, Y. *et al.* Is it best to expect the worst? Influence of patients' side-effect expectations on endocrine treatment outcome in a 2-year prospective clinical cohort study. *Ann. Oncol.* **27**, 1909–1915 (2016).
70. Kojima, Y. *et al.* Autocrine TGF-beta and stromal cell-derived factor-1 (SDF-1) signaling drives the evolution of tumor-promoting mammary stromal myofibroblasts. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20009–20014 (2010).
71. Caban, M. E. *et al.* The relationship between depressive symptoms and shoulder mobility among older women: assessment at one year after breast cancer diagnosis. *Clin. Rehabil.* **20**, 513–522 (2006).
72. Rietman, J. S. *et al.* Late morbidity after treatment of breast cancer in relation todaily activities and quality of life: a systematic review. *Eur. J. Surg. Oncol.* **29**, 229–238 (2003).
73. Levangie, P. K. & Drouin, J. Magnitude of late effects of breast cancer treatments on shoulder function: a systematic review. *Breast Cancer Res. Treat.* **116**, 1–15 (2009).
74. Hidding, J. T. *et al.* Treatment Related Impairments in Arm and Shoulder in Patients with Breast Cancer: A Systematic Review. *PLoS One* **9**, e96748 (2014).
75. Lee, T. S., Kilbreath, S. L., Refshauge, K. M., Herbert, R. D. & Beith, J. M. Prognosis of the upper limb following surgery and radiation for breast cancer. *Breast Cancer Res. Treat.* **110**, 19–37 (2008).

76. Tengrup, I., Tennvall-Nittby, L., Christiansson, I. & Laurin, M. Arm morbidity after breast-conserving therapy for breast cancer. *Acta Oncol.* **39**, 393–397 (2000).
77. Johansen, S., Fosså, K., Nesvold, I. L., Malinen, E. & Fosså, S. D. Arm and shoulder morbidity following surgery and radiotherapy for breast cancer. *Acta Oncol.* **53**, 521–529 (2014).
78. Kärki, A., Simonen, R., Mälkiä, E. & Selfe, J. Impairments, activity limitations and participation restrictions 6 and 12 months after breast cancer operation. *J. Rehabil. Med.* **1**, 1 (2004).
79. Moskovitz, A. H. *et al.* Axillary web syndrome after axillary dissection. *Am. J. Surg.* **181**, 434–439 (2001).
80. Torres Lacomba, M. *et al.* Axillary web syndrome after axillary dissection in breast cancer: a prospective study. *Breast Cancer Res. Treat.* **117**, 625–630 (2009).
81. Koehler, L. A. Axillary Web syndrome ongoing medical evaluation. (University of Minnesota, 2013).
82. Fourie, W. J. & Robb, K. A. Physiotherapy management of axillary web syndrome following breast cancer treatment: discussing the use of soft tissue techniques. *Physiotherapy* **95**, 314–320 (2009).
83. Dalberg, K. *et al.* A randomised study of axillary drainage and pectoral fascia preservation after mastectomy for breast cancer. *Eur. J. Surg. Oncol.* **30**, 602–609 (2004).
84. Salmon, R. J., Berry, M. & Hamelin, J.-P. A novel treatment for postoperative mondor's disease: manualaxial distraction. *Breast J.* **15**, 381–384 (2009).
85. Hagiwara, Y. *et al.* Coexistence of fibrotic and chondrogenic process in the capsule of idiopathic frozen shoulders. *Osteoarthr. Cartil.* **20**, 241–249 (2012).
86. Hand, C., Clipsham, K., Rees, J. L. & Carr, A. J. Long-term outcome of frozen shoulder. *J. Shoulder Elb. Surg.* **17**, 231–236 (2008).
87. Hakim, A. J., Cherkas, L. F., Spector, T. D. & MacGregor, A. J. Genetic associations between frozen shoulder and tennis elbow: a female twin study. *Rheumatology (Oxford).* **42**, 739–42 (2003).
88. Hand, G. C. R., Athanasou, N. A., Matthews, T. & Carr, A. J. The pathology of frozen shoulder. *J. Bone Joint Surg. Br.* **89-B**, 928–932 (2007).
89. Robinson, C. M., Seah, K. T. M., Chee, Y. H., Hindle, P. & Murray, I. R. Frozen

shoulder. *J. Bone Joint Surg. Br.* **94-B**, 1–9 (2012).

90. Bunker, T. D. & Anthony, P. P. The pathology of frozen shoulder. A Dupuytren-like disease. *J. Bone Joint Surg. Br.* **77**, 677–83 (1995).
91. Hagiwara, Y. *et al.* Coexistence of fibrotic and chondrogenic process in the capsule of idiopathic frozen shoulders. *Osteoarthr. Cartil.* **20**, 241–9 (2012).
92. Xue, M., Gong, S., Dai, J., Chen, G. & Hu, J. The Treatment of Fibrosis of Joint Synovium and Frozen Shoulder by Smad4 Gene Silencing in Rats. *PLoS One* **11**, e0158093 (2016).
93. Kabbabe, B., Ramkumar, S. & Richardson, M. Cytogenetic analysis of the pathology of frozen shoulder. *Int. J. Shoulder Surg.* **4**, 75 (2010).
94. Adstrum, S., Hedley, G., Schleip, R., Stecco, C. & Yucesoy, C. A. Defining the fascial system. *J. Bodyw. Mov. Ther.* **21**, 173–177 (2017).
95. Katz, a *et al.* The influence of pathologic tumor characteristics on locoregional recurrence rates following mastectomy. *Int. J. Radiat. Oncol. Biol. Phys.* **50**, 735–42 (2001).
96. Sandelin, K., Wickman, M. & Billgren, A. M. Oncological outcome after immediate breast reconstruction for invasive breast cancer: a long-term study. *Breast* **13**, 210–218 (2004).
97. Vallejo da Silva, A., Rodriguez, F. R., Loures, C. M. & Gloria Silami Lopes, V. Mastectomy in the era of implant-based reconstruction: Should we be removing the pectoralis fascia? *The Breast* **21**, 779–780 (2012).
98. Westbury, C. B. & Yarnold, J. R. Radiation Fibrosis — Current Clinical and Therapeutic Perspectives. *Clin. Oncol.* **24**, 657–672 (2012).
99. Giotopoulos, G. *et al.* The late radiotherapy normal tissue injury phenotypes of telangiectasia, fibrosis and atrophy in breast cancer patients have distinct genotype-dependent causes. *Br. J. Cancer* **96**, 1001–1007 (2007).
100. Wright, E. G. & Coates, P. J. Untargeted effects of ionizing radiation: implications for radiation pathology. *Mutat. Res.* **597**, 119–132 (2006).
101. Rödel, F., Frey, B., Multhoff, G. & Gaipl, U. Contribution of the immune system to bystander and non-targeted effects of ionizing radiation. *Cancer Lett.* **356**, 105–113 (2015).
102. Klammer, H., Mladenov, E., Li, F. & Iliakis, G. Bystander effects as manifestation of intercellular communication of DNA damage and of the cellular oxidative

status. *Cancer Lett.* **356**, 58–71 (2015).
103. Martin, M., Lefaix, J.-L. & Delanian, S. TGF-β1 and radiation fibrosis: a master switch and a specific therapeutic target? *Int. J. Radiat. Oncol.* **47**, 277–290 (2000).
104. Rybinski, B., Franco-Barraza, J. & Cukierman, E. The wound healing, chronic fibrosis, and cancer progression triad. *Physiol. Genomics* **46**, 223–244 (2014).
105. Wynn, T. A. & Ramalingam, T. R. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. *Nat. Med.* **18**, 1028–1040 (2012).
106. Piersma, B., Bank, R. A. & Boersema, M. Signaling in Fibrosis: TGF-β, WNT, and YAP/TAZ Converge. *Front. Med.* **2**, 59 (2015).
107. Gonzalez-Gonzalez, F. J., Chandel, N. S., Jain, M. & Budinger, G. R. S. Reactive oxygen species as signaling molecules in the development of lung fibrosis. *Transl. Res.* **190**, 61–68 (2017).
108. Fuxe, J. & Karlsson, M. C. I. TGF-β-induced epithelial-mesenchymal transition: A link between cancer and inflammation. *Semin. Cancer Biol.* **22**, 455–461 (2012).
109. Cirri, P. & Chiarugi, P. Cancer associated fibroblasts: the dark side of the coin. *Am. J. Cancer Res.* **1**, 482–497 (2011).
110. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, Inflammation, and Cancer. *Cell* **140**, 883–899 (2010).
111. Azzam, E. I., Jay-Gerin, J.-P. & Pain, D. Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury. *Cancer Lett.* **327**, 48–60 (2012).
112. Sun, J., Chen, Y., Li, M. & Ge, Z. Role of Antioxidant Enzymes on Ionizing Radiation Resistance. *Free Radic. Biol. Med.* **24**, 586–593 (1998).
113. Tengström, M. *et al. MnSOD* rs4880 and *XPD* rs13181 polymorphisms predict the survival of breast cancer patients treated with adjuvant tamoxifen. *Acta Oncol. (Madr).* **53**, 769–775 (2014).
114. Zhu, M.-L. *et al.* No association between TGFB1 polymorphisms and late radiotherapy toxicity: a meta-analysis. *PLoS One* **8**, e76964 (2013).
115. Reuther, S. *et al.* Association between SNPs in defined functional pathways and risk of early or late toxicity as well as individual radiosensitivity. *Strahlenther. Onkol.* **191**, 59–66 (2015).
116. Moullan, N. *et al.* Polymorphisms in the DNA repair gene XRCC1, breast cancer

risk, and response to radiotherapy. *Cancer Epidemiol. Biomarkers Prev.* **12**, 1168–1174 (2003).

117. Hu, J. *et al.* Amino acid substitution variants of APE1 and XRCC1 genes associated with ionizing radiation sensitivity. *Carcinogenesis* **22**, 917–922 (2001).
118. Chang-Claude, J. *et al.* Association between polymorphisms in the DNA repair genes, XRCC1, APE1, and XPD and acute side effects of radiotherapy in breast cancer patients. *Clin. Cancer Res.* **11**, 4802–4809 (2005).
119. Fernet, M. & Hall, J. Genetic biomarkers of therapeutic radiation sensitivity. *DNA Repair (Amst).* **3**, 1237–1243 (2004).
120. Pettersson, F. H. *et al.* Marker selection for genetic case–control association studies. *Nat. Protoc.* **4**, 743–752 (2009).
121. Lewis, C. M. & Knight, J. Introduction to genetic association studies. *Cold Spring Harb. Protoc.* **2012**, 297–306 (2012).
122. Worthington, J. J., Klementowicz, J. E. & Travis, M. A. TGFβ: a sleeping giant awoken by integrins. *Trends Biochem. Sci.* **36**, 47–54 (2011).
123. Gabbiani, G. The myofibroblast in wound healing and fibrocontractive diseases. *J. Pathol.* **200**, 500–503 (2003).
124. Shah, R., Hurley, C. K. & Posch, P. E. A molecular mechanism for the differential regulation of TGF-β1 expression due to the common SNP −509C-T (c. −1347C > T). *Hum. Genet.* **120**, 461–469 (2006).
125. Blobe, G. C., Schiemann, W. P. & Lodish, H. F. Role of Transforming Growth Factor β in Human Disease. *N. Engl. J. Med.* **342**, 1350–1358 (2000).
126. Anscher, M. S., Peters, W. P., Reisenbichler, H., Petros, W. P. & Jirtle, R. L. Transforming Growth Factor β as a Predictor of Liver and Lung Fibrosis after Autologous Bone Marrow Transplantation for Advanced Breast Cancer. *N. Engl. J. Med.* **328**, 1592–1598 (1993).
127. Li, C. *et al.* TGF-?1 levels in pre-treatment plasma identify breast cancer patients at risk of developing post-radiotherapy fibrosis. *Int. J. Cancer* **84**, 155–159 (1999).
128. Smith, L. R. & Barton, E. R. Regulation of fibrosis in muscular dystrophy. *Matrix Biol.* **68**–**69**, 602–615 (2018).
129. Gonzalez, D. *et al.* ALS skeletal muscle shows enhanced TGF-β signaling,

fibrosis and induction of fibro/adipogenic progenitor markers. *PLoS One* **12**, e0177649 (2017).

130. Pessina, P. & Muñoz-Cánoves, P. Fibrosis-Inducing Strategies in Regenerating Dystrophic and Normal Skeletal Muscle. *Methods Mol Biol* 73–82 (2016). doi:10.1007/978-1-4939-3810-0_7
131. Cardoso, A. L. *et al.* Towards frailty biomarkers: Candidates from genes and pathways regulated in aging and age-related diseases. *Ageing Res. Rev.* **47**, 214–277 (2018).
132. Ng, C. P., Hinz, B. & Swartz, M. A. Interstitial fluid flow induces myofibroblast differentiation and collagen alignment in vitro. *J. Cell Sci.* **118**, 4731–4739 (2005).
133. Martin, M. *et al.* Coactivation of AP-1 activity and TGF-beta1 geneexpression in the stress response of normal skin cells toionizing radiation. *Oncogene* **15**, 981–989 (1997).
134. Straub, J. M. *et al.* Radiation-induced fibrosis: mechanisms and implications for therapy. *J. Cancer Res. Clin. Oncol.* **141**, 1985–1994 (2015).
135. Boothe, D. L. *et al.* Transforming Growth Factor β-1 (TGF-β1) Is a Serum Biomarker of Radiation Induced Fibrosis in Patients Treated With Intracavitary Accelerated Partial Breast Irradiation: Preliminary Results of a Prospective Study. *Int. J. Radiat. Oncol.* **87**, 1030–1036 (2013).
136. Gault, N., Vozenin-Brotons, M. C., Calenda, A., Lefaix, J. L. & Martin, M. T. Promoter Sequences Involved in Transforming Growth Factorβ1 Gene Induction in HaCaT Keratinocytes after GammaIrradiation. *Radiat. Res.* **157**, 249–255 (2002).
137. Hageman, J. *et al.* Radiation and transforming growth factor-beta cooperate in transcriptional activation of the profibrotic plasminogen activator inhibitor-1 gene. *Clin. Cancer Res.* **11**, 5956–5964 (2005).
138. Muls, A. C. Gastrointestinal consequences of cancer treatment and the widercontext: A bad gut feeling. *Acta Oncol.* **53**, 297–306 (2014).
139. Goodier, H. C. J. *et al.* Comparison of transforming growth factor beta expression in healthy and diseased human tendon. *Arthritis Res. Ther.* **18**, 48 (2016).
140. Carl, C. *et al.* Ionizing radiation induces a motile phenotype in human carcinoma

cells in vitro through hyperactivation of the TGF-beta signaling pathway. *Cell. Mol. Life Sci.* **73**, 427–443 (2016).

141. Anscher, M. S. *et al.* Changes in plasma TGF beta levels during pulmonary radiotherapy as a predictor of the risk of developing radiation pneumonitis. *Int. J. Radiat. Oncol. Biol. Phys.* **30**, 671–6 (1994).
142. Awad, M. R. *et al.* Genotypic variation in the transforming growth factor-beta1 gene: association with transforming growth factor-beta1 production, fibrotic lung disease, and graft fibrosis after lung transplantation. *Transplantation* **66**, 1014–20 (1998).
143. Akhurst, R. J. & Hata, A. Targeting the TGFβ signalling pathway in disease. *Nat. Rev. Drug Discov.* **11**, 790–811 (2012).
144. Walton, K. L., Johnson, K. E. & Harrison, C. A. Targeting TGF-β Mediated SMAD Signaling for the Prevention of Fibrosis. *Front. Pharmacol.* **8**, 461 (2017).
145. Singer, A. J. & Clark, R. A. F. Cutaneous Wound Healing. *N. Engl. J. Med.* **341**, 738–746 (1999).
146. Grainger, D. Genetic control of the circulating concentration of transforming growth factor type beta1. *Hum. Mol. Genet.* **8**, 93–97 (1999).
147. Barnett, G. C. *et al.* Independent validation of genes and polymorphisms reported to be associated with radiation toxicity: a prospective analysis study. *Lancet Oncol.* **13**, 65–77 (2012).
148. Shimada, M. *et al.* IL-6 secretion by human pancreatic periacinar myofibroblastsin response to inflammatory mediators. *J. Immunol.* **168**, 861–868 (2002).
149. Tiggelman, A. M. *et al.* Interleukin-6 production by human liver (myo)fibroblasts inculture. Evidence for a regulatory role of LPS, IL-1 betaand TNF alpha. *J. Hepatol.* **23**, 295–306 (1995).
150. Arpin, D. *et al.* Early variations of circulating interleukin-6 and interleukin-10 levels during thoracic radiotherapy are predictive for radiation pneumonitis. *J. Clin. Oncol.* **23**, 8748–56 (2005).
151. Patrono, C., Sterpone, S., Testa, A. & Cozzi, R. Polymorphisms in base excision repair genes: Breast cancer risk and individual radiosensitivity. *World J. Clin. Oncol.* **5**, 874–882 (2014).
152. Seibold, P. *et al.* XRCC1 Polymorphism Associated With Late Toxicity After

Radiation Therapy in Breast Cancer Patients. *Int. J. Radiat. Oncol. Biol. Phys.* **92**, 1084–1092 (2015).

153. Alsbeih, G. *et al.* SNPs in genes implicated in radiation response are associated with radiotoxicity and evoke roles as predictive and prognostic biomarkers. *Radiat. Oncol.* **8**, 125 (2013).
154. Alsner, J., Andreassen, C. N. & Overgaard, J. Genetic markers for prediction of normal tissue toxicity after radiotherapy. *Semin. Radiat. Oncol.* **18**, 126–135 (2008).
155. Azria, D. *et al.* Single nucleotide polymorphisms, apoptosis, and the development of severe late adverse effects after radiotherapy. *Clin. Cancer Res.* **14**, 6284–6288 (2008).
156. Cheuk, I. W. Y., Yip, S. P., Kwong, D. L., Wan & Wu, V. W. C. Association of XRCC1 and XRCC3 gene haplotypes with the development of radiation-induced fibrosis in patients with nasopharyngeal carcinoma. *Mol Clin Oncol* **2**, 553–558 (2014).
157. Raabe, A. *et al.* Association of single nucleotide polymorphisms in the genes ATM, GSTP1, SOD2, TGFB1, XPD and XRCC1 with risk of severe erythema after breast conserving radiotherapy. *Radiat. Oncol.* **7**, 65 (2012).
158. Yin, M. *et al.* Functional Polymorphisms of Base Excision Repair Genes XRCC1 and APEX1 Predict Risk of Radiation Pneumonitis in Patients With Non–Small Cell Lung Cancer Treated With Definitive Radiation Therapy. *Int. J. Radiat. Oncol.* **81**, e67–e73 (2011).
159. Ho-Pun-Cheung, A. *et al.* A large-scale candidate gene approach identifies SNPs inSOD2 and IL13 as predictive markers of response topreoperative chemoradiation in rectal cancer. *Pharmacogenomics J.* **11**, 437–443 (2011).
160. Brown, A. L. *et al. SOD2* genetic variant associated with treatment-related ototoxicity in cisplatin-treated pediatric medulloblastoma. *Cancer Med.* **4**, 1679–1686 (2015).
161. Helmig, S., Aliahmadi, N., Stephan, P., Döhrel, J. & Schneider, J. TNF-α -308 genotypes are associated withTNF-α and TGF-β_1 mRNA expression in bloodleucocytes of humans. *Cytokine* **53**, 306–310 (2011).
162. Phillips, A. O., Topley, N., Steadman, R., Morrisey, K. & Williams, J. D. Induction of TGF-β1 synthesis in D-glucose primed humanproximal tubular cells by IL-1β

and TNFα. *Kidney Int.* **50**, 1546–1554 (1996).

163. Samad, F. *et al.* Tumor necrosis factor α is a key component in theobesity-linked elevation of plasminogen activator inhibitor 1. *Proc. Natl. Acad. Sci.* **96**, 6902–6907 (1999).
164. Sullivan, D. E., Ferris, M., Pociask, D. & Brody, A. R. The Latent Form of TGFβ1 is Induced by TNFα Through an ERK Specific Pathway and is Activated by Asbestos-Derived Reactive Oxygen Species In Vitro and In Vivo. *J. Immunotoxicol.* **5**, 145–149 (2008).
165. Barnett, G. C. *et al.* Individual patient data meta-analysis shows no association between the SNP rs1800469 in TGFB and late radiotherapy toxicity. *Radiother. Oncol.* **105**, 289–295 (2012).
166. Green, H. *et al.* Variation in the manganese superoxide dismutase gene (SOD2) is not a major cause of radiotherapy complications in breast cancer patients. *Radiother. Oncol.* **63**, 213–216 (2002).
167. Overstreet, J. M., Samarakoon, R., Cardona-Grau, D., Goldschmeding, R. & Higgins, P. J. Tumor suppressor ataxia telangiectasia mutated functions downstream of TGF-β1 in orchestrating profibrotic responses. *FASEB J.* **29**, 1258–1268 (2015).
168. Gutiérrez-Enr\'\iquez, S. *et al.* Functional consequences of ATM sequence variants for chromosomal radiosensitivity. *Genes Chromosom. Cancer* **40**, 109–119 (2004).
169. Barcellos-Hoff, M. H. & Cucinotta, F. A. New tricks for an old fox: Impact of TGF on the DNA damage response and genomic stability. *Sci. Signal.* **7**, re5–re5 (2014).
170. Overstreet, J. M., Samarakoon, R., Meldrum, K. K. & Higgins, P. J. Redox control of p53 in the transcriptional regulation ofTGF-β1 target genes through SMAD cooperativity. *Cell. Signal.* **26**, 1427–1436 (2014).
171. Lazzari, G., Natalicchio, M. I., Terlizzi, A., Perri, F. & Silvano, G. Single nucleotide polymorphisms and unacceptable late toxicity in breast cancer adjuvant radiotherapy: a case report. *Breast Cancer* **9**, 401–406 (2017).
172. Noble, P. W., Barkauskas, C. E. & Jiang, D. Pulmonary fibrosis: patterns and perpetrators. *J. Clin. Invest.* **122**, 2756–2762 (2012).
173. Biernacka, A., Dobaczewski, M. & Frangogiannis, N. G. TGF-β signaling in

fibrosis. *Growth Factors* **29**, 196–202 (2011).

174. Barnett, G. C. *et al.* No association between SNPs regulating TGF-β1secretion and late radiotherapy toxicity to the breast:results from the RAPPER study. *Radiother. Oncol.* **97**, 9–14 (2010).
175. Epperly, M. W. *et al.* Intratracheal injection of adenovirus containing the human MnSOD transgene protects athymic nude mice from irradiation-induced organizing alveolitis. *Int. J. Radiat. Oncol. Biol. Phys.* **43**, 169–181 (1999).
176. Zhang, Y. *et al.* Single Nucleotide Polymorphism rs1801516 in Ataxia Telangiectasia-Mutated Gene Predicts Late Fibrosis in Cancer Patients After Radiotherapy: A PRISMA-Compliant Systematic Review and Meta-Analysis. *Medicine (Baltimore).* **95**, e3267 (2016).
177. Zschenker, O. *et al.* Association of single nucleotide polymorphisms in ATM, GSTP1, SOD2, TGFB1, XPD and XRCC1 with clinical and cellular radiosensitivity. *Radiother. Oncol.* **97**, 26–32 (2010).
178. Edvardsen, H. *et al.* Linkage disequilibrium pattern of the ATM gene in breast cancer patients and controls; association of SNPs and haplotypes to radio-sensitivity and post-lumpectomy local recurrence. *Radiat. Oncol.* **2**, 25 (2007).
179. Ho-Pun-Cheung, A. *et al.* A large-scale candidate gene approach identifies SNPs in SOD2 and IL13 as predictive markers of response to preoperative chemoradiation in rectal cancer. *Pharmacogenomics J.* **11**, 437–443 (2011).
180. Terrazzino, S. *et al.* Common variants of GSTP1, GSTA1, and TGFβ1 are associated with the risk of radiation-induced fibrosis in breast cancer patients. *Int. J. Radiat. Oncol. Biol. Phys.* **83**, 504–511 (2012).
181. Wang, H. *et al.* Transforming growth factor-β1 gene polymorphisms are associated with progression of liver fibrosis in Caucasians with chronic hepatitis C infection. *World J. Gastroenterol.* **11**, 1929 (2005).
182. Xu, F., Liu, C., Zhou, D. & Zhang, L. TGF-β/SMAD Pathway and Its Regulation in Hepatic Fibrosis. *J. Histochem. Cytochem.* **64**, 157–67 (2016).
183. Andreassen, C. N. *et al.* Risk of radiation-induced subcutaneous fibrosis in relation to single nucleotide polymorphisms in TGFB1, SOD2, XRCC1, XRCC3, APEX and ATM; a study based on DNA from formalin fixed paraffin embedded tissue samples. *Int. J. Radiat. Biol.* **82**, 577–586 (2006).
184. Burri, R. J. *et al.* Association of single nucleotide polymorphisms in SOD2,

XRCC1 and XRCC3 with susceptibility for the development of adverse effects resulting from radiotherapy for prostate cancer. *Radiat. Res.* **170**, 49–59 (2008).

185. Barlow, C., Eckhaus, M. A., Schäffer, A. A. & Wynshaw-Boris, A. Atm haploinsufficiency results in increased sensitivity to sublethal doses of ionizing radiation in mice. *Nat. Genet.* **21**, 359–60 (1999).
186. Al-Serri, A. *et al.* The SOD2 C47T polymorphism influences NAFLD fibrosis severity: Evidence from case-control and intra-familial allele association studies. *J. Hepatol.* **56**, 448–454 (2012).
187. Little, J. *et al.* Strengthening the reporting of genetic association studies (STREGA): an extension of the strengthening the reporting of observational studies in epidemiology (STROBE) statement. *J. Clin. Epidemiol.* **62**, 597-608.e4 (2009).
188. Patterson, N. *et al.* Genetic structure of a unique admixed population: implications for medical research. *Hum. Mol. Genet.* **19**, 411–419 (2010).
189. Stats SA. Census 2011: Census in brief. *2011 South African Census* 26 (2011). Available at: http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf. (Accessed: 14th May 2020)
190. World Medical Association. WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects – WMA – The World Medical Association. *WMA DECLARATION OF HELSINKI – ETHICAL PRINCIPLES FOR MEDICAL RESEARCH INVOLVING HUMAN SUBJECTS* (2018). Available at: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/. (Accessed: 4th December 2018)
191. Harris, P. A. *et al.* Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* **42**, 377–381 (2009).
192. Mokone, G. G. *et al.* The Guanine-Thymine Dinucleotide Repeat Polymorphism within the Tenascin-C Gene is Associated with Achilles Tendon Injuries. *Am. J. Sports Med.* **33**, 1016–1021 (2005).
193. Lahiri, D. K., Nurnberger, J. I. & Jr. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res.* **19**,

5444 (1991).

194. Yuan, X. *et al.* Single Nucleotide Polymorphism at rs1982073:T869C of the *TGF β 1* Gene Is Associated With the Risk of Radiation Pneumonitis in Patients With Non–Small-Cell Lung Cancer Treated With Definitive Radiotherapy. *J. Clin. Oncol.* **27**, 3370–3378 (2009).
195. Lu, Y. *et al.* TGFB1 genetic polymorphisms and coronary heart disease risk:a meta-analysis. *BMC Med. Genet.* **13**, 39 (2012).
196. Rosenstein, B. S. Identification of SNPs associated with susceptibility for development of adverse reactions to radiotherapy. *Pharmacogenomics* **12**, 267–75 (2011).
197. Alsbeih, G., Al-Harbi, N., Al-Hadyan, K., El-Sebaie, M. & Al-Rajhi, N. Association between normal tissue complications after radiotherapy and polymorphic variations in TGFB1 and XRCC1 genes. *Radiat. Res.* **173**, 505–511 (2010).
198. Margalit, D. N. *et al.* GermLine Variation in Superoxide Dismutase-2 (SOD2) and Survival Outcomes After Radiation Therapy for Prostate Cancer: Results of a Test and Validation Set Analysis. *Clin. Genitourin. Cancer* **13**, 370-377.e1 (2015).
199. Abecasis, D. G. Genetic Association Study Power Calculator. Available at: http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/. (Accessed: 7th May 2020)
200. The Australia and New Zealand Melanoma Trials Group. Power Calculator - ANOVA - Statistical Decision Tree. (2020). Available at: https://www.masc.org.au/stats/PowerCalculator/PowerANOVA. (Accessed: 15th May 2020)
201. Oshiro, C. *et al.* Comparative evaluation of nuclear and histological grades as prognostic factors for invasive breast cancer. *Breast Cancer* (2020). doi:10.1007/s12282-020-01093-0
202. Wang, L. *et al.* Predictors of persistent pain after breast cancer surgery: A systematic review and meta-analysis of observational studies. *CMAJ* **188**, E352–E361 (2016).
203. Doong, S. H. *et al.* Associations Between Cytokine Genes and a Symptom Cluster of Pain, Fatigue, Sleep Disturbance, and Depression in Patients Prior to Breast Cancer Surgery. *Biol. Res. Nurs.* **17**, 237–247 (2015).

204. Langford, D. J. *et al.* Persistent breast pain following breast cancer surgery is associated with persistent sensory changes, pain interference, and functional impairments. *J. Pain* **15**, 1227–1237 (2014).
205. Kwak, H.-B. *et al.* MnSOD overexpression reduces fibrosis and pro-apoptotic signaling in the aging mouse heart. *J. Gerontol. A. Biol. Sci. Med. Sci.* **70**, 533–44 (2015).
206. Sutton, A. *et al.* The manganese superoxide dismutase Ala16Val dimorphism modulates both mitochondrial import and mRNA stability. *Pharmacogenet. Genomics* **15**, 311–9 (2005).
207. Shimoda-Matsubayashi, S. *et al.* Structural dimorphism in the mitochondrial targeting sequence in the human manganese superoxide dismutase gene. A predictive evidence for conformational change to influence mitochondrial transport and a study of allelic association in Parkinson's disease. *Biochem. Biophys. Res. Commun.* **226**, 561–5 (1996).
208. Bastaki, M. *et al.* Genotype-activity relationship for Mn-superoxide dismutase, glutathione peroxidase 1 and catalase in humans. *Pharmacogenet. Genomics* **16**, 279–286 (2006).
209. Kerns, S. L. *et al.* Radiogenomics Consortium Genome-Wide Association Study Meta-analysis of Late Toxicity after Prostate Cancer Radiotherapy. *J. Natl. Cancer Inst.* (2019). doi:10.1093/jnci/djz075
210. Schack, L. M. H. *et al.* Validation of genetic predictors of late radiation-induced morbidity in prostate cancer patients. *Acta Oncol.* **56**, 1514–1521 (2017).
211. Renwick, A. *et al.* ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat. Genet.* **38**, 873–875 (2006).
212. Hall, E. J. *et al.* A preliminary report: frequency of A-T heterozygotes among prostate cancer patients with severe late responses to radiation therapy. *Cancer J. Sci. Am.* **4**, 385–9
213. Cavaciuti, E. *et al.* Cancer risk according to type and location of *ATM* mutation in ataxia-telangiectasia families. *Genes, Chromosom. Cancer* **42**, 1–9 (2005).
214. Swift, M. *et al.* The incidence and gene frequency of ataxia-telangiectasia in the United States. *Am. J. Hum. Genet.* **39**, 573–83 (1986).
215. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).

216. Gurdasani, D. *et al.* The African Genome Variation Project shapes medical genetics in Africa. *Nature* **517**, 327–332 (2015).
217. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
218. Owolabi, M. *et al.* Advancing stroke genomic research in the age of Trans-Omics big data science: Emerging priorities and opportunities. *J. Neurol. Sci.* **382**, 18–28 (2017).
219. Lin, C.-H., Lin, E. & Lane, H.-Y. Genetic Biomarkers on Age-Related Cognitive Decline. *Front. Psychiatry* **8**, 247 (2017).
220. Padmanabhan, S., Aman, A. & Dominiczak, A. F. Recent Findings in the Genetics of Blood Pressure: How to Apply in Practice or Is a Moonshot Required? *Curr. Hypertens. Rep.* **20**, 54 (2018).
221. Mulder, N. *et al.* H3Africa: Current perspectives. *Pharmacogenomics and Personalized Medicine* **11**, 59–66 (2018).
222. Early and locally advanced breast cancer: diagnosis and management | Guidance and guidelines | NICE. *NICE Guidelines [NG101]* (2018). Available at: https://www.nice.org.uk/guidance/ng101/chapter/Recommendations#surgery-to-the-breast. (Accessed: 12th October 2018)
223. Boerma, M., Bart, C. I. & Wondergem, J. Effects of ionizing radiation on gene expression in culturedrat heart cells. *Int. J. Radiat. Biol.* **78**, 219–225 (2002).
224. Anscher, M. S. *et al.* Small Molecular Inhibitor of Transforming Growth Factor-β Protects Against Development of Radiation-Induced Lung Injury. *Int. J. Radiat. Oncol.* **71**, 829–837 (2008).

Appendices

Appendix A

Study information

DEPARTMENT OF HUMAN BIOLOGY

Medical School Campus
Anatomy Building
Anzio Road
Observatory
Tel: +27 (0) 21 404 7720 Fax: +27 (0) 21 448 7226
E-mail:
Internet: www.uct.ac.za

PARTICIPANT INFORMATION LEAFLET AND CONSENT FORM

Date: 27/5/12
Version: 01

TITLE OF THE RESEARCH PROJECT: Correlating clinical disease state of the shoulder after treatment for breast cancer with biomarkers of inflammation and angiogenesis and their associated genetic variants

PRINCIPAL INVESTIGATOR:Dr Delva Shamley 021 4065066

CO-INVESTIGATORS: A/Professor Malcolm Collins 021 6504574

Dr Alison September 021 6504574
A/Professor Sharon Prince 021 4066240
Dr Steve Roche 021 5065580
Dr Basil Vrettos 021 5065570

ADDRESS: Department of Human Biology, Level 2, Anatomy Building, Faculty of health Sciences, University of Cape Town, Observatory, 7925; Exercise Science and Sports Medicine (ESSM); Vincent Palotti Hospital.

Invitation to participate: We are inviting you to take part in a research project in breast cancer because you have been treated for breast cancer in the past. . Please read the information in this letter very carefully as it explains the details of the project and how you could be involved. Please ask your doctor to explain anything that you do not understand. Only you can decide if you want to take part and, if you choose to say no or to withdraw from the study at any time, it will not affect your medical treatment in any way, not now or in the future.

What is the study about? Shoulder problems are a well-documented side effect of the treatment for breast cancer. Shoulder problems experienced by some women can be as mild as slightly limited movement to more serious conditions such as frozen shoulder. Treatments like axillary surgery or radiotherapy have been shown to increase the risk of developing shoulder problems but some women still develop problems if they have not had axillary surgery. Not all women treated for breast cancer develop shoulder complications and we do not know why some women do have problems and others do not. We do know that radiotherapy can change blood flow and cause scar tissue formation. We also know that chemotherapy sets off an inflammation reaction in some tissues.
This study will be the first study to look at the proteins involved in these effects and explore how they might be related to shoulder pain and dysfunction. We will compare this to survivors with no shoulder problems.
All the information for our normal bodily functions is stored in our genetic material, DNA, which is found in all our cells. The genetic information in the DNA which is packaged into units called genes is passed from one generation to the next.
Research on genes has shown that if a small change is present in the gene(s) then it could weaken or predispose the body to illness. We are also doing research to find out which genes might predispose people to develop shoulder complications.
Research has shown that exercises and physiotherapy can help to reverse the pain and dysfunction experienced by some patients. We would like to be able to identify those patients at risk of developing shoulder complications in order to provide the correct rehabilitation advice and treatment.

Ethical Approval: This study has been approved by the University of Cape Town Human Research Ethics Committee. This study will be carried out according to the Declaration of Helsinki which is a set of guidelines informing researchers working with people how to do their research so that the people are always treated in the best possible way. It promotes respect for the subjects and protects their health and their rights.

Who will be doing the research? All of the researchers and health professionals who will be involved are experts in their field and are based at the University of Cape Town or at the Exercise Science and Sports Medicine Institute. Younger scientists who become involved will be adequately trained and supervised by a senior staff member.

Who is paying for this study? This study is funded by a University of Cape Town Development Research Grant.

Can we trust that there is no conflict of interest? In other words, can we be sure that this study is for the benefit of breast cancer patients and not to promote a hidden agenda of a sponsor or a researcher? There are NO conflicts of interest. The findings of this study are for the benefit of breast cancer survivors and will add to our knowledge of the cause of shoulder complications and how to manage them.

What if Something Goes Wrong? The University of Cape Town (UCT) promises that in the event of you suffering in health or well-being caused by you taking part in the study, it will provide immediate medical care.

Procedures and time involved:
Once we receive your reply slip we will phone you to confirm your eligibility for inclusion in the study. We will then make an appointment for you to attend the Sports Science Institute in Cape Town. This appointment will require about 1 hour of your time. This will start with a 15 minute information session about the aim of the research project. We will read through the information sheet and answer your questions about the project. If you are interested in taking part in our research project you will be asked to read and sign a consent form. If you agree to take part we will continue as follows:

1.A member of the research team will assist you to complete a questionnaire related to your personal details, your history of breast cancer. Your history of treatment by physiotherapists, your history of exercising.
2.You will fill in a short questionnaire about your current level of pain and functional ability of the arm and shoulder. You may not be experiencing any difficulty but we will still use your data to compare to those who are having difficulty.
3.A health professional will draw two small samples of blood (about 1-2 tablespoons, about 20ml) from a vein in the arm.
4.The blood will be stored at the ESSM institute, University of Cape Town. Aliquots will be transported to the Centre for Proteomic and Genomic Research for protein analysis. DNA analysis will be carried out at the institute of ESSM.

How long will my blood/DNA be stored? There is no limit as to how long the sample can be stored because we will constantly be screening the sample for different proteins and genes that may be involved in the side effects of treatment for breast cancer. However, you have the right to decide if you do not want your sample to be kept for a period longer than 10 years.

Will I get results from my tests? This is a pilot study which may identify a number of possible proteins and genes. It may take a few years before we can draw any conclusions from these results and as such we will not be sending out any results to participants.

Who can take part: To take part in this study you must be older than 18 years of age. You must have been treated for breast cancer at least 1 year ago. You must not have a history of neck or shoulder problems prior to cancer and you must not have had a mastectomy.

Who will know that I am taking part in this project: Your identity will be kept secret and your personal details will be safely stored in the office of the principal investigator, at the Department of Human Biology, University of Cape Town. Your samples will be coded with a reference number so that your name and other details will not be known to laboratory staff and will not be used in any reports or publications written about this study. The only person who will know that you are taking part is your Oncology consultant.

Will I get paid for taking part: You will not be paid to take part in the study.

Are there any risks involved in the process: Drawing blood is a standard procedure. There may be a little discomfort and pain and a small risk of bruising. If you are still concerned after a few days about a study-related injury then please contact Dr Morrell (telephone numbers are given on page 1 of this document).

What benefits are there for me: There are no direct benefits from the study. This study does not aim to provide a cure for shoulder complications but it may provide information that will be useful in the future to understand what causes the complications and how to improve treatment.

Do I have to take part? NO. It is entirely up to you if you want to take part in this study. You may decide that you do not want to take part or decide to withdraw from the project at any time. This decision will not affect any current treatment you may be receiving.

What if I have other questions now or later about the study or my rights: If you have questions about the study that are not answered in these Information Sheets please ask the interviewer to answer them. If you have any concerns related to this project in the future you may contact:

- the study investigators at the telephone numbers given on the first page

•Professor Marc Blockman, Chair, UCT Human Research Ethics Committee, Faculty of Health Sciences, may be contacted by research subjects to discuss their rights.
Office tel: 021-4066492Email address:marc.blockman@uct.ac.za

Informed consent forms

CONSENT FORM
Date: 27/5/12
Version: 01

Project Title: Correlating clinical disease state of the shoulder after treatment for breast cancer with biomarkers of inflammation and angiogenesis and their associated genetic variants

AGREEMENT TO PARTICIPATE
I, _____, have read (or had read to me) the Information Sheet for the study named 'Correlating clinical disease state of the shoulder after treatment for breast cancer with biomarkers of inflammation and angiogenesis and their associated genetic variants'
My role in the study is as a research volunteer to help the investigators collect information about the proteins and genes that may be involved in the development of arm and shoulder complications after treatment for breast cancer. I understand that the purpose of this study is not to cure shoulder problems and the genetic information may or may not be useful in designing better ways to treat these complications in the future. My questions have been answered to my satisfaction in a language that I understood. By signing this consent form I do not waive any of my rights.

I, ……., AGREE to take part in the study "Correlating clinical disease state of the shoulder after treatment for breast cancer with biomarkers of inflammation and angiogenesis and their associated genetic variants"

(initial boxes where appropriate)
I agree to give a blood sample for protein studies

I agree to give a blood sample for genetic studies related to this study only

I agree to my DNA being stored for a maximum period of 10years

I authorise my doctor to provide relevant clinical details to the
Department of Human Biology, UCT.

Research Volunteer:

Signature: _____

Date: _____

Name: _____
Please Print

Volunteer: Please sign both copies of this page

Interviewer Obtaining Consent:

Signature: _____

Date: _____

Name: _____
Please Print
Interviewer: Please sign both copies of this page and hand one copy back to the research volunteer.

Participant questionnaire

Shoulder Pain and Disability Index (SPADI)

Patient #:	
Assessment #:	
Date:	

Pain Scale

How severe is your pain?

Circle the number that best describes your pain where: **0** = no pain and **10** = the worst pain imaginable

1.At its worst?	0	1	2	3	4	5	6	7	8	9	10
2.When lying on the involved side?	0	1	2	3	4	5	6	7	8	9	10
3.Reaching for something on a high shelf?	0	1	2	3	4	5	6	7	8	9	10
4.Touching the back of your neck?	0	1	2	3	4	5	6	7	8	9	10
5.Pushing with the involved arm?	0	1	2	3	4	5	6	7	8	9	10

Clinical staff to circle:
Rating decision: 1 or 3 >5 = H **OR** 3/5≥5 =H **OR** 3/5 are 3-5 = I **OR** L

Disability Scale

How much difficulty do you have?

Circle the number that best describes your experience where: **0** = no difficulty and **10** = so difficult it requires help

1.Washing your hair?	0	1	2	3	4	5	6	7	8	9	10
2.Washing your back?	0	1	2	3	4	5	6	7	8	9	10
3.Putting on an undershirt or jumper?	0	1	2	3	4	5	6	7	8	9	10
4.Putting on a shirt that buttons down the front?	0	1	2	3	4	5	6	7	8	9	10
5.Putting on your pants?	0	1	2	3	4	5	6	7	8	9	10
6.Placing an object on a high shelf?	0	1	2	3	4	5	6	7	8	9	10
7.Carrying a heavy object of 10 pounds (4.5 kilograms)	0	1	2	3	4	5	6	7	8	9	10
8.Removing something from your back pocket?	0	1	2	3	4	5	6	7	8	9	10

Clinical staff to circle:
Rating decision: 6 >5 = H **OR** 5/8≥5 =H **OR** 5/8 are 3-5 = I **OR** L

Data Collection Sheet B: Clinical and Demographic Information

Date: 04-06-2013 Version: 01

Demographics

Participant #:		Weight (kg):	
GSH file #:		Height (m):	
RT file #:		BMI (kg/m²):	
Physiotherapy file #:		Birth date:	

Clinical

	Cancer							
Type:								
Stage:	I	II		III		IV		
Tumour type:	Invasive ductal	Tubulolobular	Mucinous	Comedo	Lobular	DCIS	Tubular	NOS
Tumour size (mm):								
	Tumour							
Tumour:	T1	T2		T3		T4		
Lymph nodes:	N1		N2		N3			
Metastasis:	M0			M1				
If M1, to where:								
# Resected nodes:								
Type of nodes:								
Receptor status:	**ER:**	Positive	Negative	**PR:**	Positive	Negative		
	Surgery							
Surgery type:								
Tissues removed:								
Procedure date:								

	Radiotherapy	**Chemotherapy**	**Hormonal therapy**
Dose intensity (Gy):			
# Fractions:			
Dose frequency:			
Total doses:			
Field of radiation:			
Start date:			
End date:			

Physiotherapy	
Start date symptoms:	
Time since surgery:	
Time 1st treatment since surgery:	
Total treatments:	
Time to resolution:	

Appendix B

Ethics approval

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room E52-24 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 404 7682 • **Facsimile** [021] 406 6411
Email: nosi.tsama@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

28 June 2017

HREC REF: 446/2017

Dr D Shamley
Clinical Research Centre
L51
Old Main Building

Dear Dr Shamley

PROJECT TITLE: A PILOT STUDY TO IDENTIFY LINKS BETWEEN GENETIC VARAITION AND SHOULDER PAIN AND DYSFUNCTION AFTER BREAST CANCER RADIOTHERAPY- LINKED TO 312/2012 (Masters candidate- Callum Mclarty)

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee for review.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study.

Approval is granted for one year until the 30th June 2018.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

We acknowledge that the student Mr C Mclarty will be involved in this study.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate institutional approval before the research may occur.

Please quote the HREC REF in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

HREC 446/2017

This serves to confirm that the University of Cape Town Human Research Ethics Committee complies to the Ethics Standards for Clinical Research with a new drug in patients, based on the Medical Research Council (MRC-SA), Food and Drug Administration (FDA-USA), International Convention on Harmonisation Good Clinical Practice (ICH GCP), South African Good Clinical Practice Guidelines (DoH 2006), based on the Association of the British Pharmaceutical Industry Guidelines (ABPI),and Declaration of Helsinki (2013) guidelines.
The Human Research Ethics Committee granting this approval is in compliance with the ICH Harmonised Tripartite Guidelines E6: Note for Guidance on Good Clinical Practice (CPMP/ICH/135/95) and FDA Code Federal Regulation Part 50, 56 and 312.

Appendix C

DNA extraction protocols

5444 Nucleic Acids Research, Vol. 19, No. 19

A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies

Debomoy K.Lahiri and John I.Nurnberger,Jr
Laboratory of Molecular Neurogenetics, Institute of Psychiatric Research, Indiana University Medical Center, 791 Union Drive, Indianapolis, IN 46202–4887

Submitted June 11, 1991

In genetic linkage studies using the restriction fragment length polymorphism (RFLP) technique, it is essential to process effectively large numbers of blood samples. One of the problems faced when extracting DNA by standard methods is the requirement of deproteinizing cell digests with hazardous organic solvents like phenol, chloroform and isoamyl alcohol (1–3). The method described in this report avoids the use of any organic solvents. This is achieved by salting out the cellular proteins by dehydration and precipitation with a saturated sodium chloride solution (4). Most of the procedures also involve prolonged incubation with proteinase K (1, 4, 5). Our procedure eliminates completely the use of proteinase K treatment. The method reported here is the most economical, safe and rapid for preparation of DNA from whole blood.

PROCEDURE

1. Collect whole blood in a Vacutainer tube (purple-stoppered) containing 100 μl of 15% EDTA.
2. Transfer 5 ml of blood into a 15 ml centrifuge tube and add 5 ml of low salt buffer containing 10 mM Tris-HCl pH 7.6, 10 mM KCl, 10 mM $MgCl_2$ and 2 mM EDTA (TKM1).
3. Add 125 μl of Nonidet P-40 (NP-40, Sigma) to lyse the cells. Mix well by inversion several times.
4. Centrifuge at 2200 RPM for 10 min at room temperature (RT) in a Beckman table-top centrifuge (model TJ-6).
5. Slowly pour off the supernatant and save the nuclear pellet (the small pellet at the very bottom of the tube) and wash the pellet in 5 ml of TKM1 buffer and centrifuge as before.
6. Gently resuspend the pellet in 0.8 ml of high salt buffer containing 10 mM Tris-HCl pH 7.6, 10 mM KCl, 10 mM $MgCl_2$, 0.4 M NaCl and 2 mM EDTA (TKM2).
7. Add 50 μl of 10% SDS and then mix the whole suspension thoroughly by pipetting back and forth several times and incubate for 10 min at 55°C.
8. Add 0.30 ml of 6 M NaCl in the tube and mix well.
9. Centrifuge at 12000 RPM for 5 min, in microcentrifuge.
10. Save the supernatant containing DNA and discard the precipitated protein pellet at the bottom of the tube.
11. To the supernatant add 2 volumes of 100% ethanol at RT and invert the tube several times until the DNA precipitates.
12. Remove the precipitated DNA strands and put them in a microcentrifuge tube containing 1 ml of ice-cold 70% ethanol.
13. Microfuge for 5 min at 12000 RPM at 4°C.
14. Dry the pellet in a Speed-vac and resuspend DNA in 0.5 ml of 10 mM Tris-HCl, 1 mM EDTA, pH 8.0 at 65°C for 15 min.
15. Measure the concentration of DNA by taking A_{260} and A_{280} and checking the quality of DNA by agarose gel electrophoresis.

RESULTS AND DISCUSSION

1. In our hands, this procedure yields more DNA than any other methods (1–5), generally in the range of 130–160 μg from 5 ml of blood. We have also found more yield and less time to extract DNA from whole blood by this 'Rapid method' than using an A.S.A.P. DNA isolation kit (5). Our method has worked successfully with 20 samples of blood from different persons.
2. The DNA preparation is free of RNA, protein and degrading enzymes. The uncut DNA is seen as a typical slow-migrating, high molecular weight and undegraded species in an ethidium bromide-stained agarose gel. The DNA produced is of good quality and is suitable for restriction enzyme digestion. We have used *Bam*HI, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Pvu*II, *Sac*I and *Taq*I restriction enzymes to digest the DNA successfully. We have found that the 'Rapid method' can yield high molecular weight DNA that can be analyzed in standard Southern blot experiments.
3. Sometimes the plasma is used for other purposes and only cells are available for DNA extraction. We have obtained by this 'Rapid method' an adequate amount of DNA from packed cells.
4. This procedure works equally well with fresh blood samples, with those that are stored at 4°C, and with those stored at −70°C.

SIGNIFICANCE

1. This procedure eliminates completely the use of any toxic reagents such as phenol, chloroform and isoamyl alcohol.
2. The 'Rapid method' is complete in less than an hour.
3. It eliminates the step of prolonged digestion of samples with proteinase K, thus saving the cost and time of operation.
4. This method yields 150 μg of DNA from 5 ml of blood.

ACKNOWLEDGEMENTS

Thanks to Dr M.E.Hodes, S.Bye and F.Sargent. Supported by the Indiana Department of Mental Health Grant to the Institute.

REFERENCES

1. Gross-Bellard,M. *et al.* (1973) *Eur. J. Biochem.* **36**, 32–38.
2. John,S.W.M. *et al.* (1991) *Nucl. Acids Res.* **19**, 408.
3. Madisen,L. *et al.* (1987) *Am. J. Med. Genet.* **27**, 379–390.
4. Miller,S.A. *et al.* (1988) *Nucl. Acids Res.* **16**, 1215.
5. *Technical Manual* from Boehringer Mannheim Company.

Appendix D

Supplementary statistical tables and R code

Gene	Genotype	Mean	SD	Median	1st quartile	3rd quartile
rs472	AG	14.983	16.702	10.750	0.000	33.800
	GG	16.066	20.740	6.200	0.000	26.200
rs058	CC	16.185	21.050	6.200	0.350	25.200
	CT	14.295	15.398	6.200	0.000	30.800
rs469	AA	9.896	14.894	2.700	0.385	19.600
	AG	15.421	20.321	5.775	0.000	26.200
	GG	17.097	20.439	6.900	0.700	30.800
rs1516	AG	13.604	17.695	6.925	0.750	18.850
	GG	16.238	20.907	6.200	0.000	26.200
rs4880	AA	15.870	20.77	5.400	0.000	26.700
	AG	15.066	19.88	7.700	0.770	19.200
	GG	20.374	22.71	13.070	0.770	34.000
rs25487	CC	16.04	21.49	6.200	0.000	26.700
	CT	16.050	20.78	6.900	0.385	24.600
	TT	17.639	21.35	7.700	2.285	26.200

ssrs472 Summary

```
sub$rs472:
  spadi.score       rs472
 Min.   : 0.00   Length:67
 1st Qu.: 1.50   Class :character
 Median :12.30   Mode  :character
 Mean   :20.06
 3rd Qu.:26.20
 Max.   :77.70
 NA's   :18
------------------------------------
sub$rs472: AG
  spadi.score        rs472
 Min.   : 0.000   Length:6
 1st Qu.: 0.375   Class :character
 Median :10.750   Mode  :character
 Mean   :14.983
 3rd Qu.:30.350
 Max.   :34.600
------------------------------------
sub$rs472: GG
  spadi.score       rs472
 Min.   : 0.00   Length:270
 1st Qu.: 0.00   Class :character
 Median : 6.20   Mode  :character
 Mean   :16.07
 3rd Qu.:26.20
 Max.   :98.50
 NA's   :5

 Descriptive statistics by group
group:
            vars  n  mean    sd median trimmed   mad min  max range skew kurtosis   se
spadi.score    1 49 20.06 23.32   12.3   16.91 16.75   0 77.7  77.7 1.16     0.01 3.33
rs472*         2 67   NaN    NA     NA     NaN    NA Inf -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------
group: AG
            vars n  mean   sd median trimmed   mad min  max range skew kurtosis   se
spadi.score    1 6 14.98 16.7  10.75   14.98 15.94   0 34.6  34.6 0.18    -2.13 6.82
rs472*         2 6   NaN   NA     NA     NaN    NA Inf -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------
group: GG
            vars   n  mean    sd median trimmed  mad min  max range skew kurtosis   se
spadi.score    1 265 16.07 20.74    6.2   12.09 9.19   0 98.5  98.5  1.6     2.12 1.27
rs472*         2 270   NaN    NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
```

rs058 Summary

```
sub$rs058:
  spadi.score       rs058
 Min.   : 0.00   Length:83
 1st Qu.: 1.50   Class :character
 Median :12.30   Mode  :character
 Mean   :18.71
 3rd Qu.:26.20
 Max.   :77.70
 NA's   :18
------------------------------------------------------------
sub$rs058: CC
  spadi.score       rs058
 Min.   : 0.000   Length:253
 1st Qu.: 0.525   Class :character
 Median : 6.200   Mode  :character
 Mean   :16.186
 3rd Qu.:25.100
 Max.   :98.500
 NA's   :5
------------------------------------------------------------
sub$rs058: CT
  spadi.score       rs058
 Min.   : 0.00   Length:7
 1st Qu.: 1.15   Class :character
 Median : 6.20   Mode  :character
 Mean   :14.30
 3rd Qu.:29.25
 Max.   :33.07
```

```
 Descriptive statistics by group
group:
           vars  n  mean    sd median trimmed   mad min  max range skew kurtosis   se
spadi.score    1 65 18.71 21.83   12.3   15.08 18.24   0 77.7  77.7 1.23     0.39 2.71
rs058*         2 83   NaN    NA     NA     NaN    NA Inf -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------
group: CC
           vars   n  mean    sd median trimmed  mad min  max range skew kurtosis   se
spadi.score    1 248 16.19 21.05    6.2   12.16 9.19   0 98.5  98.5  1.6     2.07 1.34
rs058*         2 253   NaN    NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------
group: CT
           vars n mean   sd median trimmed  mad min   max range skew kurtosis   se
spadi.score    1 7 14.3 15.4    6.2    14.3 9.19   0 33.07 33.07 0.21    -2.14 5.82
rs058*         2 7  NaN   NA     NA     NaN   NA Inf  -Inf  -Inf   NA       NA   NA
```

rs469 Summary

```
sub$rs469:
  spadi.score       rs469
 Min.   : 0.00   Length:135
 1st Qu.: 1.25   Class :character
 Median :10.80   Mode  :character
 Mean   :18.15
 3rd Qu.:24.23
 Max.   :98.50
 NA's   :20
------------------------------------------------------
sub$rs469: AA
  spadi.score         rs469
 Min.   : 0.0000   Length:9
 1st Qu.: 0.5775   Class :character
 Median : 2.7000   Mode  :character
 Mean   : 9.8963
 3rd Qu.:12.5000
 Max.   :33.8000
 NA's   :1
------------------------------------------------------
sub$rs469: AG
  spadi.score        rs469
 Min.   : 0.000   Length:123
 1st Qu.: 0.000   Class :character
 Median : 5.775   Mode  :character
 Mean   :15.422
 3rd Qu.:25.900
 Max.   :93.070
 NA's   :1
------------------------------------------------------
sub$rs469: GG
  spadi.score        rs469
 Min.   : 0.000   Length:76
 1st Qu.: 0.735   Class :character
 Median : 6.900   Mode  :character
 Mean   :17.097
 3rd Qu.:30.400
 Max.   :88.500
 NA's   :1
```

```
 Descriptive statistics by group
group:
            vars   n  mean    sd median trimmed   mad min  max range skew kurtosis   se
spadi.score    1 115 18.15 22.65   10.8   13.86 16.01   0 98.5  98.5  1.5     1.39 2.11
rs469*         2 135   NaN    NA     NA     NaN    NA Inf -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------------
group: AA
            vars n mean    sd median trimmed mad min  max range skew kurtosis   se
spadi.score    1 8  9.9 14.89    2.7     9.9   4   0 33.8  33.8 0.89    -1.25 5.27
rs469*         2 9  NaN    NA     NA     NaN  NA Inf -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------------
group: AG
            vars   n  mean    sd median trimmed  mad min   max range skew kurtosis   se
spadi.score    1 122 15.42 20.32   5.78    11.6 8.56   0 93.07 93.07 1.58     2.04 1.84
rs469*         2 123   NaN    NA     NA     NaN   NA Inf  -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------------
group: GG
            vars  n mean    sd median trimmed   mad min  max range skew kurtosis   se
spadi.score    1 75 17.1 20.44    6.9   13.76 10.23   0 88.5  88.5 1.29     1.05 2.36
rs469*         2 76  NaN    NA     NA     NaN    NA Inf -Inf  -Inf   NA       NA   NA
```

rs1516 Summary

```
sub$rs1516:
  spadi.score       rs1516
 Min.   : 0.00   Length:66
 1st Qu.: 1.50   Class :character
 Median :12.70   Mode  :character
 Mean   :20.91
 3rd Qu.:35.73
 Max.   :77.70
 NA's   :18
------------------------------------------
sub$rs1516: AA
  spadi.score    rs1516
 Min.   :0    Length:1
 1st Qu.:0    Class :character
 Median :0    Mode  :character
 Mean   :0
 3rd Qu.:0
 Max.   :0
------------------------------------------
sub$rs1516: AG
  spadi.score        rs1516
 Min.   : 0.000   Length:28
 1st Qu.: 1.125   Class :character
 Median : 6.925   Mode  :character
 Mean   :13.604
 3rd Qu.:18.675
 Max.   :67.700
------------------------------------------
sub$rs1516: GG
  spadi.score       rs1516
 Min.   : 0.00   Length:248
 1st Qu.: 0.35   Class :character
 Median : 6.20   Mode  :character
 Mean   :16.24
 3rd Qu.:26.20
 Max.   :98.50
 NA's   :5
```

```
 Descriptive statistics by group
group:
           vars  n  mean    sd median trimmed   mad min  max range skew kurtosis  se
spadi.score    1 48 20.91 23.54   12.7   17.85 18.83   0 77.7  77.7 1.06     -0.2 3.4
rs1516*        2 66   NaN    NA     NA     NaN    NA Inf -Inf  -Inf   NA       NA  NA
------------------------------------------------------------------------------------------
group: AA
           vars n mean sd median trimmed mad min  max range skew kurtosis se
spadi.score    1 1    0 NA      0       0   0   0    0     0   NA       NA NA
rs1516*        2 1  NaN NA     NA     NaN  NA Inf -Inf  -Inf   NA       NA NA
------------------------------------------------------------------------------------------
group: AG
           vars  n mean   sd median trimmed   mad min  max range skew kurtosis   se
spadi.score    1 28 13.6 17.7   6.93   10.97 10.27   0 67.7  67.7 1.51     1.44 3.34
rs1516*        2 28  NaN   NA     NA     NaN    NA Inf -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------
group: GG
           vars   n  mean    sd median trimmed  mad min  max range skew kurtosis   se
spadi.score    1 243 16.24 20.91    6.2   12.22 9.19   0 98.5  98.5  1.6     2.14 1.34
rs1516*        2 248   NaN    NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
```

rs4880 Summary

```
sub$rs4880:
  spadi.score       rs4880
 Min.   : 0.00   Length:92
 1st Qu.: 1.50   Class :character
 Median :10.00   Mode  :character
 Mean   :18.17
 3rd Qu.:25.96
 Max.   :83.80
 NA's   :18
------------------------------------------------
sub$rs4880: AA
  spadi.score       rs4880
 Min.   : 0.00   Length:94
 1st Qu.: 0.00   Class :character
 Median : 5.40   Mode  :character
 Mean   :15.87
 3rd Qu.:26.57
 Max.   :98.50
 NA's   :4
------------------------------------------------
sub$rs4880: AG
  spadi.score       rs4880
 Min.   : 0.00   Length:117
 1st Qu.: 0.77   Class :character
 Median : 7.70   Mode  :character
 Mean   :15.07
 3rd Qu.:19.20
 Max.   :93.07
------------------------------------------------
sub$rs4880: GG
  spadi.score        rs4880
 Min.   : 0.000   Length:40
 1st Qu.: 0.785   Class :character
 Median :13.070   Mode  :character
 Mean   :20.375
 3rd Qu.:33.900
 Max.   :88.500
 NA's   :1
```

```
 Descriptive statistics by group
group:
            vars  n  mean    sd median trimmed   mad min  max range skew kurtosis   se
spadi.score    1 74 18.17 22.46     10   14.06 13.64   0 83.8  83.8 1.36     0.76 2.61
rs4880*        2 92   NaN    NA     NA     NaN    NA Inf -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------
group: AA
            vars  n  mean    sd median trimmed  mad min  max range skew kurtosis   se
spadi.score    1 90 15.87 20.77    5.4   11.96 8.01   0 98.5  98.5 1.52     1.94 2.19
rs4880*        2 94   NaN    NA     NA     NaN   NA Inf -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------
group: AG
            vars   n  mean    sd median trimmed   mad min   max range skew kurtosis   se
spadi.score    1 117 15.07 19.88    7.7   11.04 11.42   0 93.07 93.07 1.84     3.05 1.84
rs4880*        2 117   NaN    NA     NA     NaN    NA Inf  -Inf  -Inf   NA       NA   NA
------------------------------------------------------------------------------------------
group: GG
            vars  n  mean    sd median trimmed   mad min  max range skew kurtosis   se
spadi.score    1 39 20.37 22.71  13.07   17.67 19.38   0 88.5  88.5 0.92     0.11 3.64
rs4880*        2 40   NaN    NA     NA     NaN    NA Inf -Inf  -Inf   NA       NA   NA
```

rs25487 Summary

```
sub$rs25487:
  spadi.score      rs25487
 Min.   : 0.00   Length:90
 1st Qu.: 1.50   Class :character
 Median :12.30   Mode  :character
 Mean   :18.35
 3rd Qu.:26.16
 Max.   :77.70
 NA's   :18
------------------------------------------------------------
sub$rs25487: CC
  spadi.score      rs25487
 Min.   : 0.00   Length:145
 1st Qu.: 0.00   Class :character
 Median : 6.20   Mode  :character
 Mean   :16.04
 3rd Qu.:26.70
 Max.   :98.50
 NA's   :3
------------------------------------------------------------
sub$rs25487: CT
  spadi.score       rs25487
 Min.   : 0.000   Length:88
 1st Qu.: 0.385   Class :character
 Median : 6.900   Mode  :character
 Mean   :16.050
 3rd Qu.:23.450
 Max.   :83.800
 NA's   :1
------------------------------------------------------------
sub$rs25487: TT
  spadi.score       rs25487
 Min.   : 0.000   Length:20
 1st Qu.: 2.285   Class :character
 Median : 7.700   Mode  :character
 Mean   :17.639
 3rd Qu.:26.200
 Max.   :66.900
 NA's   :1
```