EXAMINATION OF GENE DELETION AS A MECHANISM OF RBM5 DOWNREGULATION IN TOBACCO SMOKE-ASSOCIATED LUNG CANCERS

by

Sarah C. Hunt

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc) in Chemical Sciences

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
Laurentian Université/Université Laurentienne
Faculty of Graduate Studies/Faculté des études supérieures

Title of Thesis Titre de la thèse	EXAMINATION OF GENE DELETION AS A MECHANISM OF RBM5 DOWNREGULATION IN TOBACCO SMOKE-ASSOCIATED LUNG CANCERS
Name of Candidate Nom du candidat	Hunt, Sarah
Degree Diplôme	Master of Science
Department/Program Département/Programme	Chemical Sciences
Date of Defence Date de la soutenance	October 13, 2017

APPROVED/APPROUVÉ

Thesis Examiners/Examinateurs de thèse:

Dr. Leslie Sutherland
(Supervisor/Directrice de thèse)

Dr. Carita Lanner
(Committee member/Membre du comité)

Dr. Thomas Merritt
(Committee member/Membre du comité)

Dr. William Lockwood
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

ACCESSIBILITY CLAUSE AND PERMISSION TO USE

Abstract

RBM5 is a tumour suppressor gene with substantially decreased expression in most lung cancers, especially in smokers, that may result from gene deletion. The objective of this study was to determine if significantly decreased levels of RBM5 expression in the lung cancers of smokers was related to *RBM5* deletion. Using DNA from patient lung tissue, *RBM5* gene copy number was quantified by quantitative polymerase chain reaction. Although deletions were detected in the lung specimens, *RBM5* gene copy number was not significantly decreased in lung tumours. *RBM5* deletions were detected in the non-tumour DNA of smokers, but not in the never-smokers, indicating that *RBM5* gene deletion might be related to smoking. In conclusion, it appeared as though RBM5 was downregulated by more than one mechanism, which included gene deletion. Further analyses must be carried out to examine other mechanisms by which RBM5 is downregulated.

Keywords

RBM5, non-small cell lung cancer, small cell lung cancer, gene copy number, gene deletion, expression, tobacco smoke exposure, smokers, never-smokers

Acknowledgments

First and foremost, I would like to thank my thesis supervisor, Dr. Leslie Sutherland. When I started working in your lab as an undergraduate student, you presented me with the amazing opportunity to perform cancer research on clinical samples, helping me to accomplish one of my lifetime goals. As a supervisor, I would like to thank you for your insight and guidance, for believing in my abilities as a researcher, and for allowing me my independence in the lab. You are truly dedicated to ensuring your students have a meaningful learning experience, and you constantly present us with opportunities to grow and better ourselves. I would also like to thank you for being not just a supervisor, but also a friend. You have always been there to listen, and to talk whenever I've needed it. You are truly a positive role model, and I sincerely admire your dedication towards your research.

I would also like to thank my committee members, Dr. Carita Lanner and Dr. Thomas Merritt. The time you have spent participating on my thesis committee is greatly appreciated. Thank you for providing me with strong input and direction throughout my studies.

I would like to thank current and former members of the Sutherland Lab including Julie Loiselle, Jose Knee, Justin Roy and Benjamin Koenderink. Each of you was a joy to work with, and you helped make even the most stressful days in the lab more enjoyable. I also would like to thank you for your experimental expertise, advice and support. In addition, I would like to thank the other researchers and students at the Health Sciences North Research Institute for your camaraderie, and making the lab a very pleasant place to pursue research.

A big thank you goes to my parents, Ray and Carolyn, for your constant love, support and encouragement throughout my studies. In particular, I would like to thank Justin Roy for your

unwavering support throughout my graduate studies. When I experienced success, you shared in the celebration. When I was challenged or frustrated, you were there to listen, motivate me and brighten my day. Thank you for being patient with me and believing in me.

Most importantly, I would like to dedicate this thesis to my aunt, Betty Ozols, who passed away from lung cancer during my Master's studies. The strength and optimism Betty demonstrated throughout her entire battle with lung cancer was truly inspiring. I will be forever grateful for all of the love, laughter, encouragement and prayers she brought to my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

°C – degrees Celsius

µg – micrograms

µl – microliters

µM – micromolar

ACTB – β-actin

ADC – adenocarcinoma

bp – base pairs

CSE – cigarette smoke extract

Ct – threshold cycle

DTT – dichlorodiphenyltrichloroethane

E – efficiency

ES – extensive stage

HSN – Health Sciences North

kb – kilobases

KD – knockdown

kDa – kilo Daltons

KH – K Homology

LOH – loss of heterozygosity

LS – limited stage

M – molar

mg – milligrams

mM – millimolar

NSCLC – non-small cell lung cancer

NTC – no template control

PTM – post-translational modification

PUM1 – Pumilio RNA Binding Family Member 1

qPCR – quantitative polymerase chain reaction

RBM5 – RNA Binding Motif 5

RBP – RNA binding protein

RNP – ribonucleoprotein

RPLP0 – Ribosomal Protein Lateral Stalk Subunit P0

RT-qPCR – reverse transcription quantitative polymerase chain reaction

SCLC – small cell lung cancer

SqCC – squamous cell carcinoma

TSG – tumour suppressor gene

ZnF – Zinc Finger

Chapter 1

1 Introduction

1.1 RNA Binding Proteins

RNA binding proteins (RBPs) are regulatory elements within the cell that play an important role in the post-transcriptional processing of RNA [1,2]. It is estimated that the human genome codes for over 500 RBPs, each with essential functions in RNA metabolism in both the nucleus and the cytoplasm [3,4]. In order for a protein to be considered a RBP, it must contain at least one RNA binding domain [1]. The most commonly observed RNA binding domains include RNA Recognition Motif (RRM) domains, K Homology (KH) domains and Zinc Finger (ZnF) domains [1,4]. These domains can undergo a level of post-translational modification (PTM), usually through acetylation or phosphorylation [5,6]. This mechanism of RBP regulation suggests that RBPs function as a part of complex metabolic and signalling networks [6]. Using their functional domains, RBPs can bind with their RNA target, forming a ribonucleoprotein (RNP) complex [7]. Generally, RBPs will interact with the untranslated regions of RNA transcripts, but they have also been documented to interact with coding regions and even non-protein-coding RNA [4,5]. Once bound to their target RNA, RBPs regulate many aspects of RNA metabolism including alternative splicing, polyadenylation, mRNA stabilization, editing and repair, nucleocytoplasmic transport, translation and RNA degradation [1,4,7]. Due to the vast functions of RBPs, and the sheer number of RBPs within a given cell, these proteins are critical for controlling gene expression at the post-transcriptional level.

Dysregulation of RBPs is linked to various diseases, including cancer. RBPs can exhibit a cancerous phenotype following mutations or alterations in expression [8]. Mutations in RBPs,

including binding site mutations, can lead to altered binding affinities and changes in subcellular localization [4]. RBPs harbouring non-silent mutations in their RNA binding domain either can have a reduced ability to bind to RNA targets, or, have an enhanced ability to bind to RNA targets [9]. Both reduced binding ability and increased binding stability between RBPs and their RNA target can lead to the downregulation of target transcripts. Chromosomal translocations of RBPs have also been reported in cancer, leading to fusion proteins that display oncogenic properties [5].

In regards to altered RBP expression, both increased and decreased expression of RBPs have been observed [5]. Overexpression of RBPs, through processes such as gene amplification, demonstrates enhanced RBP activity and abnormal function [5,8]. Conversely, decreased expression of RBPs hinders proper functioning [5]. Considering the multitude of functions that RBPs possess, any abnormalities in RBP function or expression can lead to defects in cell differentiation, cell division, integrity checkpoints and response to stimuli [4,5]. Currently, it is not understood whether mutations or changes in expression of RBPs are involved in cancer initiation, or are a by-product of cancer, meaning that further investigation of RBPs in cancer is necessary.

1.2 RBM5

RNA Binding Motif 5 (RBM5) is a RNA binding protein that maps to 3p21.3, and is considered a putative tumour suppressor gene (TSG) [10]. It was first cloned in 1996 as LUCA-15, but has also been cloned as H37, RBM5 and LUCA15 [11-14]. The *RBM5* gene extends approximately 30 kilobases (kb). Full-length RBM5 mRNA consists of over 2500 base pairs (bp), containing 25 exons, ranging between 61 bp and 627 bp [13,15]. RBM5 mRNA also has alternative splice variants, but for the purposes of this study, we will be focusing solely on full-length RBM5. The RBM5

protein consists of 815 amino acids and has a calculated molecular weight of 90 kilo Daltons (kDa) [13]. Using the most accurate RBM5 antibody, the RBM5 protein is observed on Western blots as ~113 kDa, suggesting that the RBM5 protein is post-translationally modified [16]. In fact, RBM5 is a known phosphoprotein, with the dephosphorylated state being associated with apoptosis [17]. Post-translational modification of the RBM5 protein may, therefore, play an important role in RBM5 function.

1.2.1 RBM5 Function

RBM5 regulates the cell cycle and modulates apoptosis through the alternative splicing of pre-mRNA in the nucleus. It was reported that the downregulation of RBM5 affects transcript levels of 35 transcripts involved in cell proliferation and apoptosis [18]. The two RRM domains on the RBM5 protein have been shown to bind to pre-mRNA, resulting in the alternative splicing of various transcripts [19]. Full-length RBM5 has a role in the alternative splicing of apoptosis-associated transcripts such as Caspase 2, Fas and c-FLIP [20-22]. RBM5 has been demonstrated to promote apoptosis through a number of pathways including Fas, TNF-α, TRAIL and p53 [23,24]. RBM5-mediated apoptosis was correlated with an increase in expression of a number of pro-apoptotic factors such as BAX, mitochondrial cytochrome c in the cytosol, and Caspases 3 and 9, as well as a decrease in the expression of anti-apoptotic Bcl-2 and Bcl-x [25-27]. With regards to cell cycle regulation, RBM5 was shown to arrest cells in G_1 phase, thereby promoting apoptosis [27,28].

A number of studies have examined the function of RBM5 *in vitro*, especially using the human lung adenocarcinoma A549 cell line. RBM5 overexpression suppressed growth in A549 cells, along with other cell lines, including HT108, RBM5-null MCF-7, H1299 cells, PC-3 cells and mouse A9 cells [14,27-31]. In two of those studies, overexpression of RBM5 in the A549 cell line

also demonstrated a higher proportion of cells arrested in the G_1 phase, leading to an increase in apoptosis [27,28]. This finding was also confirmed in the CEM-C7 cell line [27,28]. Overexpression of RBM5 in A549 also resulted in reduced expression of EGFR mRNA and protein expression, thereby preventing EGFR-mediated cell proliferation [32]. Furthermore, RBM5 overexpression in A549 cells was also associated with decreased expression of Cyclin A, a protein responsible for cell cycle progression, and with decreased expression of phosphorylated RB, which, while phosphorylated, is considered inert and allows for cell cycle progression [27]. A549 cells treated with cigarette smoke extract (CSE) demonstrated decreased RBM5 expression levels and increased Wnt and β-catenin protein expression levels, suggesting that RBM5 may also regulate the Wnt pathway [33]. To support this finding, RBM5 knockdown (KD) in A549 cells was associated with increased β-catenin, and this observation was confirmed in other cell lines including Calu-6, BEAS-2B, H1299 and MCF-10A [34]. In regards to treatment outcomes, overexpression of RBM5 in A549 cells led to reduced resistance to the chemotherapeutic agent, Cisplatin [35]. When A549 cells overexpressing RBM5 were injected intraperitoneally into immunocompromised mice, tumours exhibited retarded growth *in vivo,* further demonstrating the role of RBM5 as a putative tumour suppressor [27]. From the collective work completed in the A549 cell model, it can be concluded that RBM5 plays an important role in downregulating cell cycle progression, leading to growth arrest and apoptosis.

RBM5 studies have also been completed in the human BEAS-2B normal bronchial epithelial cell line. When RBM5 was overexpressed in CSE-transformed cells, proliferation, invasion and migration was inhibited [36]. An increase in G_1/S arrest and apoptosis was observed in the CSE-transformed cells overexpressing RBM5, further confirming that the growth inhibitory properties of RBM5 are completed through cell cycle arrest and the induction of apoptosis, as was seen in

A549 studies mentioned previously [36]. To expand on this finding, RBM5 overexpression was related to an increase in mRNA transcripts that regulate apoptosis, including pro-apoptotic Caspases 9 and 3 and BAX, and was also related to a decrease in mRNA expression of anti-apoptotic Bcl-2 [36]. Overexpression of RBM5 in these cells was also related to reduced mRNA transcript levels of CDK4, CDK6, Cyclin D1 and Cyclin A, which are all involved in cell cycle progression, and increased transcript levels of p53 and p21, further promoting apoptosis [36]. In response to cell invasion and migration, increased RBM5 overexpression was associated with a significant decrease in VEGF, HIF-α, MMP2 and MMP9 mRNA levels, all of which are involved in metastasis [36]. Supporting the role of RBM5 as a tumour suppressor *in vivo*, it was also demonstrated that tumour growth was inhibited in xenografts overexpressing RBM5 in nude mice, similar to what was seen in similar A549 studies [36]. This study in CSE-transformed BEAS-2B cells highlights the importance of RBM5 expression during the transformation of normal bronchial epithelial cells to a cancerous phenotype, especially in tobacco smoke-associated lung cancers.

One recent study examined the effect of re-introducing RBM5 expression at varying levels into an RBM5-null cell line, GLC20. Through RNA sequencing, it was determined that 12% of the transcriptome was differentially expressed when RBM5 was introduced at low levels, and 18% of the transcriptome was differentially expressed when RBM5 was introduced at high levels [16]. In both low and high RBM5-expressing cell lines, 50% of differentially expressed genes were shared between cell lines [16]. It was reported that RBM5 expression was important to the maintenance of the non-transformed state [16]. This phenomenon was completed through the regulation of pathways involved in the cell cycle and apoptosis [16]. Functional studies were completed to confirm that these effects were observed *in vitro*. Indeed, it was demonstrated that

increased RBM5 expression was associated with a decrease in cell proliferation and membrane integrity and an increase in apoptosis, especially in the presence of Cisplatin, a chemotherapeutic agent commonly used in the treatment of SCLC [16].

1.2.2 RBM5 Expression

RBM5 is ubiquitously expressed in all primary tissue. It exists at high levels in tissue such as the pancreas, heart and skeletal muscle, and exists at low levels in the liver and lung [29,37-39]. For the purpose of this study, we will be focusing solely on RBM5 expression in lung cancer. In a subset of lung cancers, referred to as non-small cell lung cancer (NSCLC), RBM5 showed reduced mRNA expression in 65-82% of NSCLC tumours, compared to adjacent non-tumour tissue [29,40]. Looking at specific NSCLC subtypes, decreased RBM5 expression occurred in both adenocarcinoma (ADC) and squamous cell carcinoma (SqCC), but occurred more frequently in SqCC [29]. A separate study examining small cell lung cancer (SCLC) RNA found 50% reduction in RBM5 expression between non-tumour and tumour RNA, but this was only analyzed in two patients [16]. At the protein level, RBM5 expression showed downregulation in approximately 70% of NSCLC tumours [27,40]. Downregulation of the RNA transcript and protein was noted in a number of lung cancer studies, illustrating the importance of loss of RBM5 expression in lung cancer. Of even greater interest, it was observed that downregulation of RBM5 in lung cancer occurs more often in smokers, compared to never-smokers [40]. The link between decreased RBM5 expression and tobacco smoke exposure observed in patients was also demonstrated *in vitro*, as CSE-transformed BEAS-2B cells had reduced expression of RBM5 compared to wild-type BEAS-2B cells [36]. In addition to the relationship between RBM5 expression and tobacco smoke exposure, decreased RBM5 expression was also correlated with metastasis [34,40,41]. One study, completed in 2003, found that decreased RBM5 mRNA expression was part of a 17-gene

signature associated with metastasis in 12 solid tumour types, including lung ADC [41]. Two separate studies also demonstrated that decreased RBM5 expression was associated with lymph node metastasis [34,40]. With all of these studies taken as a whole, it appears as though loss of RBM5 expression is important for the development of most lung cancers, especially tobacco smoke-associated lung cancers, and is also important for lung cancer progression towards a metastatic phenotype.

1.2.3 RBM5 Regulation

The mechanism responsible for the frequent downregulation of RBM5 in lung cancer has been investigated previously, but, currently, the exact mechanism of RBM5 downregulation is unknown. One of the first studies, completed in 1999, involved the sequencing of cDNA from 14 SCLC and two NSCLC cell lines. The purpose of this study was to identify any aberrations in the cDNA that would disrupt the structure and functionality of the resultant RBM5 protein, promoting lung cancer development. Northern and Southern blot analyses did not detect any abnormalities in the RNA or the cDNA of the 16 cell lines [13]. Only one mutation was noted in an adenocarcinoma cell line, GLCA2, demonstrating a G→T transversion at codon 398 [13]. This mutation resulted in an amino acid substitution from a glycine to a valine, which was not likely to alter the protein structure and function [13]. Mutations in RBM5 cDNA were investigated a second time in 2007. In this study, it was hypothesized that RBM5 downregulation in tumours, at the RNA and/or protein level, was the result of an inactivating mutation following Knudson's two-hit hypothesis, similar to classic tumour TSGs such as *RB* and *TP53*. To examine mutations that were of functional relevance, cDNA from 17 NSCLC tumour and non-tumour pairs were sequenced. Upon analysis, there were not any protein-altering mutations present in any of the

samples, indicating that, once again, mutation was unlikely to be the mechanism of RBM5 downregulation in lung cancer [42].

Promoter hypermethylation has also been investigated as a mechanism of RBM5 downregulation. This type of epigenetic alteration involves the addition of methyl groups to cytosine bases at various CpG sites in the promoter region [43]. A 949 bp CpG island in the *RBM5* promoter region, containing 69 CpG dinucleotides, was examined in DNA from 11 NSCLC tumour and non-tumour pairs, as well as three NSCLC cell lines (A549, H460 and H520) [44]. This study did not find any evidence of promoter hypermethylation; however, there were experimental limitations that could have contributed to this finding. First, only two regions of the 949 bp CpG island were amplified: a 204 bp region that included the transcription start site and 14 CpG sites, and a 164 bp region downstream of the transcription start site containing 18 CpG sites. The remaining 541 bp and 37 CpG sites were not examined, preventing a complete investigation of the *RBM5* CpG island of interest. Second, tobacco smoke exposure, which has been shown to influence methylation in NSCLC, was not taken into account. Given the limitations of this study, promoter hypermethylation cannot be ruled out as a mechanism of RBM5 downregulation.

As mentioned previously, the *RBM5* gene resides at the locus 3p21.3. This region frequently undergoes loss of heterozygosity (LOH) in 70% of NSCLCs and 95% of SCLCs [45-47]. It is one of the earliest preneoplastic alterations observed in the lung, and has even been found in normal bronchial epithelium of smokers [47]. It is, therefore, possible that gene deletion is one of the mechanisms by which RBM5 is downregulated in lung cancer, especially in lung cancers associated with tobacco smoke exposure. *RBM5* gene deletions were first noted in SCLC cell lines GLC20, NCI-H740 and NCI-H1450 [39]. Each of these cell lines demonstrated homozygous

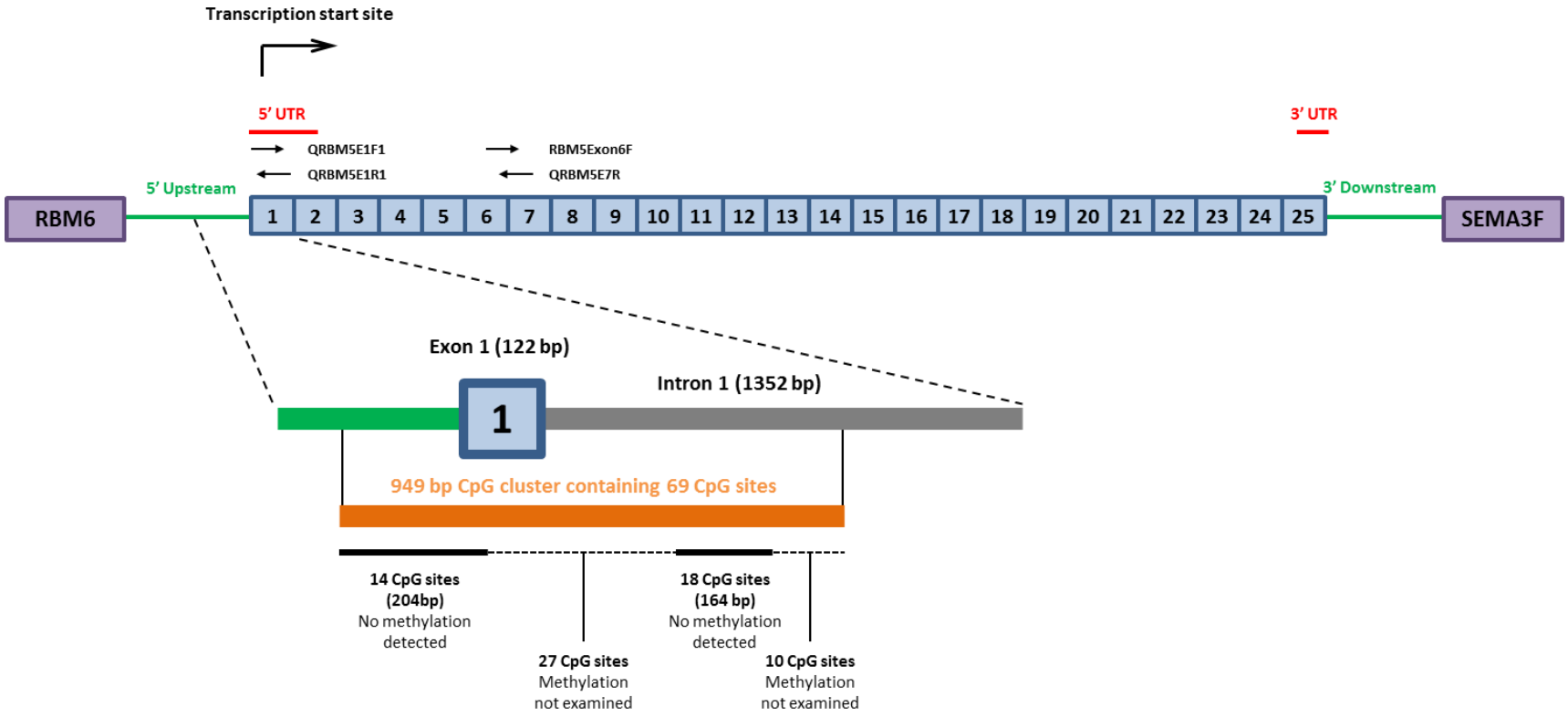


Figure 1 Schematic of *RBM5* gene. This figure depicts the *RBM5* gene located between *SEMA3F* (towards the centromere) and *RBM6* (towards the telomere). The *RBM5* gene contains 25 exons. The 5' untranslated region (UTR) spans exon 1 and part of exon 2. QRBM5E1F1 and QRBM5E1R1 primers used for *RBM5* gene copy number analyses were located in exon 1. The primers RBM5Exon6F and RBM5E7R that were used for RBM5 mRNA expression analyses spanned exons 6 and 7. Also depicted in this figure is the 949 bp region containing 69 CpG sites, located at the end of the 5' upstream region, in exon 1 and in part of the first intron. The two regions of the CpG island that were analyzed by Oh *et al.* in 2008 are also depicted.

deletions in a ~370 kb deletion region within the 3p21.3 locus [39]. In this deletion region, the *RBM5* gene was located proximal to the telomeric breakpoint within the *RBM6* gene [39]. Another study, completed in 2004, examined deletions in the ~370 kb critical deletion region in 23 lung cancer cell lines (16 SCLC, four ADC, one SqCC, one large cell carcinoma and one mixed NSCLC). This study demonstrated that the *RBM5* gene was located distal to the telomeric border of the smallest deletion region common to all of the samples examined and, therefore, not included in the deletion region [48]. This study further examined deletions in all lung cancer cell lines by examining copy number of a gene located within the deletion region, *GNAI2*. Homozygous deletions of *GNAI2* were detected in only two of the 16 SCLC cell lines, and none of the NSCLC cell lines [48]. Interestingly, hemizygosity of *GNAI2* was detected in 10 of the 23 lung cancer cell lines (one mixed NSCLC cell line, three ADC cell lines, one SqCC cell line and five SCLC cell lines) [48]. This finding suggests that loss of the ~370 kb critical deletion region, from at least one chromosome, is important to lung cancer. Taking both studies into account, the *RBM5* gene lies close to the telomeric breakpoint and is sometimes deleted in lung cancer, thereby substantiating the possibility that gene deletion is at least one mechanism by which RBM5 is downregulated in lung cancer.

1.3 Tumour Suppressor Genes and Lung Cancer

TSGs must be downregulated in order for cancer to progress. TSGs regulate various aspects of the cell cycle including cell proliferation, DNA repair and apoptosis [49]. Direct suppression of proliferation is controlled by numerous TSGs throughout the cell cycle in response to stimuli signaling for growth inhibition or in response to metabolic imbalances and DNA damage [50]. The inactivation of any of these TSGs can lead to uncontrolled cell proliferation, an important hallmark of cancer [51]. A second class of TSGs are responsible for proofreading DNA and

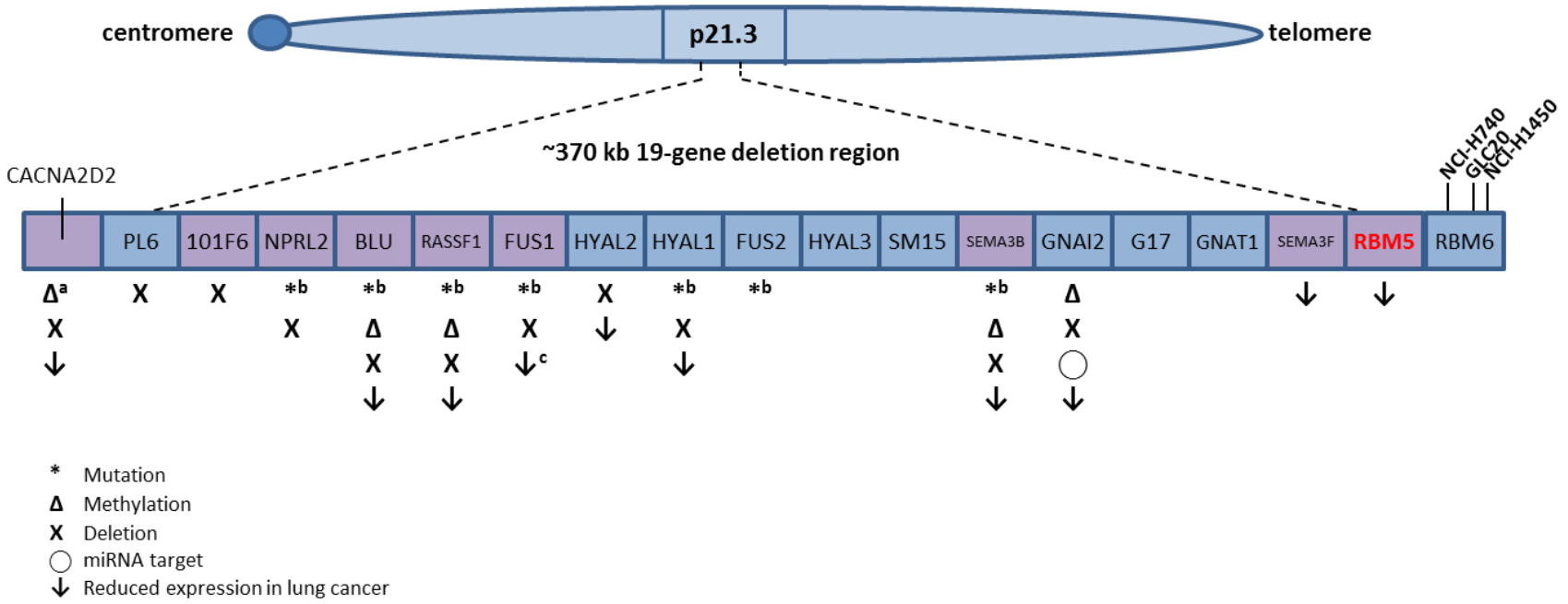


Figure 2 Schematic of the common 3p21.3 deletion region and the common molecular aberrations that occur. This figure shows the 19 genes within the common 3p21.3 deletion region. Genes that are considered TSGs are coloured in purple. Common molecular aberrations that have been detected include mutations (*****), methylation (Δ), and deletion (**X**). Genes that are downregulated in lung cancers are also indicated (↓). Expression of GNAI2 was also shown to be regulated by microRNA (◯).

[a] = frequently occurs in cell lines, but not in primary tumours

[b] = has been observed, but is a rare event

[c] = downregulated at the protein level, but not at the mRNA level

repairing damaged DNA during cell division, as errors in DNA replication can contribute to carcinogenesis [50,52]. When TSGs involved in DNA proofreading and repair are inactivated, cells with cancer-promoting mutations will progressively divide and accumulate additional detrimental mutations, leading to genomic instability and cancer [51,52]. Generally, in a normal cell, if there is too much DNA damage that is beyond the repairing capabilities of the DNA repair genes, the cell undergoes apoptosis [52]. Some TSGs are involved in triggering apoptosis [50]. If these apoptosis-initiating TSGs are impaired, apoptosis does not occur and damaged cells continue to proliferate, leading to cancer [51].

Knudson's two-hit hypothesis is currently used to describe the downregulation of TSGs in both hereditary and sporadic cancers. This hypothesis states that in order cancer to occur, both alleles of a TSG must be inactivated [51,53]. In hereditary cancers, patients exhibit a germline mutation in one allele of a TSG that will predispose them to cancer (first hit), but tumour formation will only occur if the remaining functional allele is inactivated by a somatic mutation during the patients' lifetime (second hit) [51,53]. Conversely, sporadic cancers occur after two somatic mutations inactivate both functional alleles of a TSG in a single cell [54]. Although most TSGs follow Knudson's two-hit hypothesis, emerging evidence suggests that some TSGs do not require the inactivation of both functional alleles to result in tumour initiation or progression. Knudson states that if only one allele is inactivated, the remaining allele is able to maintain the function of the TSG in the cell [53]. New evidence suggests that some TSGs might be haploinsufficient, meaning that inactivation of one allele of a TSG is enough to prevent proper regulation and function [54,55]. There are three proposed mechanisms by which the phenomenon of haploinsufficiency occurs. In the first mechanism, reduced expression levels of gene product are

unable to effectively inhibit the activity of downstream targets involved in cell growth [51,54]. The second mechanism of haploinsufficiency involves dominant-negative effects whereby the mutant allele produces a gene product that acts as an antagonist against the wild-type gene products [51,54]. Lastly, the third mechanism of haploinsufficiency describes transcriptional silencing of wild-type alleles through alterations in the expression or regulation of transcriptional machinery [51]. As downregulation of TSGs are crucial in tumour initiation and progression, it is important to understand the mechanisms by which TSG expression decreases.

Downregulation of TSGs generally occurs as a result of mutation, gene deletion, epigenetic silencing, or a combination of these events [54,56]. Missense mutations involve the substitution of a nucleic acid that alters the amino acid composition of a protein. This type of event can alter the structure of the protein, which, in turn, can prevent proper functioning of the protein, rendering it inactive [56]. Secondly, nondisjunction events can occur through errors in cell division. The missegregation of chromosomes can result in entire chromosomal deletions or amplifications. Abnormal mitotic recombination during cell division can also result in deletions, insertions and translocations that inactivate TSGs. Interestingly, it has been reported that cancerous cells will lose an average of 25-30% of alleles that were present in normal cells, with some cancer cells losing up to 75% of alleles that were present in normal cells, demonstrating the high prevalence of LOH events in cancer [57]. Epigenetic silencing of TSGs occurs through CpG methylation [58,59]. This mechanism involves the addition of methyl groups to cytosine bases at various CpG sites in the promoter region [43]. The hypermethylation of TSG promoters prevents transcriptional machinery from binding to the promoter region, thereby silencing transcription [60]. In addition, recent studies have implicated microRNA activity in the downregulation of TSGs in human

cancers, but this mechanism is still being investigated [61]. We can speculate that RBM5 downregulation is the result of at least one, or a combination, of these events.

1.4 Lung Cancer

Lung cancer is the leading cause of cancer-related deaths in Canada, accounting for over 25% of all cancer deaths [62]. In Canada, 27% of lung cancer patients are diagnosed at Stage 3, while 48% of lung cancer diagnoses are made when patients' cancers are at Stage 4 and exhibit distant metastasis [63]. Although patients with early-stage lung cancer typically have a better response to treatment, 35-50% of those patients relapse within 5 years [64]. Late stage diagnosis and disease relapse are responsible for a 5 year survival rate of 17% among lung cancer patients [63]. Due to the high mortality rates associated with this disease, it is prudent that we broaden our knowledge and understanding of the molecular events that occur in lung cancer.

1.4.1 Lung Cancer and Smoking

Tobacco-smoke exposure is the leading cause of lung cancer, as 85% of lung cancer patients are current or former smokers [65]. Through the act of smoking, smokers are exposed to over 4000 chemical constituents, with approximately 100 of those constituents being known mutagens and carcinogens [65,66]. There are two ways in which smokers are exposed to these various compounds. The primary method of exposure is through mainstream smoke, when the smoker inhales air through the cigarette [65]. Mainstream smoke consists of both vapour and particulate phases [65,66]. The vapour phase contains over 500 chemical compounds, while the particulate phase consists of over 3500 chemical compounds [65,66]. Sidestream smoke is the second route of exposure, involving the inhalation of smoke in the surrounding air between puffs [65]. This type of smoke also accounts for second-hand smoke exposure in non-smokers, however, second-hand smoke

exposure is responsible for only 1.6% of lung cancers [67]. The mutagens in tobacco smoke, whether in mainstream smoke or sidestream smoke, inflict DNA damage though the formation of DNA adducts, transversions (generally G→T) within DNA sequences, and DNA methylation (especially hypermethylation of the promoter region of tumor suppressor genes) [65,66].

Although lung cancer generally occurs in current and former smokers, only 10% of smokers develop lung cancer [68]. The reason for this is unknown, but it has been postulated that lung cancer risk in smokers is related to factors such as the type of cigarette smoked, the presence of a filter, the depth of inhalation, the length of inhalation, as well as genetic factors [65,69]. In fact, differences in smoking habits may even contribute to differences in clinical presentation, lung cancer histology, and even treatment outcomes [70]. Lung cancer risk can be reduced upon smoking cessation [69]. Despite this fact, former smokers continue to exhibit aberrant lesions and allelic losses in lung tissue, suggesting that many molecular changes that occur due to cigarette smoke exposure are not reversible [71]. Conversely, some molecular changes, such as DNA methylation, are indeed reversible and can be reduced upon smoking cessation, however, aberrant methylation patterns are still demonstrated in former smokers [71,72]. This further suggests that lung cancer caused by tobacco smoke exposure is going to be an outcome so long as cigarettes are available.

1.4.2 Subtypes of Lung Cancer

1.4.2.1 Small Cell Lung Cancer

SCLC accounts for 10-15% of all lung cancers [65,73]. It is considered the most aggressive type of lung cancer because the disease has generally metastasized at the time of diagnosis. Due to its aggressive nature, approximately 95% of patients with this diagnosis succumb to the disease [74,75]. SCLC is characterized as a neuroendocrine tumour that originates in the epithelial cells of the

bronchi [65,76]. SCLC harbours many genetic alterations, with mutation rates reported to be as high as 7.4 ± 1 protein-altering mutations per million base pairs [77]. This trend is likely due to tobacco smoke exposure, as patients with SCLC are almost exclusively current or former smokers [75,76]. Very little is known about molecular events causing SCLC. The major contributing factor to the absence of knowledge regarding genetic alterations in SCLC is due to the lack of available tissue specimens, as SCLC is generally not treated with surgery [77].

Downregulation of TSGs is important for SCLC initiation, as *TP53* and *RB1* are mutated in over 80% of cases, and considered possible driver genes [77]. While these TSGs are downregulated by inactivating mutations, *RBM5* generally does not harbour any inactivating mutations, indicating that downregulation of RBM5 does not occur through the same mechanism as other TSGs that are considered important to SCLC. Interestingly, one of the most common genetic alterations in SCLC involves the deletion of regions on the short arm of chromosome 3, in nearly 100% of cases [76]. As *RBM5* maps to this region, it is possible that gene deletion is the mechanism of decreased RBM5 expression in SCLC. This notion is substantiated by the fact that 3p21.3 undergoes LOH in 95% of SCLCs [45-47].

1.4.2.2 Non-small Cell Lung Cancer

NSCLC is the most frequently diagnosed subtype of lung cancer, accounting for 85-90% of all lung cancers [67,76]. NSCLC can be further categorized into ADC, SqCC and large cell carcinoma [76]. For the purpose of this study, we will be focusing solely on ADC and SqCC categories.

1.4.2.2.1 Adenocarcinoma

ADC is the most frequently diagnosed subtype of NSCLC in North America [78]. This type of NSCLC originates in epithelial cells located on the periphery of the lungs, and is the least related

to smoking [69,79,80]. In fact, never-smokers account for 15% of patients within this category, and many former smokers also develop ADC instead of other lung cancer subtypes [69]. This type of NSCLC also predominantly presents itself in women instead of men [70]. A study of 183 ADC tumour and non-tumor pairs revealed that the mean exonic mutation rate in ADC is 12.0 events/megabase [78]. Mutation rates in ADC are generally higher in smokers, most often in the form of C→A transversions [81]. Smokers most often present *KRAS* activating mutations in 33% of patients [81]. Never-smokers and light smokers generally will not harbour *KRAS* mutations, but instead will express *EGFR* mutations that are mutually exclusive with *KRAS* mutations [81].

Mutations in TSGs are also common in ADC. Mutations of *TP53* are the most common, and have been reported to appear in 46% of ADC tumours [81]. In addition, ADC tumours may also exhibit mutations in TSGs such as *STK11* (17%), *KEAP1* (17%), *NF1* (11%), *RB1* (4%) and *CDK2NA* (4%) [78,81]. Some ADC tumours will also exhibit translocations in *ALK*, *RET* and *ROS*, but in a very small percentage of patients [78,81]. Over 2000 gene deletions have been identified in ADC tumours [78]. The most common deletion regions have been reported on chromosomes 3p, 4q, 5q, 6q, 8p, 9 and 13q [76]. Given the fact that deletions on chromosome 3p are common in ADC, LOH at 3p21.3 is a likely mechanism of RBM5 downregulation in this subtype of NSCLC.

1.4.2.2.2 Squamous Cell Carcinoma

SqCC is the subtype of NSCLC that is most related to smoking, as over 90% of patients diagnosed with this disease are either current or former smokers [76]. Patients with SqCC are generally also heavier smokers, in comparison to current and former smokers that develop ADC [69,70]. Unlike ADC, SqCC predominately occurs in males [70]. This subtype of lung cancer generally arises in the squamous epithelial cells of the bronchi [69,76,80]. A study completed by The Cancer

Genome Atlas (TCGA) identified an average of 360 exonic mutations, 323 altered copy number segments and 165 genomic rearrangements per tumour [82]. Interestingly, the number of copy number alterations detected in SqCC was higher than other molecular profiling endeavours by the TCGA [82]. This finding has also been noted in other studies that have found a greater number of DNA deletions in SqCC compared to ADC [79]. Deletions in SqCC are common, most often occurring on chromosome 3p, but also occurring on chromosomes 4q, 5q, 8p, 9p, 10q, 11p, 13q, 17p, 18q and 21q [76]. Since deletions on the short arm of chromosome 3p are common in both SqCC and ADC, it is possible that the *RBM5* gene is deleted in both subtypes of NSCLC.

1.4.2.3 Classification of Lung Cancer Subtypes

1.4.2.3.1 SCLC Staging

SCLC is not staged according to the TNM classification like most other cancers. Instead, SCLC is classified using a two-stage system devised by the Veteran's Administration Lung Study Group [73]. This two-stage system encompasses both limited stage (LS) and extensive stage (ES) classifications [73,75,76,83]. LS includes SCLC tumours that are localized to the chest and within a single field of radiation [73,75,83]. In order to fit within this criteria, the tumor must be isolated to a single lung, the mediastinum, and/or the lymph nodes located nearby and on the same side of the body as the lung tumour [73,75,83]. Approximately 40% of SCLC patients fit within this category [74,75]. Alternately, the remaining 60% of SCLC patients diagnosed with ES have tumours that extend beyond a single field of radiation as the cancer has generally metastasized to either the other lung, brain, liver, bone, or lymph nodes located away from the original tumour [74,75,83]. For both stages, the survival outcomes are generally not favourable as over 95% of patients will eventually succumb to the disease [74,75]. Despite the fact that 80% of LS patients will show a complete response to treatment, the 5-year survival rate is 10-17%, with the median survival time

being approximately 17 months [74,75,84]. Due to the extent of metastasis in ES patients, only 20% of patients will show a response to treatment. Since the majority of ES patients will not respond to treatment, the 5-year survival rate is only 2%, with the median survival time being approximately 7 months [74,75,85]. Since most patients receive an ES diagnosis, it is necessary that we identify methods to screen at-risk populations, whether through improved imaging technologies or through novel biomarkers that can help to detect SCLC in early stages, so that patients can receive more effective treatments and increase survival outcomes.

1.4.2.3.2 NSCLC Staging and Pathology

NSCLC is classified according to the TNM Classification System (Table 1) [76]. The primary tumour (T) is staged based on metrics such as the size of the primary tumour and the location of the primary tumor, followed by examination of metastasis to regional lymph nodes (N) and distant metastasis (M) to other organs or other regions of the lung [76]. Based on the results of the TNM staging, patients are organized into prognostic groupings, ranging from Stages I-IV, with patients in earlier stages having a more favourable prognosis [64]. In addition to use of the TNM Classification System, NSCLC tumours are also graded using a four-tier grading system, allowing clinicians to better predict how the cancer will grow and spread [76,86]. The grade of the primary tumour is evaluated based on how differentiated tumour cells appear in comparison to normal lung cells. Grade 1, also known as “well-differentiated” cells are considered low grade as the cancer cells appear to be very similar to normal lung tissue [76,86]. These types of tumours are likely to be slower growing and less aggressive [76,86]. Grade 2 tumours are considered intermediate grade or “moderately differentiated” [76,86]. The most aggressive tumours are those that are grades 3 and 4, also known as “poorly differentiated” and “undifferentiated,”

respectively [76,86]. These high-grade tumours do not appear similar to normal lung tissue, and are likely to grow quickly and spread, contributing to a poor prognosis [76,86].

Table 1 TNM Classification of Lung Tumours

Primary Tumor (T)	
X	Cannot be assessed
0	No evidence of primary tumour
1	Tumour is ≤3 cm in the greatest dimension, surrounded by the lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than the main bronchus
1a	Tumour is ≤2 cm in the greatest dimension
1b	Tumour is >2 cm but ≤3 cm in the greatest dimension
2	Tumour is >3 cm in greatest dimension with involvement of the main bronchus, ≥2 cm distal to the carina, invades visceral pleura, or associated with atelectasis or obstructive pneumonitis that extends to the hilar region without involving the entire lung
2a	Tumour is >3 cm but ≤5 cm in the greatest dimension
2b	Tumour is >5 cm but ≤7 cm in the greatest dimension
3	Tumour is >7 cm and invades the chest wall, diaphragm, mediastinal pleura, or parietal pericardium **OR** Tumour <2 cm in the main bronchus distal to the carina without involvement of the carina **OR** Tumour is associated with atelectasis or obstructive pneumonitis of the entire lung
4	Tumour of any size that invades the mediastinum, heart, great vessels, trachea, esophagus, vertebral body, or carina **OR** Separate tumour nodules in a different ipsilateral lobe
Regional Lymph Nodes (N)	
X	Cannot be assessed
0	No lymph node metastasis
1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes
2	Metastasis in ipsilateral mediastinal and/or subcarnial lymph nodes
3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or subclavicular lymph nodes
Distant Metastasis (M)	
X	Cannot be assessed
0	No distant metastasis
1	Distant metastasis, including separate tumor nodules in a different ipsilateral or contralateral lobe

Note: Table was adapted from WHO Classification of Tumours of the Lung[76]

1.5 Study Hypothesis and Objectives

In the context of what has been reported in the literature, we know that RBM5 is downregulated in most lung cancers, more so in the lung cancers of smokers compared to never-smokers [40]. This finding suggests that there are two mechanisms that regulate RBM5 expression, one of which is smoking related. To date, these mechanisms have not been elucidated. Given the fact that the genes located in the common 3p21.3 deletion region are rarely mutated, and protein-altering mutations have not been detected in *RBM5*, mutation is unlikely to be a mechanism of RBM5 downregulation [13,39,42]. Since tobacco smoke exposure has been shown to influence promoter hypermethylation, and RBM5 is more frequently downregulated in lung tumours from smokers compared to lung tumours from never-smokers, it is conceivable that promoter hypermethylation may be one of the mechanisms by which RBM5 is downregulated [65,66,71,72]. In support of this, many of the TSGs in the common 3p21.3 deletion region have been shown to undergo promoter hypermethylation, leading to reduced and/or loss of expression [39,46,87]. Although the sole RBM5 promoter hypermethylation study found no evidence of promoter hypermethylation, there were a number of CpG sites that were not examined, and tobacco smoke exposure was not taken into account [10,44]. Therefore, at this point in time, promoter hypermethylation cannot be ruled out as a mechanism of RBM5 downregulation. For many TSGs within the common 3p21.3 deletion region, both promoter hypermethylation and allele loss have been observed [39,46,87]. As previously mentioned, the *RBM5* gene resides in the region of 3p21.3 that undergoes LOH in most lung cancers and even in the normal bronchial epithelium of smokers [39,47]. These observations suggest that gene deletion may be another mechanism by which RBM5 is downregulated, potentially in relation to tobacco smoke.

The working **hypothesis** for this thesis was that deletion of the *RBM5* gene was one of the mechanisms by which RBM5 was downregulated in the lung cancers of smokers. The **objective** of this study was to determine if significantly decreased levels of RBM5 expression in the NSCLCs of smokers was related to *RBM5* deletion. To accomplish this, we obtained non-tumour and tumour tissues, and tobacco smoke exposure data from patients diagnosed with NSCLC, collected at Health Sciences North (HSN) over the course of one year. DNA and RNA were extracted from these tissue specimens. Firstly, the study cohort was confirmed as representative, having downregulated RBM5 expression in tumour, compared to non-tumour tissue. Secondly, *RBM5* gene copy number studies were carried out. Statistical analyses were carried out to determine if *RBM5* gene copy number was (1) significantly decreased in tumours, and (2) associated with tobacco smoke exposure. Statistical power for the analyses relating to *RBM5* gene copy number and tobacco smoke exposure was limited due to the fact that we received tissue samples from only six never-smokers diagnosed with NSCLC. Since *RBM5* deletions were not frequently detected in the HSN NSCLC cohort, the study objectives were broadened to examine *RBM5* gene copy number in SCLC DNA samples, obtained through collaboration.

For further confirmation that RBM5 was downregulated in NSCLC, changes in RBM5 protein expression between non-tumour and tumour pairs from patients diagnosed with NSCLC were examined. The change in RBM5 protein expression in these samples was more complex than had been anticipated, and, because protein expression levels were not required to enable resolution of the objectives, protein data were provided as an appendix.

Chapter 2

2 Materials and Methods

2.1 Test Samples

2.1.1 Test Sample Procurement

The studies completed in this thesis were carried out using samples from two sources. Tissue samples were obtained from HSN, while DNA was obtained from a collaborator.

2.1.1.1 Health Sciences North Lung Tissue

Lung tissue was procured from 103 patients treated at HSN in Sudbury, Ontario, Canada between May 2014 and May 2015. All tissue was acquired with informed consent, according to the Institutional Research Ethics Board guidelines. All tissue was obtained from thoracotomies, and most was representative of invasive carcinoma. Non-tumour tissue was obtained from the periphery of each tumour at the time of surgery, and designated as non-tumour at that time. Procured tissue was immediately snap frozen in liquid nitrogen for 5 minutes, and then stored at -80°C. The procured tissue was accompanied with the corresponding pathology report, and a comprehensive smoking and mining questionnaire.

A total of 188 tissue samples were received from 103 patients in the HSN cohort. A flow-chart describing the HSN cohort is included in Figure 3. These patients received various diagnoses that included ADC, SqCC, mixed NSCLC, metastatic carcinoma, among other medical conditions. For the purpose of this study, we focused solely on the patients diagnosed with ADC and SqCC. A total of 56 patients were diagnosed with ADC, and a total of 24 patients were diagnosed with SqCC. Six of the patients in the ADC group were never-smokers, while the remaining 50 ADC

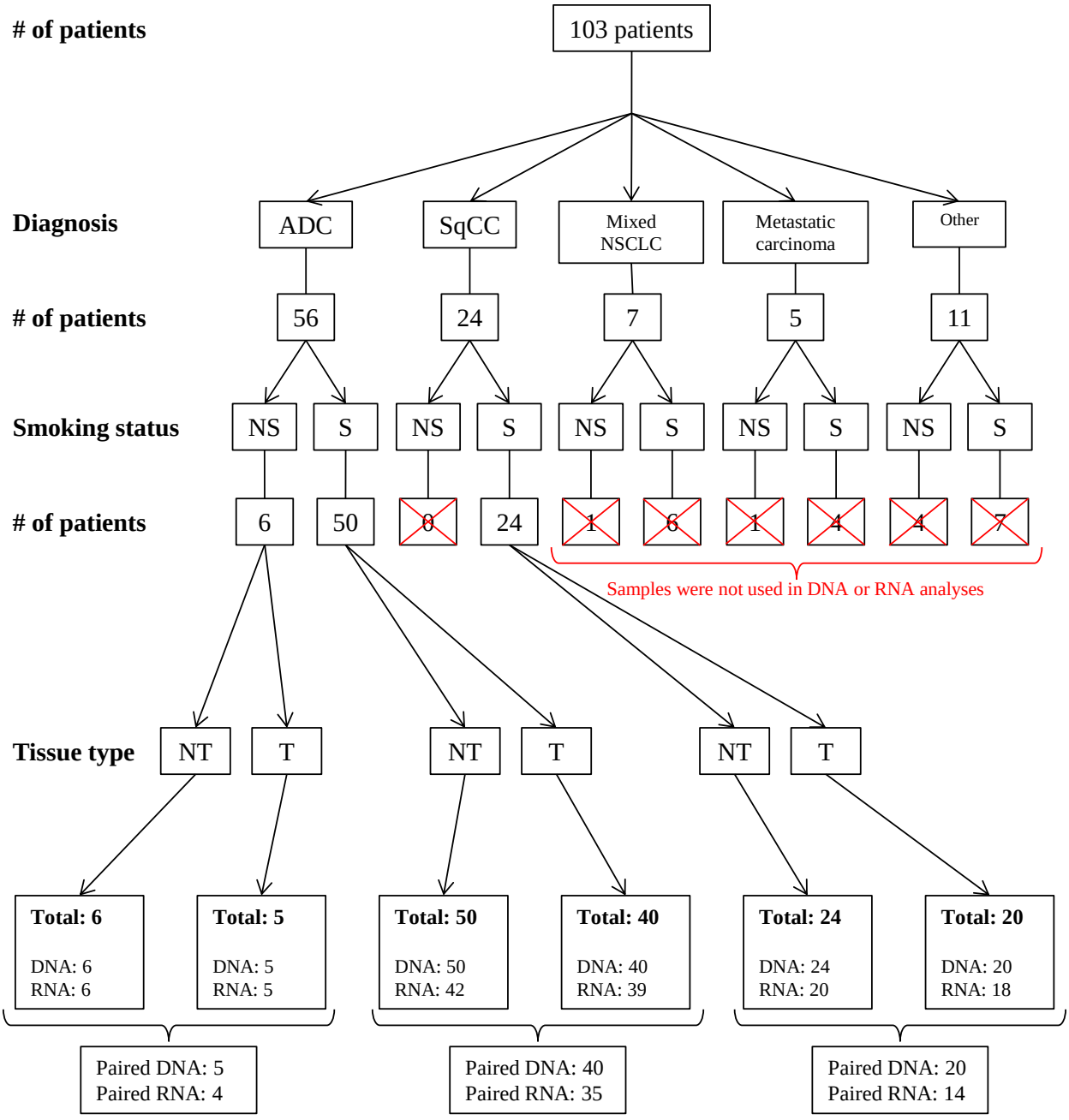


Figure 3 Flow-chart of test samples procured from HSN. Test samples were received from 103 patients diagnosed with ADC, SqCC, mixed NSCLC, metastatic carcinoma, among other medical conditions. A total of 56 patients were diagnosed with ADC, including six never-smokers and 50 smokers. There were also 24 patients that were diagnosed with SqCC, all of whom were smokers. Tissue samples from patients diagnosed with mixed NSCLC, metastatic carcinoma, and other medical conditions, were not included in DNA or RNA analyses. In total, 80 non-tumour and 65 tumour specimens were received. There were 65 patients with paired non-tumour and tumour specimens. NS = never-smoker, S = smoker, NT = non-tumour, and T = tumour.

patients, and the 24 SqCC patients, were either current or former smokers. Never-smokers were defined as patients who had smoked less than 100 cigarettes in their lifetime, while the smokers were defined as patients who were current or former smokers, exposed to ≥100 cigarettes [71]. Due to the fact that we had only received tissue samples from six never-smokers diagnosed with ADC, our statistical capabilities were limited when analyzing the role of tobacco smoke exposure on *RBM5* copy number and expression in tumours. For the patients that were considered smokers, tobacco smoke exposure was measured in pack-years, which was defined as the number of packs smoked per day multiplied by the number of years the patient smoked [70,80]. A total of 80 non-tumour tissues and 65 non-tumour tissues were received. Paired non-tumour and tumour samples were received from 65 patients. DNA was obtained from all of the tissues, while RNA was obtained for most of the tissues.

2.1.1.2 Slovakian SCLC DNA

DNA from 44 SCLC patients was obtained from Dr. Erika Halasova of the Comenius University, in Bratislava, Slovakia. A flow-chart describing the Slovakian cohort is included in Figure 4. Only tumour DNA was obtained for this cohort. In regards to tobacco smoke exposure, eight patients were never-smokers, 34 patients were smokers and two patients had an unknown smoking history.

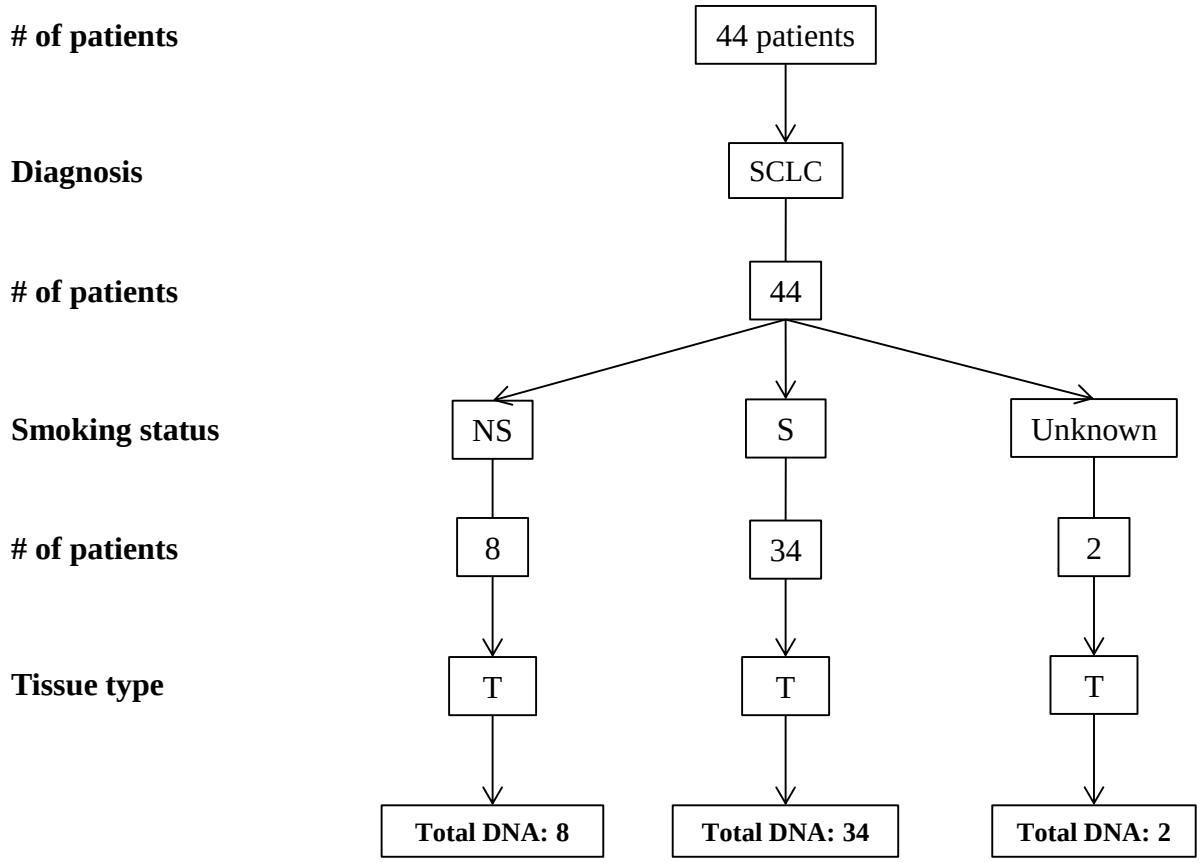


Figure 4 Flow-chart of test samples procured from Slovakia. A total of 44 SCLC DNA samples were received from Slovakia. This cohort provided SCLC DNA from eight never-smokers, 34 smokers, and two patients of unknown smoking history. NS = never-smoker, S = smoker, T = tumour.

2.1.2 Tissue Preparation

DNA, RNA and protein were isolated from the HSN lung specimens, using the Allprep DNA/RNA/Protein Mini Kit (Qiagen, Mississauga, Ontario, Canada), according to the manufacturer's instructions, and as described immediately below.

2.1.2.1 Pulverization and Homogenization

Approximately 20 milligrams (mg) of lung tissue was cut using a sterile blade and transferred to a tissue pulverizer (Bessman, Fisher Scientific, Ottawa, Ontario, Canada) that had been

immersed in liquid nitrogen for 10 minutes. The mortar was then placed in the tissue pulverizer and smashed with a hammer until the tissue was broken up into fine powder-like consistency. Tissue was transferred to a microfuge tube containing 600 microliters (μl) of Buffer RLT (Qiagen) and 0.14 millimolar (mM) β-mercaptoethanol (Sigma, Oakville, Ontario, Canada), and subsequently homogenized using a Polytron PT 1300 D Homogenizer (Kinematica, Lucerne, Switzerland). The resulting lysate was centrifuged for 3 minutes at 17 000 x g at 21°C, to pellet any tissue that did not homogenize in solution. The supernatant was subsequently transferred to an Allprep DNA spin column (Qiagen).

2.1.2.2 Genomic DNA Purification

DNA was extracted, according to the manufacturer's instructions. The Allprep DNA spin column containing the supernatant was centrifuged at 8 000 x g for 30 seconds at 21°C. The resulting flow-through was placed at 21°C for subsequent RNA purification, while the DNA on the spin column was washed with Buffers AW1 and AW2 (Qiagen). Wash buffers were eluted from the spin column by centrifugation at 8 000 x g for 30 seconds and 17 000 x g for 2 minutes at 21°C, for Buffers AW1 and AW2, respectively. To elute the DNA, Buffer EB (Qiagen), pre-heated to 70°C, was applied to the Allprep DNA spin column and incubated at room 21°C for 2 minutes. Following incubation, DNA was eluted by centrifugation at 8 000 x g for 1 minute at 21°C, and subsequently stored at -80°C. DNA was quantified, using the Nanodrop 2000c Spectrophotometer (Thermo Scientific, Waltham, Massachusetts, U.S.A.).

2.1.2.3 Total RNA Purification

RNA was extracted according to the manufacturer's instructions. The initial flow-through, which had been stored at 21°C during DNA purification, was combined with 100% ethanol, to precipitate the RNA, and transferred to the RNeasy spin column (Qiagen). The spin column was

then centrifuged at 8 000 x g for 15 seconds at 21°C. The flow-through, containing the protein, was saved for subsequent protein precipitation, while the RNA that remained within the column was subjected to subsequent wash steps. The RNA was washed with Buffer RW1 (Qiagen), by centrifugation at 8 000 x g for 15 seconds at 21°C. Flow-through was discarded and the RNeasy column was then washed twice with Buffer RPE and centrifuged at 8 000 x g, for 30 seconds and 2 minutes at 21°C, respectively. RNase-free water was then added to the column and RNA was eluted by centrifugation at 8 000 x g for 1 minute at 21°C. The eluted RNA was quantified, using the Nanodrop 2000c Spectrophotometer (Thermo Scientific), then stored at -80°C.

2.1.2.4 Total Protein Precipitation

Protein was extracted according to the manufacturer's instructions. The protein from the flow-through, which had been stored at 21°C during RNA purification, was precipitated by incubation in Buffer APP (Qiagen) and at 21°C for 10 minutes. The mixture was then centrifuged at 17 000 x g for 10 minutes at 21°C, to form a protein pellet. The supernatant was discarded and the resulting protein pellet was washed with 70% ethanol then centrifuged at 17 000 x g for 1 minute at 21°C. Any excess ethanol was removed using a pipette tip, and the pellet was left to dry at 21°C for 15 minutes. Since it was previously determined that the resulting protein pellets were difficult to resuspend, the protein pellet was then resuspended in 0.2% RapiGest (Waters, Mississauga, Ontario) in 50 mM ammonium bicarbonate to increase protein solubility. This solution was heated at 100°C for 5 minutes, and then left on ice for an additional 5 minutes. The solution was centrifuged at 17 000 x g for 1 minute at 21°C, to form a protein pellet. The remaining RapiGest was discarded and the protein pellet was resuspended in Buffer ALO (Qiagen). This solution was heated at 95°C for 5 minutes, and then centrifuged at 17 000 x g at 21°C for 1 minute. Protein was quantified using the Bio-Rad DC Protein Assay Kit (Bio-Rad

Laboratories, Hercules, U.S.A.), using bovine serum albumin (Life Technologies) as a standard. Once protein was quantified, it was stored at -80°C.

2.2 Control Samples

The BEAS-2B human bronchial epithelial cell line was obtained from ATCC (catalogue number: CRL-9609). These cells were used as a control sample for all DNA copy number calculations and mRNA expression calculations, following confirmation of a 2n *RBM5* gene copy number (see Chapter 3). As such, *RBM5* copy number in the test samples was compared to the BEAS-2B cells. For consistency, RBM5 mRNA expression in the test samples was also compared to the BEAS-2B cells.

2.2.1 Cell Culture

2.2.1.1 Preparation for Cell Culture

Prior to growing cells, 10 cm^2 plates were coated with 0.01 mg/ml BSA (Life Technologies, Burlington, Ontario, Canada), 0.03 mg/ml Collagen from rat tail (Sigma), 0.01 mg/ml Fibronectin (Sigma) and LHC-9 medium (Life Technologies), and placed in an incubator at 37°C overnight. In the morning, the coated plates were placed in a tupperware container, and placed in a drawer at room temperature until they were needed, within 48 hours. Prior to thawing cells, plates were washed twice with 1X PBS (Life Technologies) and then sterilized for 30 minutes under ultraviolet light.

2.2.1.2 Thawing

BEAS-2B cells with a passage number of 41 were thawed and 1 mL of the freezing medium was removed by dilution in 9 mL LHC-9 media and centrifugation for 5 minutes at 200 x g at 21°C. The supernatant was removed and the cell pellet was resuspended in LHC-9 media, and then

transferred to sterilized coated plates. Cells were placed in an incubator at 37°C and grown over a period of 2-3 days until cells were approximately 70-80% confluent.

2.2.1.3 Passaging

To pass cells, medium was removed from the plates and the cells were washed twice with 1X PBS. Trypsin/EDTA/PVP solution (0.25% Trypsin [Life Technologies], 0.53 mM EDTA [Bio-Rad] and 0.5% PVP [Sigma]) was then added to the plate and placed in the incubator at 37°C for 5-10 minutes, until the cells appeared to have lifted from the coating. Trypsin inhibitor was then added to the plate, followed by a volume of LHC-9 medium. The cells were removed from the plate, using a pipette, and transferred to a conical tube that was centrifuged at 200 x g for 5 minutes at 21°C. The supernatant was discarded and the pellet was resuspended in LHC-9 medium then dispensed at a 1:10 dilution of LHC-9 medium into sterilized coated plates. Cells were then placed in the incubator at 37°C and left to grow over a 2-3 day period. BEAS-2B cells with passage numbers of 43 and 44 were harvested, and the DNA, RNA and protein were extracted using the DNA/RNA/Protein Mini Kit (Qiagen), to use as a control for gene copy number and expression analyses.

2.2.2 *RBM5* Gene Copy Number in BEAS-2B Cells

For *RBM5* copy number analyses, it was important to identify a control sample that retained a 2n *RBM5* copy number from which *RBM5* copy number in the test samples could be compared to. The BEAS-2B cell line is a "normal" non-tumorigenic lung epithelial cell line [36]. As such, the cell line was expected to retain a normal 2n copy number of *RBM5*. To confirm that the BEAS-2B cell line retained a normal (2n) *RBM5* copy number, the *RBM5* copy number in the BEAS-2B cell line was compared to the *RBM5* copy number in non-tumour and SCLC tumour tissues from two patients that were obtained from the Ontario Tumour Bank (OTB), and two cell lines. DNA

and RNA were extracted from the OTB samples, and RNA-seq was performed on the RNA [16]. The RNA-seq analysis found that RBM5 mRNA expression in the SCLC tumours decreased by approximately 50%, compared to the non-tumour lung tissue, in both patients [16]. Loss of 50% of RBM5 mRNA expression in both patients suggested loss of one *RBM5* allele in the tumour samples, making the OTB samples an appropriate *RBM5* copy number comparator. Additionally, *RBM5* copy number in the BEAS-2B cell line was compared to the GLC20 SCLC cell line [88] and the BT-474 (ATCC, catalog number: HTB-20) human breast ductal carcinoma cell line. The GLC20 cell line was chosen as a comparator because it has a homozygous deletion of *RBM5* [16,39]. The BT-474 cell line was also chosen as a comparator because it does not contain any alterations on the short arm of chromosome 3 (ATCC, catalog number: HTB-20), the region to which *RBM5* maps to, and, therefore, is likely to retain a normal (2n) *RBM5* copy number.

DNA, RNA and protein were extracted from the cell lines and the tissue samples using the Allprep DNA/RNA/Protein Mini Kit (Qiagen), described in section 2.1.2. *RBM5* copy number analyses were completed using the qPCR technique and primers described in section 2.5.1, and the calculation described in section 2.4.3.

2.3 Reference Gene Selection

A literature review was completed to identify potentially suitable reference genes for this study, meaning genes did not change expression (for RNA) or copy number (for DNA) in the tumour, compared to the non-tumour, or in the smokers, compared to the never-smokers.

2.3.1 Reference Genes for RNA Expression Analyses

As the first objective of this study was to confirm that RBM5 was downregulated in NSCLC tumours, to prove that our cohort was representative of the literature, reference genes that did not

change expression between non-tumour and tumour tissue were necessary for accurate quantification of RBM5 expression. Two reference genes that were consistently shown to be equally expressed in NSCLC tumours and normal lung tissue were Ribosomal Protein Lateral Stalk Subunit P0 (RPLP0) and β-actin (ACTB) [89-93]. The use of two reference genes, with the intention of averaging the two for normalization purposes, was acceptable for accuracy in mRNA expression calculations [91,93].

The *RPLP0* gene maps to 12q24.23, and codes for a protein that is a component of the 60S subunit of the ribosome. The RPLP0 primers used in this study were taken from an article written by Gresner and colleagues, published in 2009 [89]. The purpose of this study was to identify stably expressed reference genes for the reliable interpretation of SYBR qPCR data when comparing NSCLC tumours to normal lung tissues. In this study, RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) from 20-25 mg of non-tumour and NSCLC tumour, from nine patients diagnosed with ADC and 12 patients diagnosed with SqCC. The extraction protocol utilized in this study was very similar to our protocol, and the researchers were also using a SYBR-based qPCR method, which we had also intended to use, making this study a suitable platform for reference gene consideration. Through the examination of Ct values, it was demonstrated that there was a slight difference of approximately one cycle in median Ct values between non-tumour and tumour RNA, but the difference was not great enough to cause inequality in expression values [89]. The suitability of RPLP0 as a reference gene was demonstrated using three separate analyses, through the examination of relative expression, and through use of softwares including NormFinder and GeNorm [89]. The suitability of RPLP0 as a reference gene for comparing normal lung tissue to NSCLC tumour tissue was also demonstrated in other analyses, using a similar platform as the Gresner *et al.* study [90-92].

The *ACTB* gene, on the other hand, maps to 7p22.1, and codes for a protein that is a component of the cytoskeletal structure. One of the reasons ACTB was chosen as a reference gene for mRNA analyses was because ACTB was also used as a normalization control in the protein analyses, included in the Appendix section. It was important to maintain consistency in normalization at both mRNA and protein levels, which is why ACTB was chosen as a normalization control for both mRNA and protein analyses. ACTB was shown to be a suitable reference gene for qPCR analyses, albeit for Taqman-based qPCR analyses, in a study completed by Saviozzi and colleagues in 2006 [91]. The purpose of the Saviozzi *et al.* study was to evaluate a number of reference genes to determine which ones were the most suitable to use for accurate normalization of gene expression in paired non-tumour and NSCLC tumour tissue [91]. Paired fresh-frozen non-tumour and NSCLC tumour samples were obtained from 18 patients diagnosed with NSCLC, including seven patients diagnosed with ADC, 11 patients diagnosed with SqCC, and one patient diagnosed with bronchioloalveolar carcinoma [91]. RNA was isolated from samples using two methods: the RNeasy 96 Kit (Qiagen) and the Biorobot 8000 (Qiagen) [91]. Looking at difference in median Ct values, it appeared as though ACTB demonstrated a difference of approximately one cycle between non-tumour and tumour tissue [91]. Since the data were shown to follow a normal distribution, it can be assumed that the median Ct values are equal to the mean Ct values [91]. This difference of one Ct value was not shown to influence expression as the fold-change in expression levels between non-tumour and tumour tissue was relatively low and demonstrated less variation, compared to other reference genes [91]. From this study, we were convinced that ACTB was a suitable reference gene for mRNA expression analyses. As the ACTB primers specific to the Saviozzi study were part of a commercially available Taqman primer/probe kit, but we intended on using a SYBR-based qPCR assay, it was

necessary to look elsewhere for primers for our study. ACTB primers for this study were obtained from a publication by Radonic *et al.* in 2004, a study that also provided information on selecting suitable reference genes for qPCR [93].

As this study also took tobacco smoke exposure history into consideration for subsequent analyses, it was also important to confirm that expression of the reference genes was not altered by tobacco smoke exposure. Although there were no studies that examined changes in reference genes after exposure to tobacco smoke, both RPLP0 and ACTB have been used in combination as reference genes in the literature for studies that examined MUC5AC RNA expression in the airway epithelium of smokers and never-smokers, as well as global miRNA expression in alveolar macrophages of smokers and never-smokers, indicating their suitability as reference genes for tobacco smoke-related expression analyses, especially when used in combination [94,95].

2.3.2 Reference Genes for DNA Copy Number Analyses

As the objective of this study was to determine if the significantly decreased RBM5 expression in the NSCLCs of smokers was the result of gene deletion, it was necessary to identify reference genes that maintained a normal (2n) copy number in non-tumour vs tumour tissue and in never-smokers vs smokers. Upon review of the literature, it was clear that there were not any publications that examined suitable reference genes for quantification of gene copy number in lung. As such, publications that examined reference genes at the mRNA level in lung tissue were utilized. We inferred that if expression of reference genes were consistent and equal between non-tumour and NSCLC tumour mRNA, it was unlikely that the DNA was altered. Following identification of candidates that did not change expression, additional review of the literature was carried out to ensure that the chromosomal region to which each gene mapped was not

commonly altered in cancer. These genes were then tested for suitability in subsequent analyses (Chapter 3).

Pumilio RNA Binding Family Member 1 (*PUM1*), which maps to 1p35.2, is a type of RBP that is involved in the translational regulation of various mRNA targets as a repressor of translation. Its RNA was shown to be stably expressed in non-tumour and NSCLC tumour tissue [92]. In fact, a study completed by Soes *et al.* in 2013, found that PUM1 was one of the most stably expressed genes among 23 candidate reference genes. The short arm of chromosome 1 was not shown to be commonly altered in NSCLCs, including ADC and SqCC, and was not shown to be commonly altered in SCLC either, as chromosome 1 is not one of the common deletion or duplication regions of the lung cancer genome [76]. Alterations in chromosome 1 have been observed in some lung cancers, including the amplification of 1q21-q25 in SqCC and LOH at 1p36 in both SCLC and NSCLC, but the 1p35 region was not altered in lung cancers, meaning that the *PUM1* gene was likely neither deleted nor amplified, and the gene retained a normal (2n) copy number in lung cancers [76,96].

In order to keep a level of consistency in reference genes across DNA copy number and mRNA expression analyses, *RPLP0* was also considered as a candidate reference gene for DNA copy number analyses. As previously mentioned, RPLP0 was chosen as a reference gene for mRNA expression analyses because it was considered to be stably expressed in non-tumour and NSCLC tumour tissue in multiple studies [89-92]. *RPLP0* maps to the chromosomal region 12q24.23. The long arm of chromosome 12 was not included in the commonly deleted or amplified regions of the lung cancer genome [76]. Only the short arm of chromosome 12 was shown to be amplified, in SqCC, specifically [97]. Given this information, it did not appear as if the *RPLP0* gene was likely

to undergo copy number gains or deletions, and was, therefore, likely suitable to use as a reference gene for copy number analyses, in combination with *PUM1*.

Since *RBM5* copy number was compared between smokers and never-smokers in this study, it was crucial to confirm that aberrations at chromosome 1p35.2 and 12q24.23 did not occur as a result of tobacco smoke exposure. Since molecular aberrations in these chromosomal regions did not occur in lung cancers that were associated with tobacco smoke exposure, such as SqCC or SCLC, we inferred that these regions likely remained unaltered in patients with tobacco smoke exposure.

2.4 RBM5 mRNA Expression Analysis

RNA expression was analysed in HSN test samples relative to the BEAS-2B control sample. RNA was reverse-transcribed into cDNA, and the cDNA was amplified using a SYBR-based qPCR method. RBM5 primers spanning exons 6 and 7 were used to measure RBM5 expression, while RPLP0 and ACTB primers spanning exons 2 and 3 and exons 1 and 2, respectively, were used as normalization controls.

2.4.1 Reverse Transcription

RNase-free water, containing 0.5 micrograms (µg) of RNA, was combined with 10 mM Deoxynucleotide Solution Mix (New England Biolabs, Pickering, Ontario, Canada) and 100 µg/µl oligo-dT primer (Alpha DNA, Montreal, Quebec, Canada). This solution was incubated at 65°C for 5 minutes, and then placed on ice. While on ice, 5X First Strand Buffer (Invitrogen) and 0.1 molar (M) dichlorodiphenyltrichloroethane (DTT) (Invitrogen) was added to the solution, then incubated at 42°C for 2 minutes. Finally, 200 U of MMLV Reverse Transcriptase

(Invitrogen) was added to the solution and incubated at 42°C for 50 minutes, followed by 70°C for 15 minutes. Resulting cDNA was stored at -20°C.

2.4.2 Quantitative PCR (qPCR)

For mRNA expression analyses, a SYBR-based qPCR assay was used. In this method, SYBR dye intercalates with double-stranded DNA, and its fluorescence increases with each successive cycle, as PCR product is amplified. When amplification of a gene reached an exponential phase, the cycle at which the exponential phase began, and rose above the background fluorescence, was referred to as the threshold cycle (Ct) [98-101]. The Ct values were used for all expression calculations. Standard curves for each of the amplified genes were also constructed, and the slope of each standard was used to measure efficiency of the standard curve. The efficiency of the standard curve is important for understanding how efficiently the PCR amplicon doubled during the exponential phase [100]. Ideally, the efficiency of the standard curve should have been close to 100%, but this was not always the case [100]. Using both the Ct values and the efficiency of the standard curve, mRNA expression was calculated using a comparative Ct method. Furthermore, it was also important that the standard curve was linear across all data points, as denoted by an R^2 value that was greater than 0.9 [100]. Any plates in which a standard curve had an R^2 value below 0.9 were not used for mRNA expression calculations.

For each patient sample, three technical replicates for each gene were examined on a single 96-well plate, to control for technical errors. Duplicate assays were completed, to control for plate-to-plate variability. Mean RBM5 mRNA expression values were calculated using the average of all six technical replicates. Outliers within each set of six technical replicates were identified using the Grubb's test, which is capable of identifying only one outlier in each data set [102]. When

a significant outlier was detected, the outlier value was removed from the calculation of the mean RBM5 mRNA expression value.

cDNA samples were diluted in water and combined with iTaq Universal SYBR Green Supermix (Bio-Rad, Mississauga, Ontario, Canada), and 0.75 micromolar (μM) (RBM5, RPLP0 and ACTB) forward and reverse primers (Alpha DNA) (Table 2). Standard curves were prepared using a serial dilution of BEAS-2B cDNA, with dilutions ranging between 1:1 and 1:625. The qPCR reaction was performed on the Aria Mx Realtime PCR System (Agilent, Mississauga, Ontario, Canada) using the following program: 95°C for 5 minutes, followed by 40 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 45 seconds, and concluded with one dissociation curve cycle consisting of temperatures ranging between 95°C and 60°C. The dissociation curve was used to confirm the specificity of PCR products and identify the presence of primer dimers and contaminants [100].

Table 2 Primers used for mRNA qPCR analysis

Gene	Primer	Primer Sequence (5' > 3')	Exon	Size of Amplicon (bp)	Annealing Temperature (°C)
RBM5	RBM5 Exon6F RBM5E7R	F: GTG TAA GCC GTG GTT TCG CCT TC R: TTG CAA TGT GCT TTC CTT GA	6 7	108	60
RPLP0	RPLP0E2F RPLP0E3R	F: CTG ATG GGC AAG AAC ACC AT R: GTG AGG TCC TCC TTG GTG AA	2 3	115	60
ACTB	ACTB-F ACTB-R	F: CAG AGC CTC GCC TTT GCC GA R: GAA GCC GGC CTT GCA CAT GC	1 2	111	60

RBM5 primers were previously used in the Sutherland Lab for the measurement of RBM5 expression by qPCR [103]. RPLP0 primers were adapted from an article by Gresner and colleagues, and ACTB primers were adapted from an article by Radonic and colleagues, published in 2009 and 2004, respectively [89,93]. The specificity of the primers was confirmed using the online OligoAnalyzer 3.1 Tool (Integrated DNA Technologies, Coralville, Iowa, U.S.A.) and the efficiency of the primers was confirmed (Chapter 3). Reference genes were also validated for stable expression in non-tumour and tumour tissue (Chapter 3).

2.4.3 Calculation of RBM5 mRNA Expression

For calculation of RBM5 mRNA expression, based on a fold-change between a test sample and a control sample, the Pfaffl method was used, with the Hellemans correction [104,105]. The Pfaffl method measures the expression ratio of a test gene in a test sample and a control sample, and compares this ratio to the expression ratio of a single reference gene in a test sample and a control sample, and corrects for changes in standard curve efficiencies [99,101,104,105]. Since the Pfaffl method only considers one reference gene, and more than one reference gene is now considered preferable for accurate quantification in qPCR analyses, the Hellemans correction was used [105].

BEAS-2B RNA represented the standardized expression of RBM5 and reference genes, serving as the control against which all the HSN test samples, both non-tumour and tumour, were measured. Ct values were compared between control and test sample RNA, to calculate the fold-changes in expression. These Ct comparisons were made for both the gene of interest, RBM5, and the reference genes, RPLP0 and ACTB.

Prior to using the Pfaffl method, the efficiency of the standard curve for each gene primer pair needs to be calculated (Equation 1).

$$\text{Equation 1: Efficiency} = 10^{(-1/\text{slope})}$$

Using Equation 1, the calculated efficiency value that is reflective of 100% efficiency is an efficiency value of 2, as reflected by a slope of approximately -3.32 [100]. Reduced efficiency will result in an efficiency value below 2, while increased efficiency will result in an efficiency value above 2, according to Equation 1.

Standard curve efficiencies between the primer pairs amplifying a test gene and the primer pairs amplifying each reference gene will likely not be identical, as may the standard curve efficiencies between the same primer pair used in plate replicates. When efficiency is below the ideal of 100%, and amplicon doubling capacity is decreased, the resultant Ct values would be higher than expected if the primers were 100% efficient. Alternatively, when efficiency is higher than 100%, the amplicon doubling capacity is increased, and the resultant Ct values would be lower than expected if the primers were 100%. Therefore, interpretation of Ct values is meaningless unless primer efficiencies are known.

Once efficiency values are calculated, Equation 2 is used to measure test gene expression in the test sample(s) compared to the control sample, while also normalizing test gene expression to the reference gene(s).

$$\text{Equation 2: Test expression} = \frac{\text{Efficiency (test)}^{\text{Ct control-Ct test}}}{\sqrt[2]{\text{Efficiency (Ref 1)}^{\text{Ct control-Ct test}} \times \text{Efficiency (Ref 2)}^{\text{Ct control-Ct test}}}}$$

2.5 *RBM5* DNA Copy Number Analysis

A SYBR-based qPCR method was used for *RBM5* copy number analyses. *RBM5* gene copy number in the HSN and Slovakian test samples was measured relative to the BEAS-2B control sample, using a comparative Ct method. This method was nearly identical to the mRNA expression analyses described above.

2.5.1 Quantitative PCR (qPCR)

Similar to the mRNA analyses, a SYBR-based qPCR method was used. The Ct values of test samples and the control sample, as well as the efficiency of the standard curves, were necessary for calculation of *RBM5* copy number. Standard curves were required to be linear for accurate quantification of *RBM5* copy number, and, therefore, required an R^2 value that was greater than 0.9 [100].

For each patient sample, three technical replicates for each gene were examined on a single 96-well plate to control for technical errors. Duplicate assays were completed, to control for plate-to-plate variability. The 96-well plate set-up is displayed in Figure 5. Mean copy number values were calculated using the average of all six technical replicates. An outlier within each of the six technical replicates was identified using the Grubb's test, which is capable of identifying only one outlier in each data set [102]. If a significant outlier were detected, the outlier value was removed from the calculation of the mean copy number value.

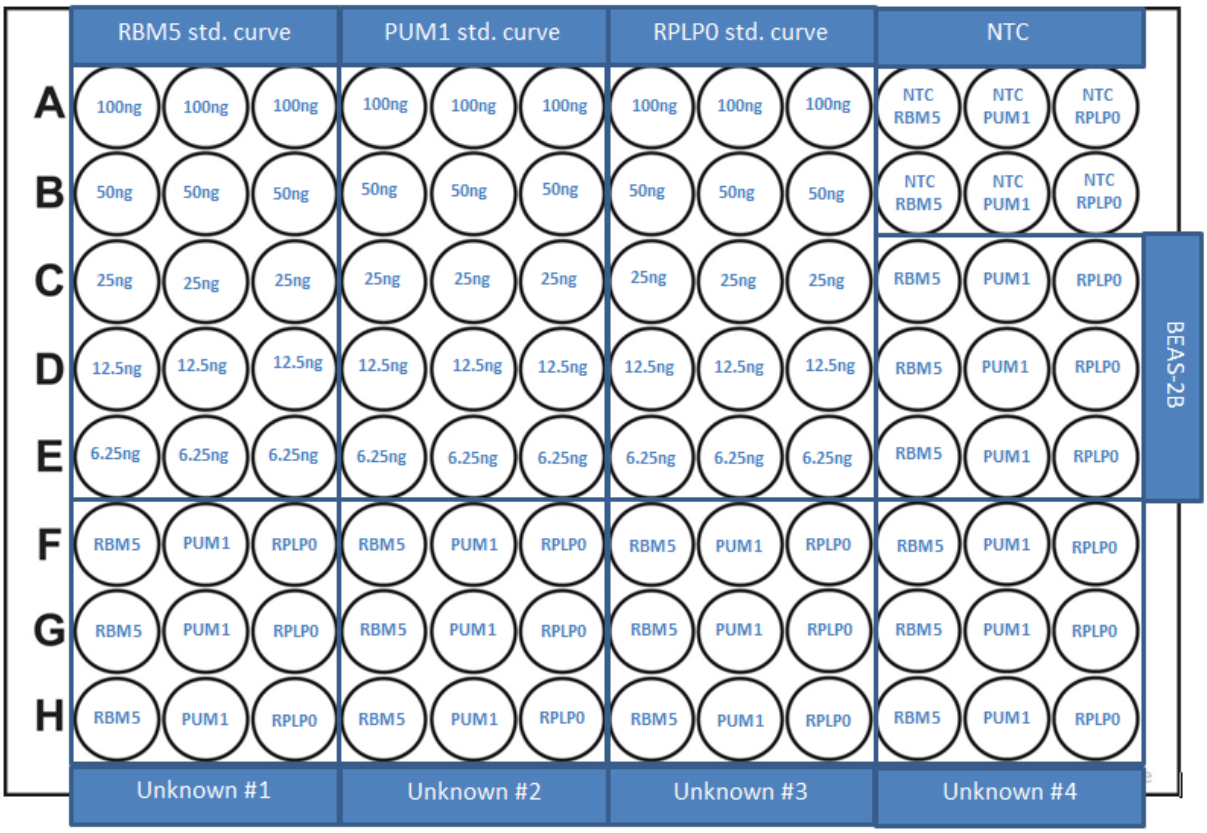


Figure 5 96-well plate set-up for the *RBM5* gene copy number qPCR assay. Three standard curves were constructed in triplicate, for *RBM5*, *PUM1* and *RPLP0*, respectively, using BEAS-2B DNA. Standard curves contained five data points, with BEAS-2B DNA quantities ranging from 100 ng-6.25 ng. Four test samples (Unknowns #1-4) and the BEAS-2B control sample were also included on the 96-well plate. *RBM5*, *PUM1* and *RPLP0* measurements were completed, in each test sample and the control sample, in triplicate, using 25 ng of DNA. No template controls (NTC) were included for each gene, in duplicate, using water in the place of DNA in the reaction mixture.

Samples were prepared using 25 ng of patient DNA, 1X All-in-One qPCR Mix (Genecopoeia, Rockville, Maryland, U.S.A.), and 1.5 µM (*RBM5*) or 0.75 µM (*PUM1* and *RPLP0*) forward and reverse primers (Alpha DNA) (Table 3). BEAS-2B DNA was used as a control sample, and prepared identically to the test samples, with 25 ng of DNA. The 25 ng addition of test and control DNA was optimal, as the amplified DNA was expected to fall within the middle of the standard curve, for accurate interpolation of data. Standard curves containing DNA quantities

between 100 ng and 6.25 ng were prepared using a serial dilution of BEAS-2B DNA. The qPCR reaction was carried out using the AriaMx Realtime PCR System (Agilent), using the following program: 95°C for 5 minutes, followed by 40 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 45 seconds. Following amplification, one dissociation curve cycle consisting of temperatures ranging between 95°C and 60°C was completed for each sample and no template control (NTC), to confirm the specificity of the PCR products and identify any primer dimers or contaminants.

RBM5 primers were designed by a former student in the Sutherland lab. Primers mapping to exon 1 of the *RBM5* gene were chosen because it was in a region of the *RBM5* gene that was transcribed, and also because it was the exon that was closest to the telomere of 3p. The placement of primers in exon 1 enabled detection of deletions that were likely due to loss of an entire allele, that was part of the 3p21.3 deletion region, and not loss of a partial segment of the *RBM5* gene, that would have resulted in a truncated transcript. Forward and reverse primers were designed in exons 13 and 2 of *PUM1* and *RPLP0*, respectively. The online OligoAnalyzer 3.1 Tool (Integrated DNA Technologies) was used to confirm the specificity of the primers and the efficiency of the primers was confirmed in Chapter 3. Reference genes were also validated for stable expression in non-tumour and tumour tissue, and in smokers and never-smokers in Chapter 3.

Table 3 Primers used for DNA qPCR analyses

Gene	Primer	Primer Sequence (5' > 3')	Exon	Size of Amplicon (bp)	Annealing Temperature (°C)
RBM5	QRBM5 E1F1 QRBM5 E1R1	F: CGG AGG CGC CAT TTT GT R: GAA GCA GCA GTA GCG GTT CTG	1	72	60
PUM1	PUM1-F PUM1-R	F: TGG ACC ATT TCG CCC TTT AG R: CAG AGA GTT GTT GCC GTA GAA	13	103	60
RPLP0	RPLP0-F RPLP0-R	F: AGA TCC GCA TGT CCC TTC G R: CCT TGC GCA TCA TGG TGT T	2	66	60

2.5.2 Calculation of *RBM5* Copy Number

Calculation of *RBM5* copy number was also completed using the Pfaffl method with the Hellemans correction, exactly how RBM5 mRNA expression was calculated [99,104,105].

The BEAS-2B cell line was used as the control sample for all copy number calculations. The cell line was experimentally determined to have two copies of the *RBM5* gene (see Chapter 3). As the BEAS-2B cells had two copies of *RBM5*, it served as a platform that was representative of the results anticipated for normal (2n) copies of *RBM5*. Comparing the test samples to the control sample allowed us to calculate a relative copy number value for each of the test samples, that was reflective of fold-changes either equating to or deviating from a normal (2n) copy number.

Equations 1and 2, as described in Section 2.4.3, were used to calculate the efficiency value and test gene copy number.

2.6 Statistical Analyses

The strategy for statistical analyses completed in this study is described in Figure 6, through use of a flow-chart. Statistical analyses were carried out using GraphPad Prism 5 software (GraphPad Software Inc., San Diego, California, U.S.A.).

Once *RBM5* copy number or RBM5 expression values were calculated for each patient, data sets were examined to determine if they followed a normal distribution. To test for normality, the Shapiro-Wilk test was used [106]. If the Shapiro-Wilk test resulted in a *p*-value of less than 0.05, the data were interpreted as not following a normal distribution. In such cases, non-parametric tests were used in subsequent analyses. If the resulting *p*-value was greater than or equal to 0.05, parametric tests were used in subsequent analyses. In this cohort, all data sets did not follow a normal distribution, and, therefore, non-parametric analyses were used.

For continuous data, such as RBM5 mRNA expression, data was displayed in the form of a boxplot. The Tukey method was chosen to identify outliers, as it identifies outliers that do not fit within interquartile ranges, defined by the parameters of the boxplot [107]. Non-parametric tests for continuous data sets included the Mann-Whitney test and the Wilcoxon signed-rank test. The Mann-Whitney test was used to compare the means of two independent variables (i.e. smoker vs never-smoker, or, patients with a 1n *RBM5* copy number vs patients with a 2n *RBM5* copy number) [108]. If the Mann-Whitney test resulted in a two-tailed *p*-value of less than 0.05, then the difference between the means of the two groups was considered to be statistically significant. Another non-parametric test that was used throughout this study was the Wilcoxon signed-rank test, used to measure statistically significant differences in paired samples (i.e. non-tumour vs tumour, from the same patients) [109]. If the difference in rank totals resulted in a *p*-value of less than 0.05, the difference was considered statistically significant.

For categorical data, such as *RBM5* gene copy number, the data were displayed in the form of a bar graph. The 3x2 Fisher's exact test was used to identify significant differences in categorical data from a contingency table [110]. For this test, the proportion of patient samples that fit into a particular *RBM5* copy number category (1n, 2n or 3n) was analysed. Differences in the proportion of samples that fit into a particular copy number category were compared between two groups (i.e. non-tumour vs tumour, smoker vs never-smoker). If the proportion of patient samples that fit into a particular copy number category was significantly different between the two groups, a two-tailed *p*-value of less than 0.05 was obtained using the 3x2 Fisher's exact test.

The Kruskal-Wallis test was used to examine the relationship between *RBM5* copy number and RBM5 mRNA expression. This test determined if the medians between two or more groups were significantly different [111]. Groups were established based on categorical data (i.e. copy number). Continuous data (i.e. expression) were then sorted according to the established groups. The Kruskal-Wallis test determined if there was a statistically significant difference between two or more groups, denoted by a two-tailed *p*-value of less than 0.05, but did not specify which groups were significantly different. In the instances where the Kruskal-Wallis test identified a significant difference among groups, the Mann-Whitney test was used to compare two groups at a time, in order to identify the specific groups that were significantly different.

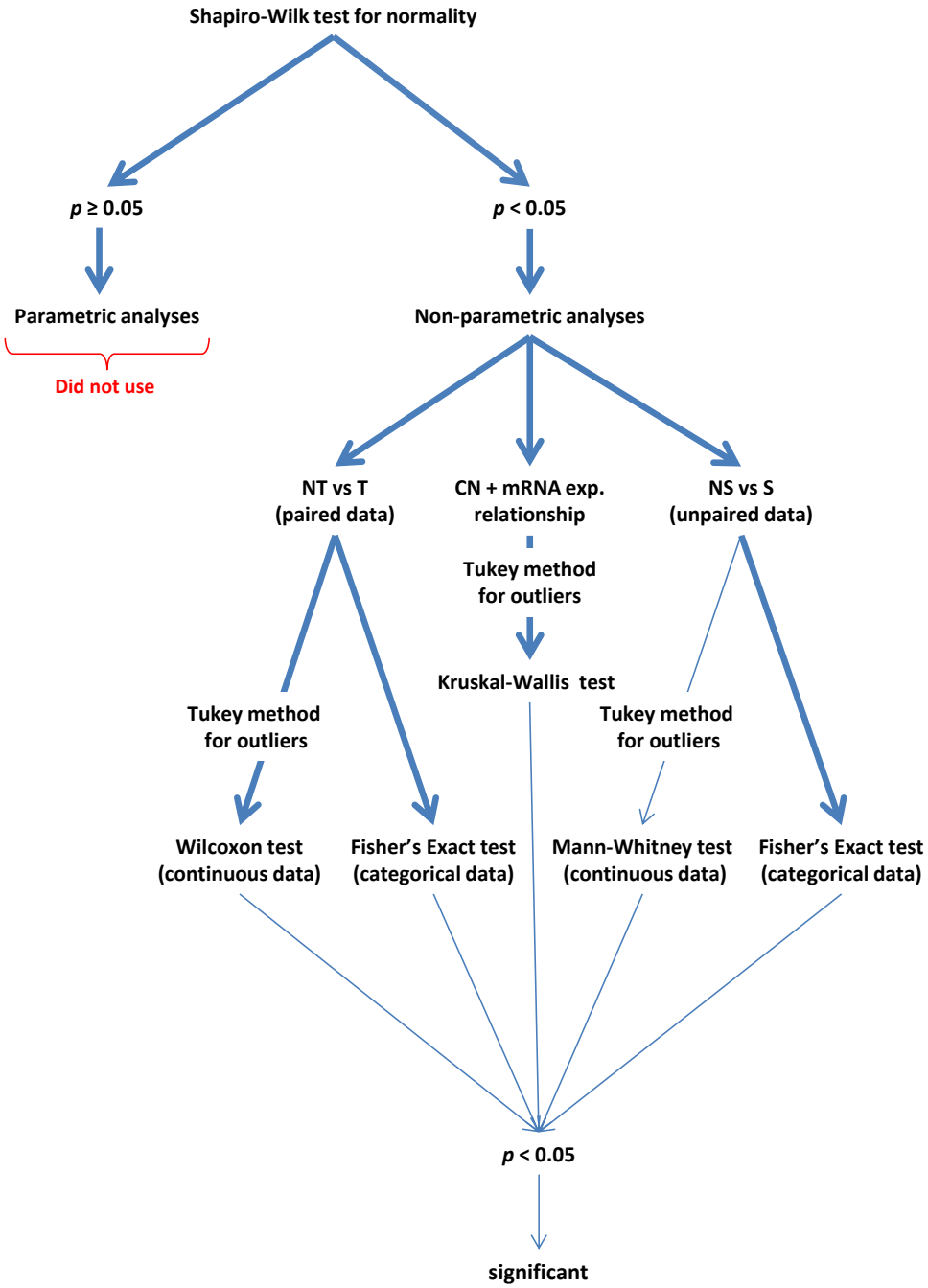


Figure 6 Flow-chart describing strategy for statistical analyses. Data sets were tested for normality using the Shapiro-Wilk test. All data sets did not follow a normal distribution, therefore, non-parametric analyses were used. For continuous data that was represented as a boxplot, the Tukey method was used to identify outliers in each data set, and outliers were excluded prior to further statistical analyses. For categorical data, the 3x2 Fisher's exact test was used. Significance was defined as having a two-tailed *p*-value of less than 0.05. NT = non-tumour, T = tumour, CN = copy number, exp = expression, NS = never-smoker, S = smoker. Bold arrows indicate which statistical tests were used in this study.

Chapter 3

3 Experimental Validation

Before carrying out *RBM5* gene copy number analyses, it was necessary to establish that the HSN cohort was representative of those previously presented in the literature [29,40], then determine the validity of the experimental procedures used in this study, including the confirmation of a normal 2n copy number in the BEAS-2B control cell line, the suitability of the qPCR primers and the suitability of the reference genes. Firstly, RBM5 mRNA expression in the HSN cohort was examined to confirm that RBM5 expression decreased in NSCLC tumours, compared to adjacent non-tumour tissue. Since the objective of this study was to determine if the previously noted significantly decreased levels of RBM5 expression in NSCLCs from smokers [40] correlated with *RBM5* deletion, it was crucial to confirm that the HSN cohort had significantly decreased RBM5 expression levels [29,40]. Secondly, *RBM5* gene copy number in the BEAS-2B cell line was compared to *RBM5* gene copy number in patient samples and cell lines to confirm that the BEAS-2B cell line contained two copies of *RBM5*, and was, therefore, a suitable control sample. Thirdly, all primer pairs, for both DNA and RNA qPCR assays, test genes and reference genes, were examined to confirm that primers were both efficient and specific. Both efficiency and specificity of primers were required for precision and reproducibility when conducting qPCR analyses [112]. Lastly, the reference genes, for both DNA and RNA analyses, were examined to confirm stable expression between non-tumour and tumour tissue, and between samples from smokers and never-smokers. Stable expression of reference genes was required for accurate normalization when *RBM5* gene copy number and mRNA expression were calculated [91,93,113].

3.1 Cohort Validation

Before analyzing *RBM5* copy number in the patient samples, it was important to show that the HSN cohort was representative of those already described in the literature, with reduced RBM5 expression occurring in tumour, compared to non-tumour, samples [29,40]. To determine if our model was representative, we examined RBM5 mRNA expression in NSCLC tumour and adjacent non-tumour tissue from 53 patients, 39 and 14 of which were diagnosed with ADC and SqCC, respectively. For each patient sample (non-tumour or tumour), we obtained expression values from six technical replicates. Expression values per sample were averaged, excluding any within-sample outlier identified by the Grubb's test. All non-tumour or tumour RBM5 expression values were then graphed, excluding any outliers identified using the Tukey method. Using the Shapiro-Wilk normality test, the normalized RBM5 expression values in the non-tumour data set and the tumour data set were found to not follow a normal distribution. Since the non-tumour and tumour samples were paired, a Wilcoxon signed-rank test was used to identify if RBM5 expression was significantly decreased in NSCLC tumours, compared to normal adjacent lung tissue.

The grouped non-tumour and tumour boxplots with statistical outcomes are presented in Figure 7. Expression results from patients diagnosed with both ADC and SqCC were combined, to represent all patients diagnosed with NSCLC. Comparing non-tumour and tumour RNA, it was shown that RBM5 expression in NSCLC tumours was significantly downregulated (Figure 7 A). ADC and SqCC were then examined separately, to confirm that RBM5 was downregulated in both NSCLC subtypes. Indeed, RBM5 expression decreased significantly in both ADC (Figure 7 B) and SqCC (Figure 7 C). Therefore, the HSN cohort was a suitable model for the examination of mechanisms contributing to RBM5 downregulation in the lung cancers of smokers.

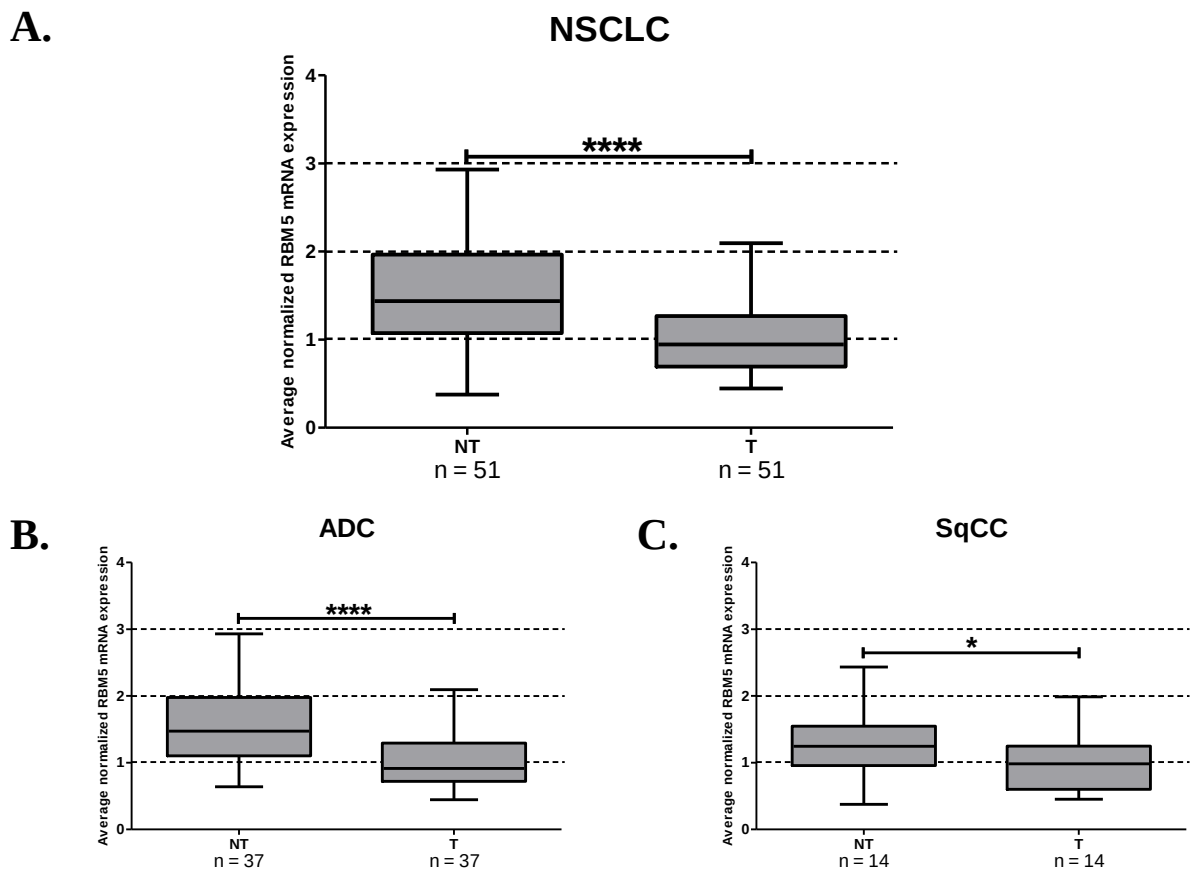


Figure 7 Confirmation of RBM5 downregulation in NSCLC tumours. Wilcoxon signed-rank tests were used to confirm that RBM5 mRNA expression decreased significantly in (A) NSCLC, (B) ADC and (C) SqCC tumours, compared to adjacent normal lung tissue. *p<0.05, ****p<0.0001.

3.2 Control Sample Validation

The BEAS-2B cell line was chosen as a control sample for this study, since it was expected to retain a normal (2n) *RBM5* copy number, because it is a "normal" non-tumorigenic cell line [36]. Before commencing *RBM5* gene copy number analyses on the HSN test samples, it was necessary to confirm that the BEAS-2B cell line indeed retained a normal two copies of the *RBM5* gene. *RBM5* gene copy number in the BEAS-2B cell line was compared to the GLC20 SCLC cell line and the BT-474 breast cancer cell line, and also to two paired non-tumour and SCLC tumour samples. As anticipated, the GLC20 cell line, which has a homozygous deletion of

RBM5, did not amplify any *RBM5* DNA by qPCR, and demonstrated a calculated *RBM5* copy number of zero, when compared to the BEAS-2B cell line [16,39] (Figure 8). Since the BT-474 cell line does not harbour any alterations on the short arm of chromosome 3 (ATCC, catalog number: HTB-20), it was anticipated that the BT-474 cell line would retain a normal (2n) *RBM5* copy number. BEAS-2B and BT-474 cell lines amplified identical quantities of *RBM5* DNA, by qPCR, indicating that both cell lines retained two copies of *RBM5*. The OTB samples were previously shown to have a 50% reduction of RBM5 mRNA expression in SCLC tumours, compared to adjacent non-tumour samples, suggesting a loss of one *RBM5* allele in both tumour samples [16]. As anticipated, both non-tumour samples amplified similar quantities of *RBM5* DNA as the BEAS-2B cell line, while both SCLC tumours amplified only half the amount of *RBM5* DNA. These results suggested that both non-tumour samples retained two copies of *RBM5*, while

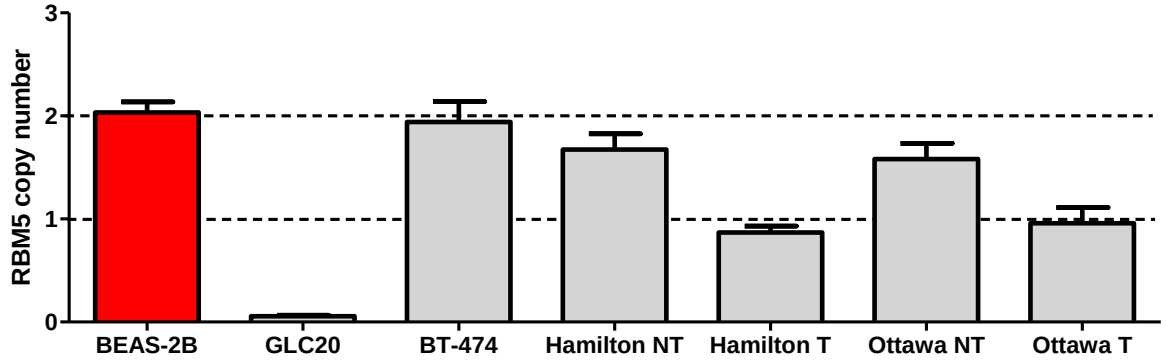


Figure 8 Validation of *RBM5* gene copy number in the BEAS-2B cell line. *RBM5* gene copy number in the BEAS-2B cell line was compared to *RBM5* gene copy number in two cell lines (GLC20 and BT-474) and two paired non-tumour and tumour samples (Hamilton and Ottawa NT and T). NT = non-tumour, T = tumour.

the SCLC tumour samples only retained one copy of *RBM5*. Taking these results as a whole, the BEAS-2B cell line was confirmed to be a reliable comparator for the calculation of *RBM5* copy number in various test samples.

3.3 Primer Validation

Before commencing with qPCR analyses, all primer pairs were validated to ensure that the chosen primers were efficient and specific [112]. Amplification efficiency describes how efficiently the PCR products doubled after each successive cycle, while specificity refers to the degree of uniqueness of primer sequence recognition.

With regards to primer efficiency, ideally, amplification efficiency should be close to 100%, as described by a standard curve slope of -3.32 [100]. In order to accurately interpret the efficiency of the primers, it is necessary that the data points on the standard curve are amplified at a linear rate, as demonstrated by a correlation coefficient (R^2) value close to 1 [100]. Due to both inter- and intra-assay variability, obtaining an efficiency of 100% is not always possible. Generally, standard curves result in efficiency values between 1.6 and over 2, corresponding to efficiencies of 60% to over 100% [114,115]. Efficiency values can fluctuate because of variability in estimated baseline fluorescence, contaminating Taq inhibitors, suboptimal annealing temperatures, amplicon secondary structures and poorly designed primers [116,117]. As such, the literature strongly recommends an efficiency-corrected mathematical model, such as the Pfaffl method, for the interpretation of data with any fluctuation in efficiencies [114].

With regards to primer specificity, primers were validated to confirm that one specific PCR product was amplified. Specificity was confirmed through melt-curve analyses, where the dissociation of double-stranded DNA PCR product should result in a single, sharp peak [100].

3.3.1 Primers for DNA Copy Number Analysis

RBM5, *PUM1* and *RPLP0* primers were used for *RBM5* copy number analyses by qPCR. Representative standard curves, efficiency values, and melt-curves are displayed in Figure 9. A total of 115 assays were completed for this study, and all standard curve efficiencies were above 60%. Average *RBM5* efficiency values fluctuated around 73.23%, while average *PUM1* and *RPLP0* efficiency values were 103.16% and 102.90%, respectively. Since *RBM5* PCR products resulted in efficiency values below 100%, it was important to use an efficiency-corrected method, such as the Pfaffl method, for quantifying *RBM5* expression [114]. Because the primers were specific for one gene, and the PCR product was amplified at a linear rate, it was confirmed that the primer pairs were acceptable for subsequent *RBM5* copy number analyses using an efficiency-corrected comparative Ct method.

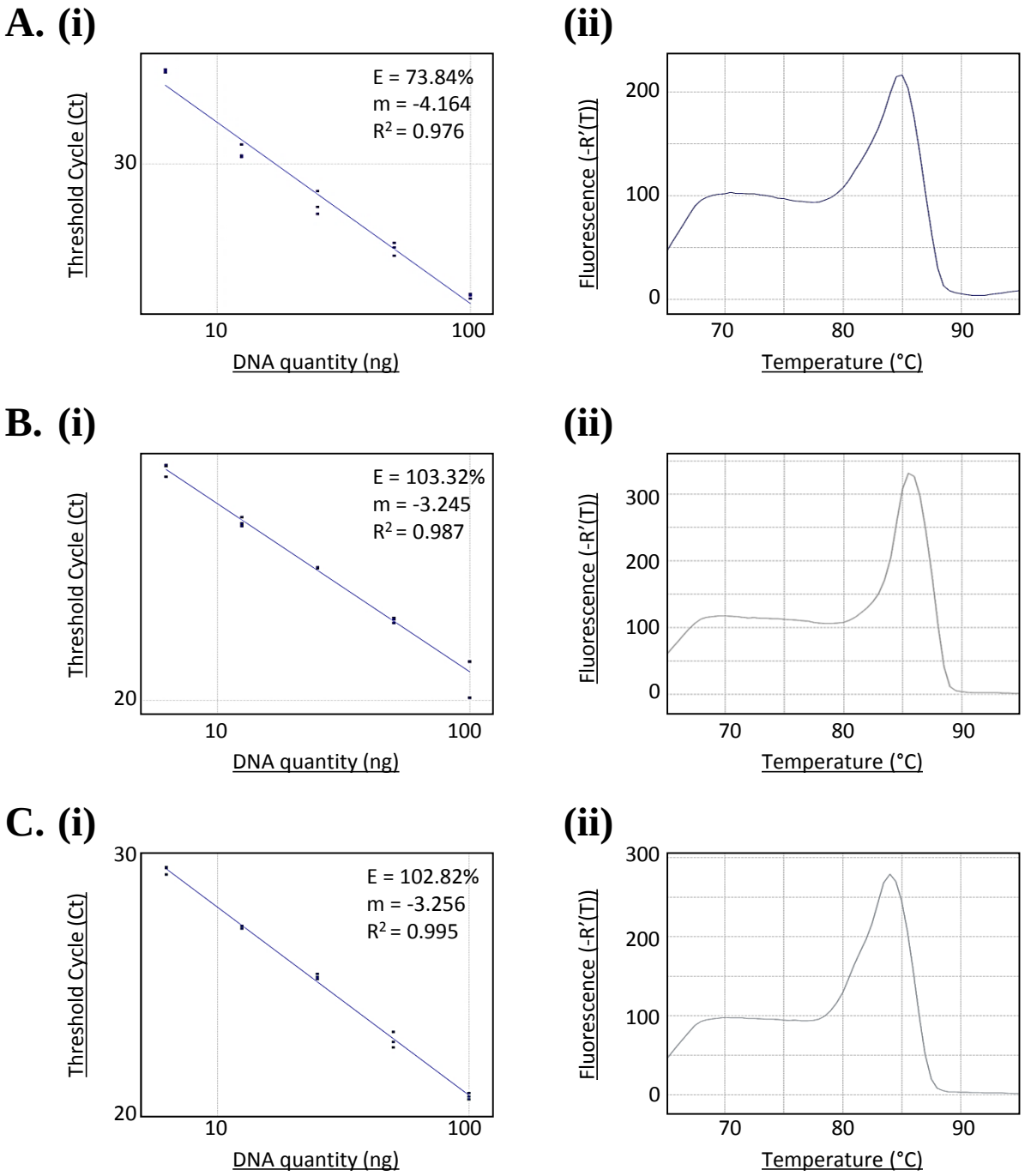


Figure 9 Confirmation of efficiency, linearity and specificity of DNA primers. (A) *RBM5*, (B) *PUM1* and (C) *RPLP0* primer pairs were assessed. (i) Efficiency, E, and linearity, R^2, were determined using standard curves of five BEAS-2B DNA dilutions between 100 ng and 6.25 ng, with technical triplicates. (ii) Specificity of primer pairs was determined using melt curves with dissociation temperatures between 65°C and 95°C.

3.3.2 Primers for RNA Expression Analysis

RBM5, RPLP0 and ACTB primers were used for RBM5 mRNA expression analyses by qPCR. Representative standard curves and melt curves for each primer pair are displayed in Figure 10. The RBM5 primers, which were distinct from the primers used for gene copy number analysis, were used previously in the Sutherland lab, and the primer efficiency was confirmed [103]. The previous validation of RBM5 primers resulted in an efficiency of 95%, using dilutions of H9c2 cDNA, derived from rat myoblasts [103]. In this study, a total of 65 assays were performed, and the average efficiency of RBM5 primers was 104.30%. This efficiency percentage was close to the ideal 100%, confirming that RBM5 PCR products were doubling with each successive cycle [100]. The RPLP0 primers, which were also distinct from the primers used for gene copy number analysis, were generated from published primer sequences [89]. In the Gresner *et al.* study, the reported efficiency of the primers was 101.3% [89]. In our study, however, the average efficiency was 113.89% for these RPLP0 primers. It was noted that the Gresner *et al.* study used an annealing temperature of 57°C, whereas this study used a slightly higher melting temperature of 60°C, which can affect the efficiency of the primers and result in the observed differences [89,117]. The ACTB primers that were used in this study were also obtained from the literature [93]. In the Radonic *et al.* study, the ACTB primers had an efficiency percentage of 99% [93]. In our study, however, the efficiency of the primers was higher, at an average of 123.29%. Once again, differences in annealing temperatures were noted. The Radonic *et al.* study used an annealing and extension temperature of 67°C, whereas this study used an annealing temperature of 60°C and an extension temperature of 72°C, which may have influenced the observed changes in primer efficiency [93,117]. Since reference gene primer efficiencies fluctuated at values slightly higher than 100%, it was important to use the efficiency-corrected Pfaffl method for all

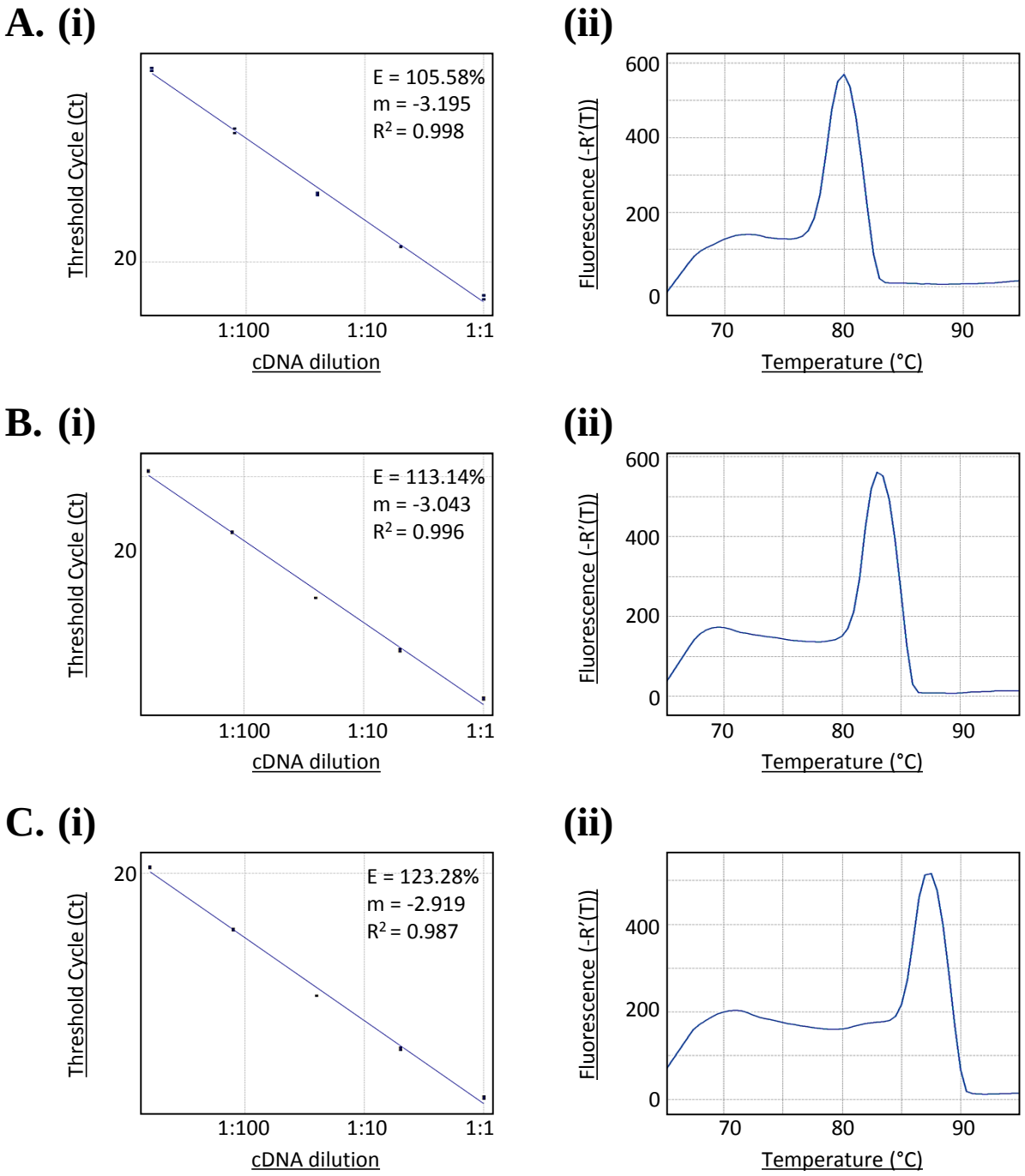


Figure 10 Confirmation of efficiency, linearity and specificity of RNA primers. (A) RBM5, (B) RPLP0 and (C) ACTB primer pairs were assessed. (i) Efficiency, E, and linearity, R^2, were determined using standard curves of five BEAS-2B cDNA dilutions between 1:1 and 1:625, with technical duplicate. (ii) Specificity of primer pairs was determined using melt curves with dissociation temperatures between 65°C and 95°C.

subsequent RBM5 expression calculations [114]. Each primer pair produced one specific gene product that amplified at a linear rate, confirming that the primers chosen for RBM5 expression analyses were acceptable.

3.4 Reference Gene Validation

In qPCR analyses, normalizing to a reference gene is essential for eliminating variability in DNA and cDNA quantity or quality [89,91,93]. As such, it is important to confirm that reference genes do not change in different tissue types or under different experimental conditions [92]. To be certain that *RBM5* gene copy number and RBM5 expression were accurately quantified by qPCR, references genes for both DNA and RNA analyses were validated. To validate the reference genes, it was necessary to confirm that the amount of *PUM1* and *RPLP0* DNA, as well as the expression of RPLP0 and ACTB, did not change between the two groups being compared (i.e. non-tumour vs tumour, and smoker vs never-smokers). Validation of each reference gene was carried out by comparing Ct values, and also by examining relative changes in DNA or RNA quantities.

3.4.1 Reference gene quantities did not change between non-tumour and tumour DNA or RNA

First, changes in the quantities of *PUM1* and *RPLP0* DNA were compared between non-tumour and tumour tissue. For this analysis, data from paired non-tumour and tumour samples, obtained from 65 patients, were used. The first analysis involved the comparison of mean Ct values. Distributions of raw Ct values have been used throughout the literature to validate reference genes [89,91]. If Ct values were equivalent, it was assumed that reference gene expression was also equivalent. For each DNA sample, Ct values from six technical replicates were averaged. Any outlier within each of the six technical replicates was identified using the Grubb's test, which is

only capable of identifying one outlier. If an outlier were detected, it was removed from the calculation of the mean Ct value for that particular sample. Mean Ct values from non-tumour and tumour tissue from each of the 65 patients were then graphed using boxplots (Figure 11 Ai). Outliers within each group were identified using the Tukey method, and excluded prior to statistical analyses. The Shapiro-Wilk test was performed to examine the distribution of data. One of the groups did not follow a normal distribution, and, as such, non-parametric analyses were used. Since the non-tumour and tumour DNA were obtained from the same patients, and therefore paired, a Wilcoxon signed-rank test was performed to identify statistically significant differences in reference gene Ct values between non-tumour and tumour DNA. It was determined that Ct values for each reference gene did not change between non-tumour and tumour DNA, confirming that the quantities of reference gene DNA was unlikely to be altered in the different tissue types.

Since Ct distributions only provide a rough estimate of amplification changes, a secondary analysis was recommended to confirm that the quantities of reference gene DNA did not change between non-tumour and tumour tissue [91]. The second analysis examined the quantity of reference gene in the tumour, relative to the non-tumour. DNA from the same 65 patients was also used in this analysis. The quantity of each reference gene was calculated using the relative standard curve quantification method, using the formula $\log 10^{[(\text{Ct reference gene} - \text{y-intercept})/(\text{-slope})]}$, where the y-intercept and slope values were taken from the *PUM1* and *RPLP0* standard curves [118,119]. This calculation was performed on each of the six technical replicates. Outliers in each of the six technical replicates were identified using the Grubb's method. For each patient sample, the quantity of reference gene was averaged from the six technical replicates, excluding any outliers. To demonstrate the suitability of using the two reference genes together, the geometric

mean of the quantities of both *PUM1* and *RPLP0* was also calculated. The mean quantity of reference gene in the tumour DNA was then normalized to the mean quantity of reference gene in the non-tumour DNA, for each patient. If the quantities of reference gene DNA remained constant between both tissue types, the normalized value was close to 1. The normalized values were then graphed using boxplots (Figure 11 Aii). Outliers were identified using the Tukey method, and excluded. As demonstrated by Figure 11 Aii, the quantities of individual reference genes, and the geometric mean of the reference genes, remained constant in most patient samples, as demonstrated by normalized values that are close to 1. Taking both the Ct analyses and the relative quantification analyses as a whole, it was confirmed that the amount of *PUM1* and *RPLP0* DNA did not change between non-tumour and tumour tissue, indicating that both genes are suitable reference genes for calculation of *RBM5* gene copy number.

The same two analyses were also completed to confirm the suitability of RPLP0 and ACTB as reference genes for RBM5 mRNA expression analyses. First the distribution of Ct values was compared between non-tumour and tumour RNA from 53 patients. For each sample, we obtained Ct values from six technical replicates. The mean Ct value was calculated for each sample, averaging the six technical replicates and excluding any outliers identified by the Grubb's test. Mean Ct values were then graphed as boxplots (Figure 11 Bi). Outliers were identified using the Tukey method, and removed prior to statistical analyses. As per the Shapiro-Wilk test, the data sets did not follow a normal distribution, and, as such, non-parametric statistical analyses were completed. Since the non-tumour and tumour pairs were obtained from the same patients, a

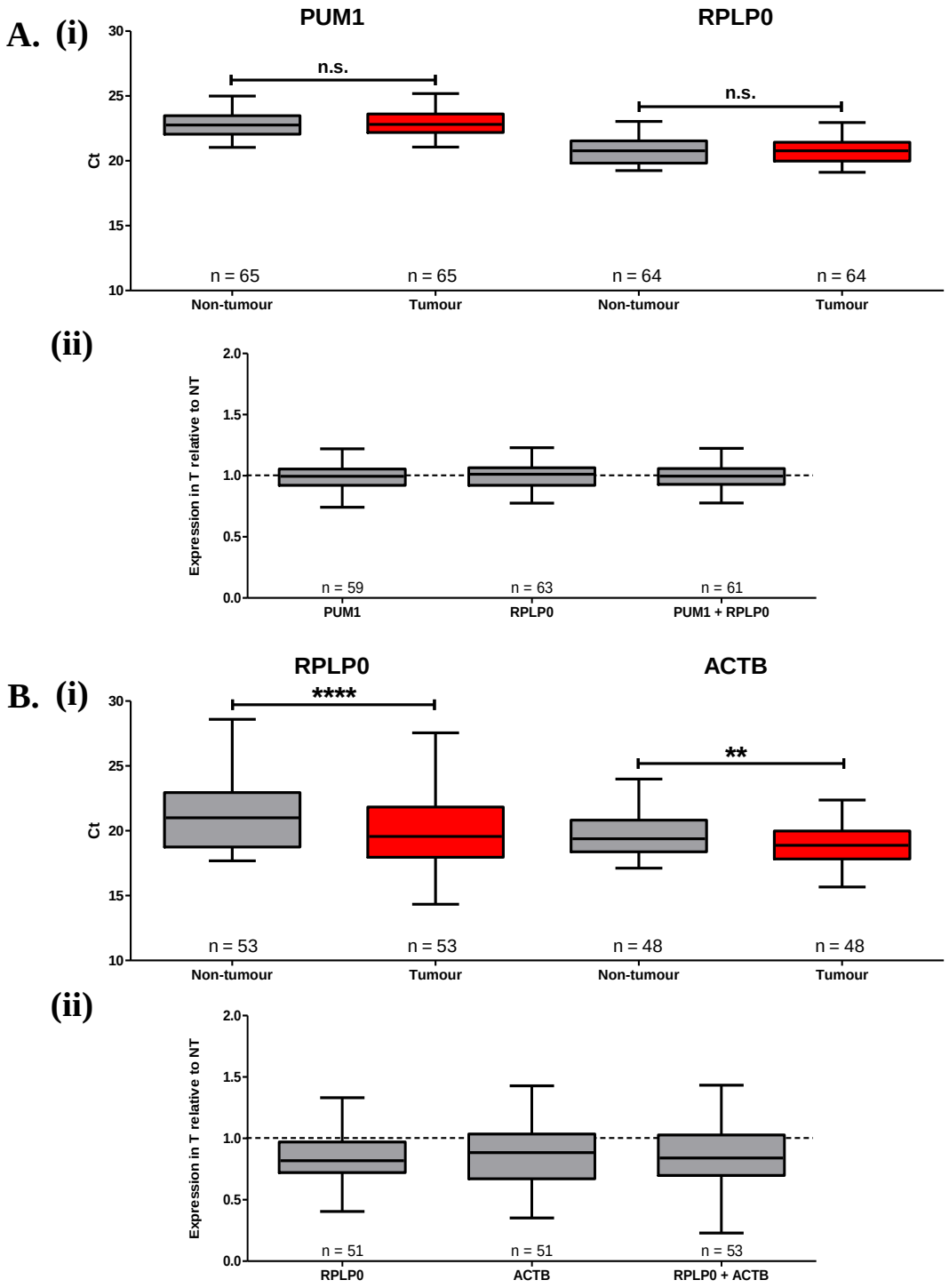


Figure 11 Analysis of changes in reference gene Ct values and the corresponding changes in reference gene quantities. Reference genes were compared between non-tumour and tumour in (A) DNA from 65 patients and (B) RNA from 53 patients. (i) Mean Ct values of each reference gene were compared using the Wilcoxon singed-rank test. ***p<0.001,****p<0.0001. (ii) Individual reference gene quantities, and the geometric mean of reference gene quantities, in tumour tissue relative to non-tumour tissue, was compared to confirm relative values close to 1, representing no changes or minimal changes in DNA or RNA quantities.

paired Wilcoxon signed-rank test was used to identify statistically significant differences between non-tumour and tumour Ct values. Both reference genes resulted in significant changes in Ct values between non-tumour and tumour RNA. For RPLP0, non-tumour and tumour RNA demonstrated a difference in median Ct values of approximately 1.5 cycles, while ACTB displayed a difference in median Ct values of approximately one cycle. In the publication from Gresner *et al.,* where the RPLP0 primers were found, the median Ct values demonstrated a difference of approximately one cycle between non-tumour and tumour, where the tumour had the lower Ct value, similar to what was observed in this study [89]. The publication stated that the difference in median Ct values was not large enough to assume changes in expression [89]. Although the Radonic *et al.* study, from which the ACTB primers were obtained, did not analyze differences in Ct values between non-tumour and tumour RNA, we can assume the same principle applies, in which a difference of one Ct is not enough to permit changes in ACTB expression [93]. As such, further analyses were required to confirm that expression of RPLP0 and ACTB did not change, despite significant changes in Ct values.

Expression of RPLP0 and ACTB was compared in tumour RNA relative to non-tumour RNA in the 53 patients from whom we obtained paired non-tumour and tumour RNA. Using the Ct values from each of the six technical replicates from the patient samples, expression of RPLP0 and ACTB was calculated using the relative standard curve quantification method described above [118,119]. The calculation was performed on each of the six technical replicates, and then the average of the six technical replicates was taken, excluding any outliers identified by the Grubb's test. To examine the suitability of using the two reference genes together, the geometric mean of the expression of RPLP0 and ACTB was also calculated. Relative expression values for RPLP0, ACTB, and the geometric mean of the two reference genes, were calculated by normalizing

expression values in the tumour RNA relative to the non-tumour RNA (Figure 11 Bii). The mean normalized expression values were 0.84, 0.88 and 0.86 for RPLP0, ACTB, and the geometric mean of the two reference genes, respectively. Although there appeared to be slightly less expression of both of the reference genes in tumour RNA, this slight difference was considered acceptable for normalization purposes. In reference to the literature, median fold-changes of less than 1.7 and mean fold-changes of less than 1.8 have been reported as acceptable [89,91]. Since the changes in expression of the reference genes, in this study, are less than the reported values in the literature, it was confirmed that RPLP0 and ACTB were suitable reference genes for the normalization of RBM5 qPCR data.

3.4.2 Reference gene quantities did not change between smoker and never-smoker DNA

As previously mentioned, the act of smoking can inflict DNA damage, causing aberrant lesions, including allelic loss [65,66,71]. Since one of the intentions of this study was to compare *RBM5* gene copy number in smokers to never-smokers, it was necessary to ensure that the reference genes did not change with tobacco smoke exposure, as allelic loss of reference genes would not allow for accurate quantification of *RBM5* gene copy number. To examine this, Ct values of *PUM1* and *RPLP0* were compared between never-smokers and a subset of smokers diagnosed with NSCLC. As the cohort contained only five never-smokers diagnosed with NSCLC, the never-smokers were compared to a small subset of smokers, so that the sample sizes would be similar. The subset of smokers included eight patients diagnosed with NSCLC. These eight patients were chosen specifically because they had the most tobacco smoke exposure, measured in pack-years. It was assumed that the patients with the most tobacco smoke exposure were the patients who

were most likely to exhibit aberrant DNA lesions in reference genes, related to smoking. The number of pack-years in the eight patients ranged between 60 and 130.

For each patient sample (non-tumour and tumour), Ct values from six technical replicates were used. The average Ct value was calculated for each patient sample, excluding any outliers identified by the Grubb's test. Average Ct values from the five never-smokers and the eight smokers were then graphed, using a boxplot (Figure 12). Outliers were identified using the Tukey method, and excluded prior to statistical analyses. Since the data sets did not follow a normal distribution, and samples were not paired, a Mann-Whitney test was used to determine if the difference in mean Ct values between never-smokers and smokers was statistically significant. As demonstrated in Figure 12, mean Ct values of both *PUM1* and *RPLP0* did not change significantly between never-smokers and smokers, in both non-tumour DNA (Figure 12 A) and tumour DNA (Figure 12 B), indicating that the DNA for both reference genes was unlikely to be altered by smoking. Through this reference gene validation, it was confirmed that *PUM1* and *RPLP0* were suitable reference genes for normalizing *RBM5* gene copy number values, in both never-smokers and smokers.

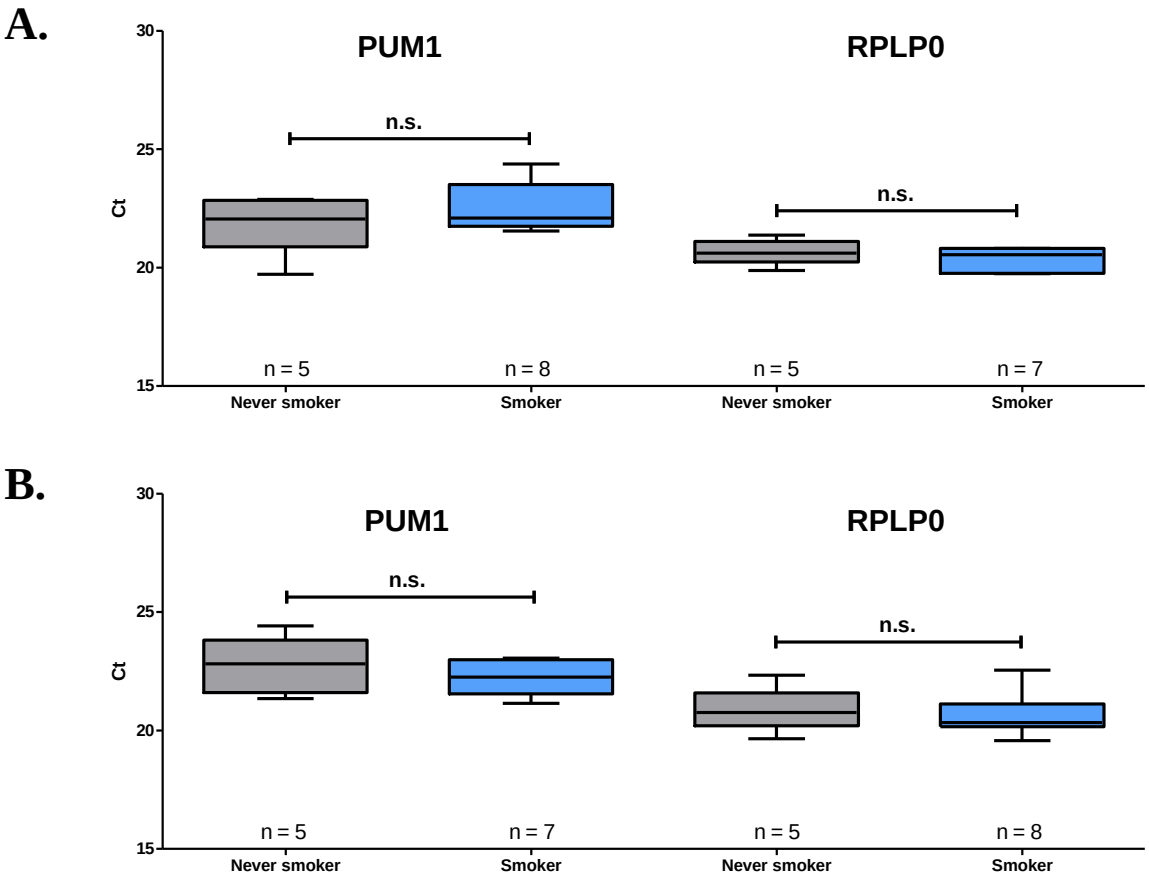


Figure 12 Comparative analyses of reference genes in never-smokers and smokers. Mann-Whitney tests were used to examine the difference in mean Ct values for *PUM1* and *RPLP0* reference genes between never-smokers and smokers in (A) non-tumour and (B) tumour DNA.

Chapter 4

4 *RBM5* Gene Copy Number Analysis

The objective of this study was to determine if the significantly decreased levels of RBM5 expression in the NSCLCs of smokers was related to *RBM5* deletion. To examine this, *RBM5* copy number was quantified in all patient DNA samples from the HSN cohort. Using these data, the number of patients that demonstrated *RBM5* deletions was determined. Next, *RBM5* copy number was compared between non-tumour and NSCLC tumour DNA, to determine if *RBM5* gene copy number was significantly decreased in tumours. Following this, statistical analyses were carried out to determine if *RBM5* gene copy number was related to the level of RBM5 mRNA expression in both non-tumour and tumour tissue. Finally, *RBM5* gene copy number was compared in never-smokers and smokers, to determine if *RBM5* gene copy number was associated with tobacco smoke exposure. After a complete analysis of the HSN cohort, this study was broadened and similar analyses were repeated using SCLC tumour DNA obtained from a collaborator in Slovakia.

4.1 HSN Cohort Study

4.1.1 *RBM5* deletions were detected in a small percentage of patients

Since RBM5 mRNA was significantly downregulated in NSCLC tumours, compared to normal lung tissue, it was expected that at least one copy of *RBM5* would be deleted in tumour DNA, contributing to the observed downregulation at the mRNA level. To test this, *RBM5* copy number was examined in the non-tumour and tumour DNA from 65 patients diagnosed with ADC or SqCC, from whom paired samples were obtained. The quantified *RBM5* copy numbers were averaged from six technical replicates per patient sample, excluding any outliers identified

by the Grubb's test. The average copy numbers in this group ranged between 0.851 and 3.142. Since quantified copy numbers were not integers (i.e. 1, 2 and 3), copy numbers were interpreted according to theoretical boundaries between copy numbers of 1, 2 and 3. These theoretical boundaries were determined using the geometric means of the integer copy number values, as suggested in an article describing accurate copy number profiling using qPCR [100]. The geometric mean of copy numbers 1 and 2 was 1.414, while the geometric mean of copy numbers 2 and 3 was 2.449. Using these theoretical boundaries, copy number results between 1.414 and 2.449 were normal (2n), while results below 1.414 were interpreted as a deletion (1n), and results above 2.449 were interpreted as a duplication (3n+) [100].

Using these integer copy number data, it was determined that more than one-fifth (21.6%) of the patients harboured *RBM5* deletions in tumour DNA (Table 4). Of the 14 patients that exhibited *RBM5* deletions in tumour DNA, seven patients also demonstrated *RBM5* deletions in non-tumour DNA. As some smokers are known to harbour 3p21.3 deletions in normal bronchial epithelium, it was not surprising that all seven of these patients were either current or former smokers [47]. Interestingly, there were eight patients who demonstrated *RBM5* deletions in non-tumour DNA, while tumour DNA was unaltered. All eight of these patients were either current or former smokers; therefore, 3p21.3 deletions in the non-tumour tissue could be anticipated [47]. The fact that the corresponding tumour DNA in these patients did not harbour any *RBM5* deletions was, however, unexpected. It was originally suspected that, perhaps, these patients might have been smokers with minimal pack-years who developed a lung cancer unrelated to smoking, such as an EGFR-positive or ALK-positive adenocarcinoma [81]. This theory, however, was proven to be incorrect. The eight patients were smokers with between 15 and 64 pack-years, smoking an average of approximately one pack per day. Three of the patients were diagnosed

with SqCC, a type of lung cancer that is generally diagnosed in heavy smokers, while the remaining five were diagnosed with ADC, none of which were EGFR-positive or ALK-positive [70,120]. Looking at additional pathological parameters, patterns driving this observation, of decreased copy number in the non-tumour samples only, were not apparent.

Table 4 *RBM5* copy number results for 65 paired samples from patients diagnosed with ADC and SqCC

RBM5 Copy Number	N value (total 65)	Percent (%)	N value Smokers (total 60)	Percent (%)	N value Never Smokers (total 5)	Percent (%)
Deletions (1n)	*22*	*33.8*	*20*	*33.3*	*2*	*40.0*
T only [a]	7	10.8	5	8.3	2	40.0
NT and T	7	10.8	7	11.7	0	0.0
NT only	8	12.3	8	13.3	0	0.0
Duplications (3n+)	*5*	*7.7*	*4*	*6.7*	*1*	*20.0*
T only	3	4.6	2	3.3	1	20.0
NT and T	1	1.5	1	1.7	0	0.0
NT only	1	1.5	1	1.7	0	0.0
Normal (2n)	*38*	*58.5*	*36*	*60.0*	*2*	*40.0*

[a] with the "only" designation, the corresponding non-tumour or tumour remained 2n

In some instances, *RBM5* duplications were observed. Tumours harboured *RBM5* duplications in 6.1% of patients. Three of these patients displayed duplications in the tumour DNA only, while only one patient demonstrated *RBM5* duplications in both non-tumour and tumour DNA. Unexpectedly, one patient displayed a duplication of *RBM5* in the non-tumour DNA only, retaining a normal (2n) copy number in the tumour DNA. Pathological parameters were

examined to determine if the patients harbouring *RBM5* duplications had any similarities that might explain these results, but there were not any outstanding similarities.

In regards to smoking status, 60 patients were either current or former smokers, while only five were never-smokers. In the smokers, 60% had a 2n *RBM5* copy number, 33.3% had deletions, and 6.7% had amplifications. In the never-smokers, three out of five *RBM5* copy number alterations in tumour DNA: two of these patients displayed *RBM5* deletions in tumour DNA, while one patient demonstrated a duplication of *RBM5* in tumour DNA. The most important observation noted amongst the five never-smokers was the fact that none of these patients harboured RBM5 alterations in non-tumour DNA, suggesting that RBM5 deletions and duplications observed in non-tumour DNA are strictly smoking-related events. Unfortunately, the discrepancy in sample sizes made it impossible to attach statistical significance to these finding.

4.1.2 *RBM5* gene copy number was not significantly reduced in NSCLC tumour DNA

As described in Chapter 3, Section 3.1, a significant downregulation of RBM5 mRNA was observed in NSCLC tumours. It was, therefore, anticipated that *RBM5* gene copy number in tumours would also be significantly decreased, demonstrating that RBM5 downregulation in NSCLC was caused by deletion of at least one copy of the *RBM5* gene. To determine significance, DNA data from ADC and SqCC patients were first grouped together, for analysis, so that *RBM5* copy number could be examined in NSCLC as a whole. Data were then analyzed separately, so that any possible differences in *RBM5* copy numbers related to specific NSCLC subtypes could be identified. Normality testing with the Shapiro-Wilk test determined that all data sets did not follow a normal distribution, allowing for non-parametric statistical tests. A 3x2

Fisher's exact test was used to examine significance differences in *RBM5* copy number between non-tumour and tumour DNA in patients diagnosed with NSCLC (Figure 13 A), ADC (Figure 13 B) and SqCC (Figure 13 C). As demonstrated in Figure 13, *RBM5* copy number was not significantly reduced in tumour DNA, compared to non-tumour DNA, in NSCLC, and in both ADC and SqCC subtypes.

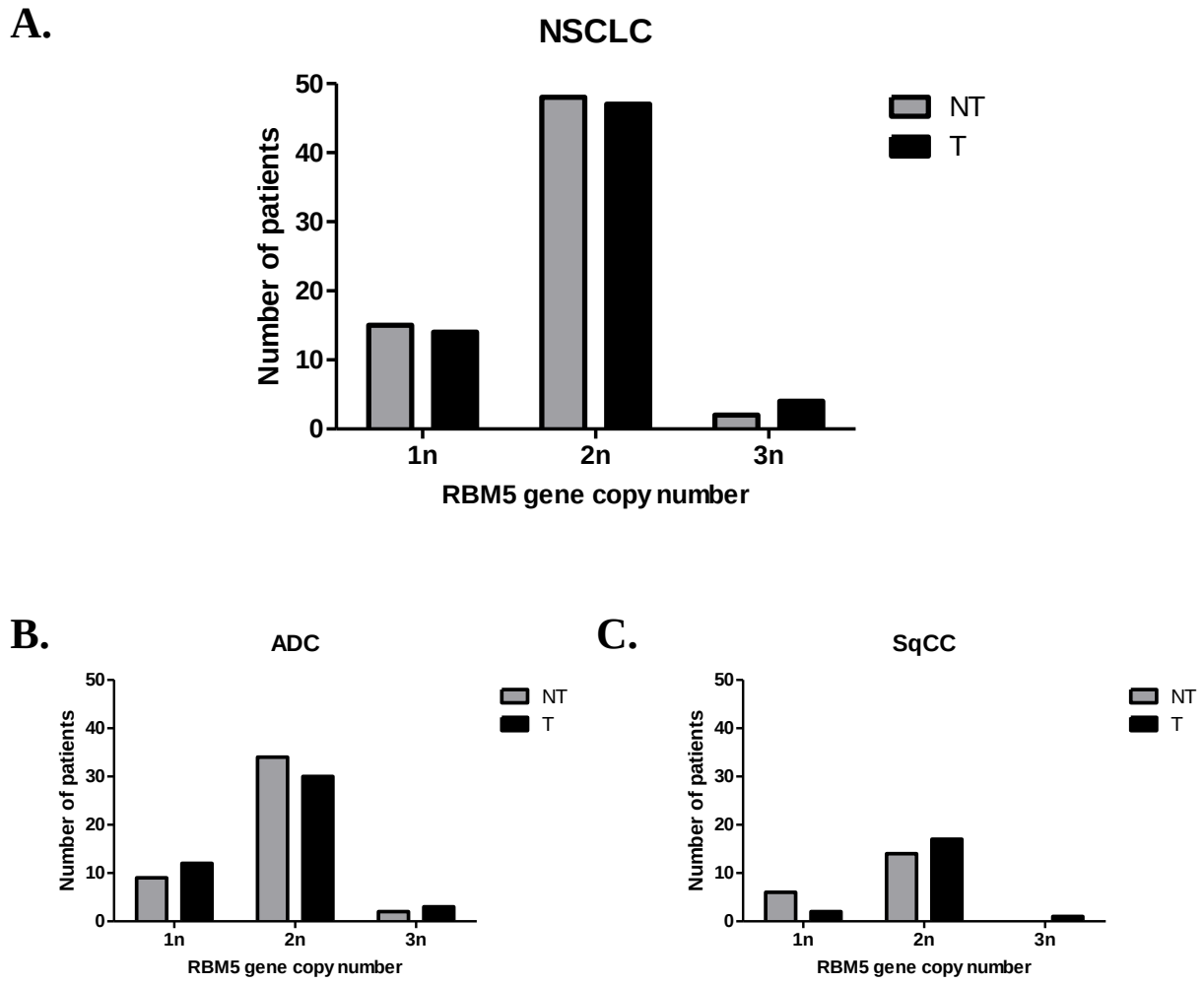


Figure 13 A comparison of *RBM5* copy number in non-tumour and NSCLC tumour DNA. A 3x2 Fisher's exact test was used to compare *RBM5* gene copy number in non-tumour and tumour DNA from patients diagnosed with (A) NSCLC, (B) ADC and (C) SqCC.

4.1.3 *RBM5* gene copy number was related to RBM5 mRNA expression in the non-tumour tissue, but not the tumour tissue

To determine if the deletion events were related to loss of RBM5 mRNA expression, in the same tissue sample, statistical analyses were carried out. For this analysis, 53 paired DNA and RNA samples were utilized from both non-tumour and tumour tissue. The relationship between *RBM5* gene copy number and average normalized RBM5 mRNA expression levels were analysed using a Kruskal-Wallis test, since the RNA data sets did not follow a normal distribution. Boxplots representing RBM5 expression levels, obtained from tissue samples with varying *RBM5* copy numbers, are presented in Figure 14 A. Outliers were identified using the Tukey method, and excluded prior to statistical analyses. In the non-tumour tissue (Figure 14 Ai), a significant difference between *RBM5* copy number and RBM5 mRNA expression levels was observed using the Kruskal-Wallis test. Since the Kruskal-Wallis test could not specify which groups were significantly different, the non-parametric Mann-Whitney test was used to compare each of the groups separately. It was determined that there was a significant difference in RBM5 mRNA expression between patients with *RBM5* copy numbers of 1n and 2n. Non-tumour tissue with a *RBM5* copy number of 1n demonstrated significantly reduced RBM5 mRNA levels, compared to non-tumour tissue with a *RBM5* copy number of 2n. In the tumour tissue (Figure 14 Aii), there were not any significant differences in RBM5 expression between tumours of varying *RBM5* copy numbers. Taking these data as a whole, it appeared as though deletion of *RBM5* influenced loss of RBM5 expression in non-tumour tissue, but deletion of *RBM5* in tumour tissue did not influence loss of RBM5 expression, since expression was reduced in majority of the samples, even in those with normal (2n) copy numbers and duplications.

Next, RBM5 mRNA expression was characterized in the groups harbouring various deletion events, and in the patients with normal (2n) copy numbers, using the categories presented in Table 4. Of the 65 patients that were included in Table 4, we obtained paired RNA samples from 53 of those patients. Non-tumour and tumour RBM5 mRNA expression values in each respective category were presented as boxplots (Figure 14 B). Outliers were identified using the Tukey method, and excluded prior to statistical analyses. The Shapiro-Wilk test determined that the small data sets did not follow a normal distribution, and, as such, non-parametric statistical analyses were used. Since the non-tumour and tumour mRNA samples were from the same patient, a Wilcoxon signed-rank test was used to determine if there was a statistically significant difference in RBM5 expression between non-tumour and tumour pairs in each of the respective categories.

Of the seven patients that exhibited a deletion in the tumour only, we obtained six paired RNA samples. All six tumours showed reduced RBM5 mRNA expression levels that were approximately half of what was observed in the non-tumour tissue (Figure 14 Bi). This result was anticipated, because, based on the hypothesis of this study, the non-tumour tissue that retained a normal (2n) *RBM5* copy number would express normal levels of RBM5, while the tumour tissue that exhibited *RBM5* deletions would express reduced levels of RBM5, reflective of loss of one allele.

Next, of the seven patients that exhibited *RBM5* deletions in both non-tumour and tumour tissue, paired RNA was obtained from six of those patients. As anticipated, RBM5 expression was reduced in both tissue types (Figure 14 Bii), suggesting that *RBM5* deletions influenced RBM5 mRNA expression levels.

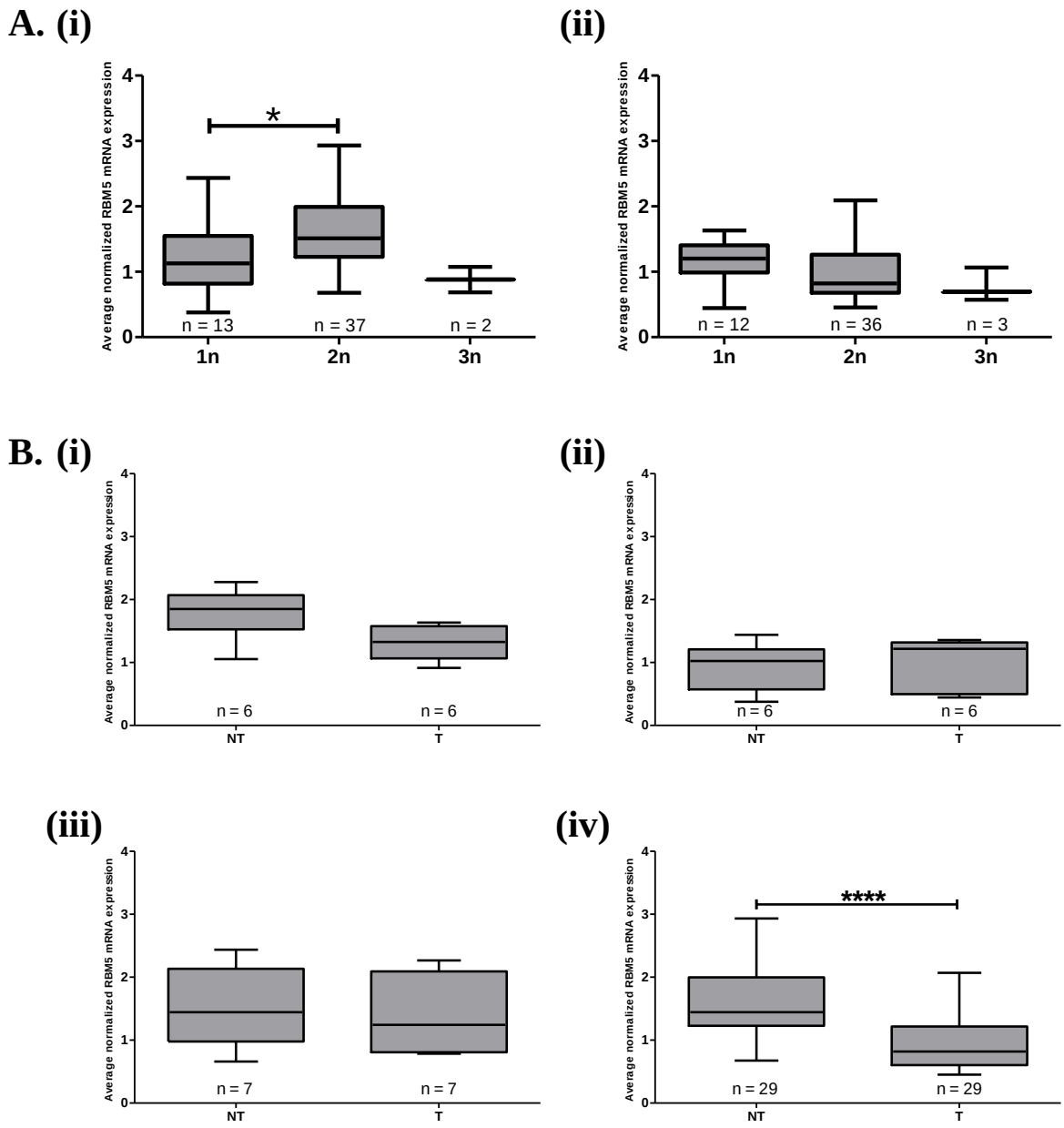


Figure 14 Associations between *RBM5* copy number and mRNA expression. (A) Boxplots were constructed to demonstrate the relationship between *RBM5* copy number and RBM5 mRNA expression in (i) non-tumour and (ii) tumour tissue. The Kruskal-Wallis test was used to evaluate the relationship between *RBM5* copy number and RBM5 mRNA expression values in all data sets. The Mann-Whitney test determined significant differences between each copy number group. *p<0.05. (B) RBM5 mRNA expression levels in patients with an *RBM5* gene copy number of 1 in (i) tumour only, (ii) non-tumour and tumour, and (iii) non-tumour only. (iv) RBM5 mRNA expression in patients with an *RBM5* gene copy number of 2. Statistically significant differences between RBM5 expression values in non-tumour and tumour tissue were evaluated using the Wilcoxon signed-rank test. ****p<0.0001.

From the eight patients that harboured deletions in the non-tumour tissue, but not the tumour tissue, seven paired RNA samples were obtained. As demonstrated by Figure 14 Biii, most of the seven patients demonstrated loss of RBM5 mRNA expression in both non-tumour and tumour tissue. This result suggests that, although deletion of *RBM5* may contribute to decreased RBM5 expression in non-tumour tissue, deletion of *RBM5* was not required for the downregulation of RBM5 in tumour tissue. Supporting this observation, the patients that demonstrated normal (2n) *RBM5* copy numbers, in non-tumour and tumour tissue, demonstrated significantly reduced RBM5 expression levels in tumour tissue, compared to non-tumour tissue (Figure 14 Biv). These results suggested that there was more than one mechanism of RBM5 downregulation in NSCLC tumours.

4.1.4 Smokers have substantially reduced *RBM5* copy numbers in non-tumour DNA, compared to never-smokers

To examine if deletion of *RBM5* was associated with smoking, *RBM5* gene copy number was compared in smokers and never-smokers, in both non-tumour and tumour DNA. For this analysis, 60 paired non-tumour and tumour DNA specimens from smokers, as well as five paired non-tumour and tumour DNA specimens from never-smokers were used. The lack of available DNA specimens from never-smokers caused a substantial discrepancy in sample sizes between smokers and never-smokers, and, therefore, limited statistical power for this analysis. The number of patients in each *RBM5* gene copy number category, from smokers and never-smokers, was graphed as bar graphs for both non-tumour (Figure 15 A) and tumour (Figure 15 B) tissue. The never-smokers did not follow a normal distribution, as per the Shapiro-Wilk test for normality, and so non-parametric statistical analyses were carried out. A 3x2 Fisher's exact test

was used to determine if there was a statistically significant difference in *RBM5* gene copy number between smokers and never-smokers, in both non-tumour and tumour tissue.

In non-tumour DNA, although not significant, only smokers demonstrated *RBM5* loss (Figure 15 A). As anticipated, all five never-smokers retained a normal (2n) *RBM5* gene copy number in non-tumour tissue. The reduction in *RBM5* gene copy number in the non-tumour DNA of smokers was anticipated, as 3p21.3 deletions have been observed in the normal bronchial epithelium of smokers [47]. These results suggest that *RBM5* deletions in non-tumour DNA are dependent on tobacco smoke exposure.

Despite the fact that substantial differences in *RBM5* gene copy number, related to tobacco smoke exposure, were observed in non-tumour DNA, the same effect was not detected in tumour DNA. There were not any significant differences in *RBM5* copy number between smokers and never-smokers (Figure 15 B). *RBM5* copy numbers were more variable in the tumour tissue from never-smokers, compared to the adjacent non-tumour tissue. Two never-smokers displayed two copies of *RBM5*, while the remaining three presented deletions (n=2) or duplications (n=1), in tumour DNA. The fact that *RBM5* deletions were observed in both smokers and never-smokers suggests that loss of one *RBM5* allele in NSCLC may have occurred independent of tobacco smoke exposure.

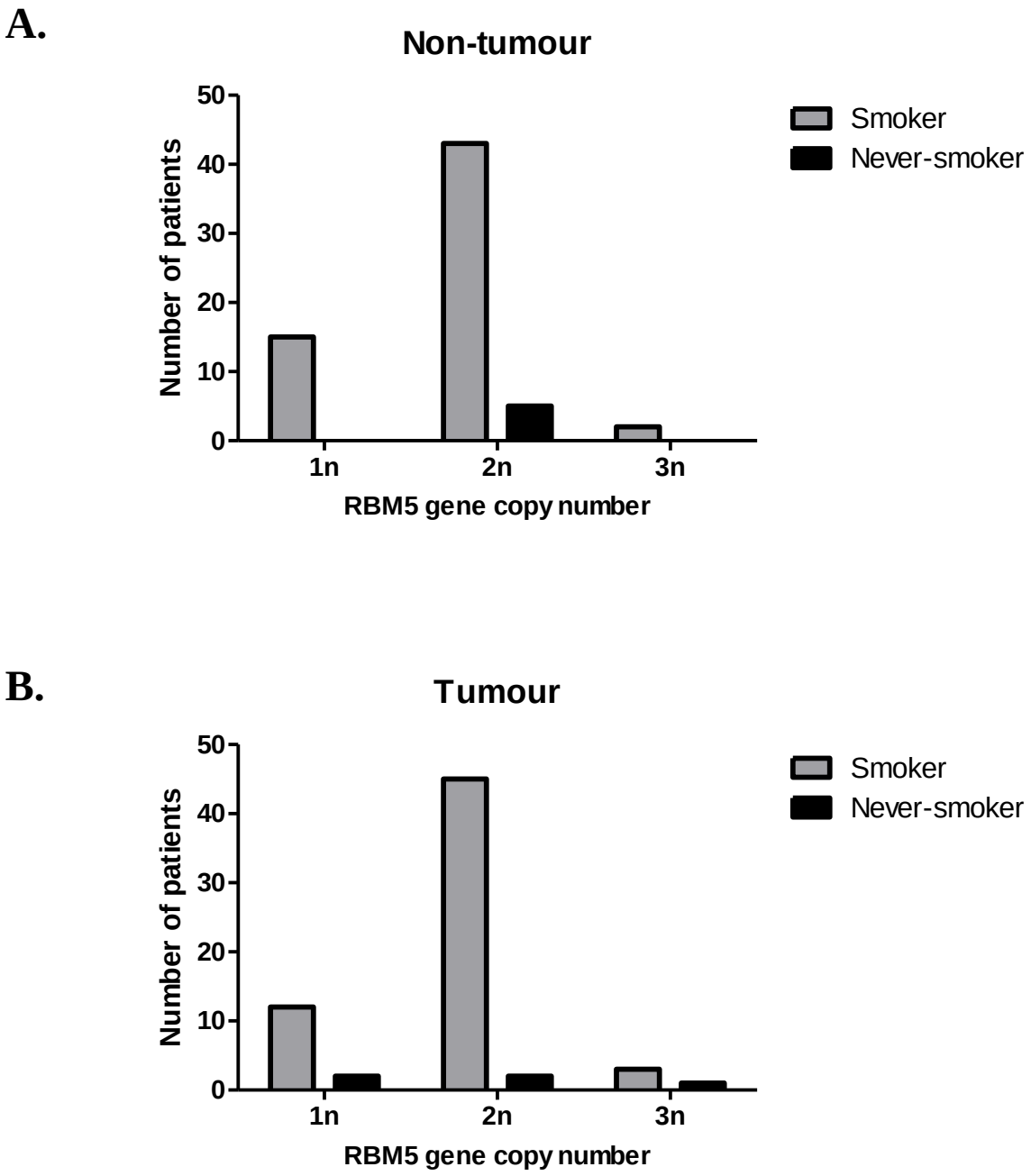


Figure 15 A comparison of *RBM5* copy number in smokers and never-smokers. A 3x2 Fisher's exact test was used to compare *RBM5* gene copy numbers in (A) non-tumour and (B) tumour DNA from smokers and never-smokers.

4.2 Slovakian Cohort Study

4.2.1 *RBM5* deletions were detected in half of the SCLC tumours

Through collaboration, tumour DNA was obtained from a Slovakian cohort, consisting of 44 patients diagnosed with SCLC. It was anticipated that, since LOH at 3p21.3 is more common in SCLC than NSCLC, and *RBM5* was shown to be homozygously deleted in three SCLC cell lines, *RBM5* would be deleted more frequently in the SCLC tumour DNA, compared to the NSCLC tumour DNA [45-47]. The quantified *RBM5* copy numbers in the Slovakian cohort ranged between 0.868 and 3.890. Once again, theoretical boundaries of 1.414 and 2.449 were used, as described previously, to distinguish normal copy number (2n) from deletions (1n) and duplications (3n+).

As anticipated, *RBM5* deletions were more frequent in the SCLC tumours, as 50.0% of patients demonstrated a loss of one *RBM5* allele (Table 5). The frequency of duplications observed in the SCLC cohort was similar to the frequency observed in the HSN NSCLC cohort. Furthermore, a large percentage (40.9%) of SCLC tumours retained both copies of the *RBM5* gene. Paired non-tumour DNA was not available, and so it was impossible to directly compare copy numbers between non-tumour and tumour DNA. As such, it could not be determined if the greater frequency of deletions in SCLC tumours was significant.

Table 5 *RBM5* copy number results for Slovakian SCLC tumour samples

RBM5 Copy Number	N value (total 44)	Percent (%)	N value Smokers (total 34)	Percent (%)	N value Never Smokers (total 8)	Percent (%)
Deletions (1n)	22	50.0	17	50.0	5	62.5
Duplications (3n+)	4	9.1	4	11.8	0	0.0
Normal (2n)	18	40.9	13	38.2	3	37.5

*The smoking status of two patients was not obtained, and, therefore, these patients could not be included in either of the smoking status categories.

4.2.2 *RBM5* deletions in SCLC tumours were not related to smoking status

Since smoking information linked to 42 of the 44 patient samples was received, it was possible to compare *RBM5* copy number in smokers and never-smokers. In the Slovakian cohort, DNA from 34 smokers and 8 never smokers was obtained. The number of patients that fit into specific copy number categories were graphed as bar graphs, for both smokers and never-smokers. Since the data did not follow a normal distribution, as per the Shapiro-Wilk normality test, the non-parametric 3x2 Fisher's exact test was used to determine if there was a significant difference between the number of patients with *RBM5* deletions in the smoking and non-smoking categories. Using this test, it was determined that there was not a statistically significant difference in *RBM5* copy number between smokers and never smokers (Figure 16).

The lack of significant differences in *RBM5* copy number was apparent when examining the percentages of smokers and never-smokers in each of the copy number categories. Similar percentages of smokers and never-smokers retained two copies of *RBM5*, as 38.2% (13/34) of smokers and 37.5% (3/8) of never-smokers displayed normal (2n) *RBM5* copy number values.

With respect to *RBM5* deletions, 50% (17/34) of smokers and 62.5% of never-smokers (5/8) displayed loss of one *RBM5* allele, further demonstrating *RBM5* copy number similarities between smokers and never-smokers. The similar percentages of *RBM5* deletions observed in both smokers and never-smokers indicated that *RBM5* deletions in SCLC tumours were likely unrelated to tobacco smoke exposure.

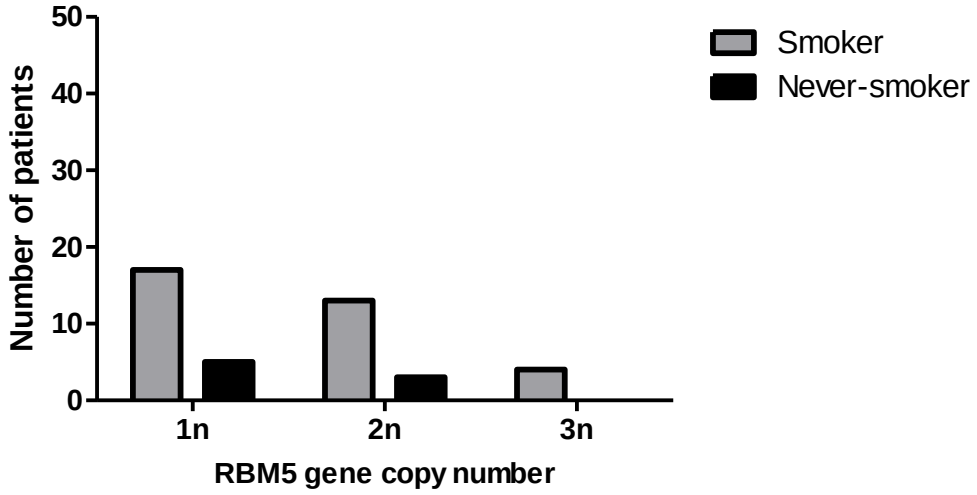


Figure 16 A comparison of *RBM5* copy number in tumour DNA of smokers and never-smokers. A 3x2 Fisher's exact test was used to determine if mean *RBM5* copy numbers in SCLC tumours were significantly different between smokers and never-smokers.

Chapter 5

5 Discussion

5.1 Experimental limitations and possible confounders

5.1.1 Only one region of the *RBM5* gene was used for copy number analyses

The major experimental limitation of this study was the fact that only one region of the *RBM5* gene was analyzed for copy number alterations. In order to be certain that there was loss of one *RBM5* allele, multiple primer pairs mapping to different regions of the *RBM5* gene should have been used. If one whole *RBM5* allele was deleted, then a copy number of one would be anticipated for all primer pairs. Alternatively, if different primer pairs, mapping to different regions of the *RBM5* gene, resulted in different copy numbers, it would suggest that the *RBM5* gene was altered through other mechanisms, in very specific regions. For instance, primers may not hybridize to sequences of the gene that have been mutated, preventing amplification of DNA, even though the gene is intact. Additionally, deletions that do not comprise the whole gene have also been observed in cancer [56]. Therefore, in order to accurately claim *RBM5* allele loss, gene deletion should be confirmed at multiple sites using multiple primer pairs, or the data should be validated by another technique. Other techniques that can be used to analyse deletion of *RBM5* include fluorescence *in situ* hybridization and next-generation sequencing [121,122].

5.1.2 Heterogeneity of crude tissue samples could have skewed *RBM5* copy number results

A second experimental limitation in this study was the fact that we were using crude tissue specimens containing heterogeneous cell types. It is estimated that approximately 50-75% of crude lung tumour biopsies consist of normal cells [123]. These normal cells, which may not contain 3p21.3 deletions, have the ability to skew *RBM5* copy number results, driving the data

towards a normal (2n) copy number. In fact, in this study, approximately half of the *RBM5* copy numbers were calculated as decimal values between 1 and 2. This observation could be indicative of heterogeneous cell populations, in which tumour cells might have harboured *RBM5* deletions, but the presence of normal cells, containing normal (2n) *RBM5* copy numbers, drove the calculated copy number towards a normal (2n) result. If this was the case, some DNA samples may have been improperly categorized as normal [124]. This type of error would have prevented us from accurately quantifying *RBM5* deletions, meaning that the frequency of *RBM5* deletions in tumours may be higher than what was observed in this study. For improved accuracy of copy number calculations, laser capture microdissection should be used to isolate homogenous tumour cells, prior to extracting DNA for copy number analyses [125]. Utilization of this technique would allow for a more accurate representation of the tumour genome, eliminating interference from confounding DNA from heterogeneous cell types.

Instances where *RBM5* deletions were observed in the non-tumour, but not the tumour, can also be explained by tumour heterogeneity. Intratumour heterogeneity has been identified within primary tumours, as populations of distinct subclones that are diverse in DNA content and chromosomal imbalances have been described [126]. In fact, one study estimated that eight or more independently altered clonal patches are present per cm^2 of lung tumour [127]. For the lung tissue used in this study, tiny pieces of non-tumour and tumour tissue, that were much less than 1 cm, constituted the 20 mg that was removed from the tissue specimens for simultaneous DNA, RNA and protein extraction, therefore, only representing a small number of clonal populations. Consequently, the subclonal population from which tumour DNA was obtained may not have been representative of a subclonal population that was derived from the altered non-tumour tissue. In fact, one study could not identify a clonal relationship between non-tumour and tumour

clonal patches present in the lungs of smokers, and the authors attributed this result to heterogeneity [128]. In order to accurately describe *RBM5* gene copy number in the NSCLC tumours used in this study, different sections of the tumour should have been analysed for alterations in copy number, to ensure that multiple subclonal populations were accounted for.

5.1.3 The lack of available DNA samples from never-smokers limited statistical power

One of the obvious experimental limitations in this study was the lack of available DNA from never-smokers. This experimental limitation is constant across many lung cancer studies. The problem with studying never-smokers with lung cancer, is the fact that never-smokers rarely develop lung cancer. As mentioned previously, only 15% of lung cancer diagnoses occur in never-smokers [65]. As a result, it is very difficult to obtain the number of never-smoker samples necessary for statistical power. In this study, of the 103 patients from which tissue samples were obtained, only 80 patients had a diagnosis of either ADC or SqCC. Of these 80 patients, only six were never-smokers, which equates to only 7.5% of the NSCLC group. This means that the HSN cohort contained less never-smokers than expected from a typical population. To complicate matters even further, paired samples were received from five of the six never-smokers. In regards to our Slovakian cohort, eight out of 44 patients were never-smokers, accounting for 18.2% of the cohort. This percentage is actually higher than what would be expected, as, generally, SCLC in never-smokers is rare [75,76].

G*Power 3.1 is a software that can compute the necessary sample size required for statistical power in a given study [129,130]. This software was used to determine the number of smokers and never-smokers that would have been necessary to detect significant differences in the proportions of patients with *RBM5* deletions. Using the G*Power 3.1 software, the appropriate sample size

calculator was chosen to examine proportions of two independent groups (smokers and never-smokers), with the intention of analysing significant differences using the Fisher's exact test. To use this calculator, the proportion of smokers with *RMB5* deletions, the proportion of never-smokers with *RBM5* deletions, and the ratio of never-smokers to smokers, was required. The ideal error probability, power, and the selection of either a one- or two-tailed test were also required. The parameters chosen for the test included an error probability of 0.05, a power of 0.95, and a two-tailed test. Using *RBM5* copy number data from the non-tumours, the proportion of smokers with a *RBM5* deletion was 0.25 (15/60), the proportion of never-smokers with a *RBM5* deletion was 0 (0/5), and the ratio of never-smokers to smokers was 0.083 (5 never-smokers/60 smokers). Using these values with the defined parameters, the G*Power 3.1 calculator stated that at least 216 smokers and 18 never-smokers were necessary to detect a significant difference in the proportions of *RBM5* deletions between never-smokers and smokers. Given these calculations, it is clear that this study did not have the sample size required to accurately detect significant differences in *RBM5* copy number between never-smokers and smokers.

5.1.4 Second-hand smoke exposure and occupational exposures could have confounded *RBM5* copy number data

As mentioned previously, second-hand smoke exposure accounts for approximately 1.6% of lung cancers [67]. It was hypothesized that deletion of *RBM5* was a tobacco smoke-related event, meaning that it was possible that second-hand smoke exposure could contribute to *RBM5* deletions. As all of the never-smokers diagnosed with NSCLC had between 18 and 64 years of second-hand smoke exposure, it is possible that the deletions observed in the tumours of never-smokers were a product of second-hand smoke exposure. In the Slovakian cohort, tumour DNA

was obtained from eight never-smokers, five of which harboured *RBM5* deletions. As we did not receive second-hand smoke exposure information, it is unknown whether *RBM5* deletions were influenced by second-hand tobacco smoke exposure. Ideally, second-hand smoke exposure would be ruled out as a confounder by comparing *RBM5* gene copy number in never-smokers with and without second-hand smoke exposure, if the DNA samples were available.

Another possible confounder that could have contributed to deletion of *RBM5* was occupational exposures to carcinogens, particularly in relation to mining and chromium processing. Mine work has been associated with a number of occupational hazards that could contribute to lung cancer including radon, asbestos, silica dust, diesel fumes and heavy metals [76]. In the HSN cohort, 24 men and one woman documented that they had previously worked in a mine. To determine if mine work was a confounder, *RBM5* gene copy number was compared between the 24 men that worked in a mine to the remaining 31 men who had never worked in a mine. It was determined that there was not a significant difference in the mean *RBM5* gene copy numbers of miners in the HSN cohort (data not shown), proving that mine exposure was not a confounder in the study of *RBM5* copy numbers in the HSN cohort. The SCLC samples that were received from Slovakia were previously part of a study that examined chromium exposure and SCLC incidence. Chromium exposure has previously been shown to inflict DNA damage, contributing to lung cancer development [131]. As information on which of the 44 SCLC patients had previous exposure to chromium was not received, chromium could not be ruled out as a confounder. Had this information been available, *RBM5* gene copy number would have been compared between patients exposed to chromium and patients who were never exposed to chromium.

5.2 Gene deletion was not the primary mechanism of RBM5 downregulation in NSCLC tumours

Upon interpretation of the collective *RBM5* gene copy number data, it was apparent that *RBM5* gene copy number was not significantly decreased in tumours. Due to the low frequency of deletion events that contrasted the high frequency of downregulation observed at the mRNA level, it was concluded that gene deletion was not the primary mechanism by which RBM5 was downregulated in NSCLC tumours. Since it has been demonstrated that TSGs in the common 3p21.3 deletion region are downregulated by more than one mechanism, it is possible that the observed *RBM5* deletions may be responsible for only a small percentage of the observed downregulation in NSCLC tumours.

It was discovered that deletion of one *RBM5* allele was much more frequent in the Sloviakian cohort, as 50.0% of the patients harboured a deletion. Non-tumour DNA from this cohort was not obtained, and non-tumour and tumour DNA could not be compared to determine if *RBM5* gene copy number significantly decreased in this cohort. As the proportion of SCLC tumours exhibiting *RBM5* deletions was more than double what was observed in the NSCLC tumours, it is possible that the greater frequency of deletions could have driven a significant difference, assuming that most non-tumour DNA was normal (2n). Normal *RBM5* gene copy numbers, however, would not be expected in the non-tumour DNA from patients diagnosed with SCLC because allelic loss in discrete regions of 3p, including 3p21.3, occur in the majority of normal bronchial epithelium that accompany SCLC tumours [132]. Furthermore, paired RNA samples were not obtained, preventing confirmation that RBM5 was downregulated in this cohort. Without performing analyses that take into account both non-tumour and tumour DNA and RNA in a

sufficient number of samples, it cannot be determined if deletion of *RBM5* is the mechanism of RBM5 downregulation in SCLC.

Although there was a greater frequency of *RBM5* deletions in the Slovakian SCLC cohort, compared to the HSN NSCLC cohort, it could not be determined whether this phenomenon was a characteristic of SCLC as a disease, or was a characteristic of this Slovakian cohort, specifically. Since LOH at 3p21.3, which may include the *RBM5* gene, occurs more frequently in SCLC (>95%), compared to NSCLC (<70%), it is conceivable that the greater frequency of *RBM5* deletions in SCLC was the result of the disease itself, and not the cohort [45,46,133]. To confirm that the frequency of *RBM5* deletions in SCLC tumours is consistent across multiple cohorts, *RBM5* gene copy number must be examined in a number of SCLC cohorts.

5.3 Smoking may contribute to deletion of *RBM5* in patients diagnosed with NSCLC, especially in non-tumour tissue

Since *RBM5* deletions in non-tumour tissue occurred only in smokers, and these deletions were related to loss of expression, it suggests that *RBM5* deletion is related to smoking, though statistical significance cannot be attributed to this observation since the sample size was too low. This smoking-related lesion may lead to loss of expression that contributes to NSCLC development, as per the sequential theory of lung cancer development [127]. In this cohort, one-quarter (15/60) of smokers diagnosed with NSCLC displayed *RBM5* deletions in non-tumour tissue, indicating that smoking-induced *RBM5* LOH may be an important mechanism of early RBM5 downregulation in a substantial number of smokers that develop NSCLC.

In regards to *RBM5* deletions in NSCLC tumour DNA, there was no relationship with smoking, but a larger never-smoker sample size was necessary to draw accurate conclusions. Since there

were only five samples with variable *RBM5* copy numbers obtained from never-smokers, it could not be determined if the proportions of the different copy number events were representative of a broader non-smoking population. Since 40% (2/5) of never-smokers displayed *RBM5* deletions in tumour DNA, and only 20% (12/60) of smokers displayed *RBM5* deletions in tumour DNA, it was speculated that the data obtained from never-smokers was not typical. Based on the evidence presented in the literature, LOH at 3p21.3, which may include the *RBM5* gene, is related to tobacco smoke exposure [133]. As a consequence, it would not be anticipated that a greater proportion of never-smokers would develop a smoking-related lesion, in comparison to smokers.

Similar to what was observed in the HSN cohort, there were not any significant differences in *RBM5* copy number between never-smokers and smokers in the Slovakian cohort, suggesting that, once again, *RBM5* copy number in tumours was not related to smoking. In this study, sample size was also a factor, as there were only eight never-smokers that were compared to 34 smokers. Smokers and never-smokers demonstrated *RBM5* deletions in 50% (17/34) and 63% (5/8) of tumours, respectively, suggesting that the loss of RBM5 may be a common characteristic of SCLC tumours that is not necessarily related to smoking, and is, instead, related to the disease itself.

5.4 Exploring alternative methods of RBM5 downregulation

Since *RBM5* deletion only accounts for approximately one-fifth of RBM5 downregulation in tumours, other mechanisms of RBM5 downregulation must be explored. Based on what has been documented in the literature, mutations are not a likely mechanism of RBM5 downregulation [13,42,134]. As mentioned previously, promoter hypermethylation cannot be ruled out as a mechanism of RBM5 downregulation. A more thorough examination of DNA methylation of the *RBM5* gene, including the promoter region, should be completed. This examination could be

completed using techniques such as bisulfite sequencing, high performance liquid chromatography – ultraviolet, or liquid chromatography coupled with tandem mass spectrometry [135].

The mechanisms by which other TSGs that map to the common 3p21.3 deletion region are downregulated in lung cancer, might provide insight into how RBM5 is also downregulated in lung cancer. One of the most commonly studied TSGs from the common 3p21.3 deletion region is *RASSF1A* [136]. RASSF1A is downregulated in the majority of lung cancers, like RBM5. Furthermore, similar to *RBM5*, the *RASSF1A* gene is rarely mutated [39,136]. The most common mechanism of RASSF1A downregulation is tumour-acquired promoter hypermethylation, occurring in approximately 90% of SCLCs and 50% of NSCLCs [87]. As a secondary mechanism of RASSF1A downregulation, the loss of one allele can be achieved through LOH at 3p21.3 [137]. One study that examined mechanisms of *RASSF1* inactivation in primary lung tumours found that *RASSF1A* was methylated in 72% of SCLCs and 34% of NSCLCs [138]. In the NSCLC group, 39% of samples demonstrated LOH at 3p21.3 and not methylation [138]. The percentage of tumours showing both methylation and LOH at 3p21.3 was 70% and 27% for SCLC and NSCLC tumours, respectively [138]. The results of the Agathanggelou *et al.* study show that both methylation and LOH at 3p21.3 are important mechanisms of RASSF1A downregulation in lung cancer. Other TSGs that map to the common 3p21.3 deletion region, such as *SEMA3B* and *CACNA2D2*, follow similar mechanisms of inactivation, suggesting that RBM5 downregulation in lung tumours is also influenced by both methylation and LOH [87].

It is also possible that RBM5 is downregulated by deficiencies in transcriptional machinery responsible for regulating RBM5 expression. To date, there have not been any studies completed that have examined regulation at the promoter region, or regulation of transcriptional machinery.

Perhaps, if the *RBM5* gene itself is not altered, there might be alterations in transcription factors or regulatory elements that bind to the *RBM5* gene and regulate its expression. Given the recent studies implicating microRNAs in the downregulation of various TSGs, it is also possible that microRNAs might play a role in the downregulation of RBM5. In fact, one recent study demonstrated that RBM5 is a target of miR-483-5p in prostate cancer [139]. The Yang *et al.* study demonstrated that miR-483-5p expression increased in prostate cancer cell lines and tissue samples, compared to normal cell lines and tissue samples, and negatively regulated RBM5, leading to a loss of RBM5 protein expression [139]. It is, therefore, possible that microRNAs may also play a role in the regulation of RBM5 expression in lung cancer.

5.5 Conclusion

The purpose of this study was to determine if the significantly decreased levels of RBM5 expression in the NSCLCs of smokers was related to *RBM5* deletion. Deletion of *RBM5* in non-tumour DNA was related to loss of RBM5 expression at the mRNA level, but the same result was not observed in the NSCLC tumour DNA. Although *RBM5* deletions were observed in a small percentage of tumours, the deletions did not occur frequently enough to account for the significant downregulation of RBM5 observed at the mRNA level. RBM5 is, therefore, downregulated by more than one mechanism. In the future, *RBM5* deletions should be confirmed in the HSN cohort, and, alternative mechanisms of RBM5 downregulation, such as promoter hypermethylation, should be examined.

In regards to tobacco smoke exposure, *RBM5* deletions in non-tumour tissue were only observed in smokers, suggesting a possible relationship between *RBM5* deletion and smoking. In the tumours, however, *RBM5* deletion was observed in both smokers and never-smokers, suggesting that *RBM5* deletion in NSCLC was not related to smoking. Similar results were observed in

SCLC tumours. This analysis, however, was complicated by the lack of available tissue specimens from never-smokers. In the future, deletion of *RBM5* should be examined in a greater number of never-smokers in order to gain statistically meaningful insight into the relationship between *RBM5* deletion and smoking.

References

1. Lee MH, Schedl T. RNA-binding proteins. WormBook 2006:1-13.

2. Kechavarzi B, Janga SC. Dissecting the expression landscape of RNA-binding proteins in human cancers. Genome Biol 2014;15:R14.

3. Anantharaman V, Koonin EV, Aravind L. Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res 2002;30:1427-64.

4. Castello A, Fischer B, Hentze MW, Preiss T. RNA-binding proteins in Mendelian disease. Trends Genet 2013;29:318-27.

5. Lukong KE, Chang KW, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. Trends Genet 2008;24:416-25.

6. Castello A, Fischer B, Frese CK, et al. Comprehensive Identification of RNA-Binding Domains in Human Cells. Mol Cell 2016;63:696-710.

7. Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett 2008;582:1977-86.

8. Wurth L. Versatility of RNA-Binding Proteins in Cancer. Comp Funct Genomics 2012;2012:178525.

9. Fredericks AM, Cygan KJ, Brown BA, Fairbrother WG. RNA-Binding Proteins: Splicing Factors and Disease. Biomolecules 2015;5:893-909.

10. Sutherland LC, Wang K, Robinson AG. RBM5 as a putative tumor suppressor gene for lung cancer. J Thorac Oncol 2010;5:294-8.

11. Wei MH, Latif F, Bader S, et al. Construction of a 600-kilobase cosmid clone contig and generation of a transcriptional map surrounding the lung cancer tumor suppressor gene (TSG) locus on human chromosome 3p21.3: progress toward the isolation of a lung cancer TSG. Cancer Res 1996;56:1487-92.

12. Oh JJ, Grosshans DR, Wong SG, Slamon DJ. Identification of differentially expressed genes associated with HER- 2/neu overexpression in human breast cancer cells. Nucleic Acids Res 1999;27:4008-17.

13. Timmer T, Terpstra P, van den Berg A, et al. A comparison of genomic structures and expression patterns of two closely related flanking genes in a critical lung cancer region at 3p21.3. Eur J Hum Genet 1999;7:478-86.

14. Edamatsu H, Kaziro Y, Itoh H. LUCA15, a putative tumour suppressor gene encoding an RNA-binding nuclear protein, is down-regulated in ras-transformed Rat-1 cells. Genes Cells 2000;5:849-58.

15. Sutherland LC, Rintala-Maki ND, White RD, Morin CD. RNA binding motif (RBM) proteins: a novel family of apoptosis modulators? J Cell Biochem 2005;94:5-24.

16. Loiselle JJ, Roy JG, Sutherland LC. RBM5 reduces small cell lung cancer growth, increases cisplatin sensitivity and regulates key transformation-associated pathways. Heliyon 2016;2:e00204.

17. Shu Y, Rintala-Maki ND, Wall VE, et al. The apoptosis modulator and tumour suppressor protein RBM5 is a phosphoprotein. Cell Biochem Funct 2007;25:643-53.

18. Bechara EG, Sebestyen E, Bernardis I, Eyras E, Valcarcel J. RBM5, 6, and 10 Differentially Regulate NUMB Alternative Splicing to Control Cancer Cell Proliferation. Mol Cell 2013;52:720-33.

19. Zhang L, Zhang Q, Yang Y, Wu C. The RNA recognition motif domains of RBM5 are required for RNA binding and cancer cell proliferation inhibition. Biochem Biophys Res Commun 2014;444:445-50.

20. Fushimi K, Ray P, Kar A, Wang L, Sutherland LC, Wu JY. Up-regulation of the proapoptotic caspase 2 splicing isoform by a candidate tumor suppressor, RBM5. Proc Natl Acad Sci U S A 2008;105:15708-13.

21. Bonnal S, Martinez C, Forch P, Bachi A, Wilm M, Valcarcel J. RBM5/Luca-15/H37 regulates Fas alternative splice site pairing after exon definition. Mol Cell 2008;32:81-95.

22. Chang DW, Xing Z, Pan Y, et al. c-FLIP(L) is a dual function regulator for caspase-8 activation and CD95-mediated apoptosis. EMBO J 2002;21:3704-14.

23. Rintala-Maki ND, Abrasonis V, Burd M, Sutherland LC. Genetic instability of RBM5/LUCA-15/H37 in MCF-7 breast carcinoma sublines may affect susceptibility to apoptosis. Cell Biochem Funct 2004;22:307-13.

24. Rintala-Maki ND, Sutherland LC. LUCA-15/RBM5, a putative tumour suppressor, enhances multiple receptor-initiated death signals. Apoptosis 2004;9:475-84.

25. Mourtada-Maarabouni M, Williams GT. RBM5/LUCA-15 - Tumour Suppression by Control of Apoptosis and the Cell Cycle? ScientificWorldJournal 2002;2:1885-90.

26. Sutherland LC, Lerman M, Williams GT, Miller BA. LUCA-15 suppresses CD95-mediated apoptosis in Jurkat T cells. Oncogene 2001;20:2713-9.

27. Oh JJ, Razfar A, Delgado I, et al. 3p21.3 tumor suppressor gene H37/Luca15/RBM5 inhibits growth of human lung cancer cells through cell cycle arrest and apoptosis. Cancer Res 2006;66:3419-27.

28. Shao C, Zhao L, Wang K, Xu W, Zhang J, Yang B. The tumor suppressor gene RBM5 inhibits lung adenocarcinoma cell growth and induces apoptosis. World J Surg Oncol 2012;10:160.

29. Oh JJ, West AR, Fishbein MC, Slamon DJ. A candidate tumor suppressor gene, H37, from the human lung cancer tumor suppressor locus 3p21.3. Cancer Res 2002;62:3207-13.

30. Kobayashi T, Ishida J, Musashi M, et al. p53 transactivation is involved in the antiproliferative activity of the putative tumor suppressor RBM5. Int J Cancer 2011;128:304-18.

31. Zhao L, Li R, Shao C, Li P, Liu J, Wang K. 3p21.3 tumor suppressor gene RBM5 inhibits growth of human prostate cancer PC-3 cells through apoptosis. World J Surg Oncol 2012;10:247.

32. Su Z, Yin J, Zhao L, et al. Lentiviral vector-mediated RBM5 overexpression downregulates EGFR expression in human non-small cell lung cancer cells. World J Surg Oncol 2014;12:367.

33. Hao YQ, Su ZZ, Lv XJ, et al. RNA-binding motif protein 5 negatively regulates the activity of Wnt/beta-catenin signaling in cigarette smoke-induced alveolar epithelial injury. Oncol Rep 2015;33:2438-44.

34. Oh JJ, Taschereau EO, Koegel AK, et al. RBM5/H37 tumor suppressor, located at the lung cancer hot spot 3p21.3, alters expression of genes involved in metastasis. Lung Cancer 2010;70:253-62.

35. Li P, Wang K, Zhang J, et al. The 3p21.3 tumor suppressor RBM5 resensitizes cisplatin-resistant human non-small cell lung cancer cells to cisplatin. Cancer Epidemiol 2012;36:481-9.

36. Lv XJ, Du YW, Hao YQ, et al. RNA-binding motif protein 5 inhibits the proliferation of cigarette smoke-transformed BEAS-2B cells through cell cycle arrest and apoptosis. Oncol Rep 2016;35:2315-27.

37. Sutherland LC, Edwards SE, Cable HC, et al. LUCA-15-encoded sequence variants regulate CD95-mediated apoptosis. Oncogene 2000;19:3774-81.

38. Drabkin HA, West JD, Hotfilder M, et al. DEF-3(g16/NY-LU-12), an RNA binding protein from the 3p21.3 homozygous deletion region in SCLC. Oncogene 1999;18:2589-97.

39. Lerman MI, Minna JD. The 630-kb lung cancer homozygous deletion region on human chromosome 3p21.3: identification and evaluation of the resident candidate tumor suppressor genes. The International Lung Cancer Chromosome 3p21.3 Tumor Suppressor Gene Consortium. Cancer Res 2000;60:6116-33.

40. Liang H, Zhang J, Shao C, et al. Differential expression of RBM5, EGFR and KRAS mRNA and protein in non-small cell lung cancer tissues. J Exp Clin Cancer Res 2012;31:36.

41. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet 2003;33:49-54.

42. Oh JJ, Koegel AK, Phan DT, Razfar A, Slamon DJ. The two single nucleotide polymorphisms in the H37/RBM5 tumour suppressor gene at 3p21.3 correlated with different subtypes of non-small cell lung cancers. Lung Cancer 2007;58:7-14.

43. Bird A. The essentials of DNA methylation. Cell 1992;70:5-8.

44. Oh JJ, Boctor BN, Jimenez CA, et al. Promoter methylation study of the H37/RBM5 tumor suppressor gene from the 3p21.3 human lung cancer tumor suppressor locus. Hum Genet 2008;123:55-64.

45. Kok K, Naylor SL, Buys CH. Deletions of the short arm of chromosome 3 in solid tumors and the search for suppressor genes. Adv Cancer Res 1997;71:27-92.

46. Ji L, Minna JD, Roth JA. 3p21.3 tumor suppressor cluster: prospects for translational applications. Future Oncol 2005;1:79-92.

47. Wistuba II, Behrens C, Virmani AK, et al. High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. Cancer Res 2000;60:1949-60.

48. Senchenko VN, Liu J, Loginov W, et al. Discovery of frequent homozygous deletions in chromosome 3p21.3 LUCA and AP20 regions in renal, lung and breast carcinomas. Oncogene 2004;23:5719-28.

49. Sherr CJ. Principles of tumor suppression. Cell 2004;116:235-46.

50. Sun W, Yang J. Functional mechanisms for human tumor suppressors. J Cancer 2010;1:136-40.

51. Macleod K. Tumor suppressor genes. Curr Opin Genet Dev 2000;10:81-93.

52. Motoyama N, Naka K. DNA damage tumor suppressor genes and genomic instability. Curr Opin Genet Dev 2004;14:11-6.

53. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68:820-3.

54. Santarosa M, Ashworth A. Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. Biochim Biophys Acta 2004;1654:105-22.

55. Payne SR, Kemp CJ. Tumor suppressor genetics. Carcinogenesis 2005;26:2031-45.

56. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med 2004;10:789-99.

57. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature 1998;396:643-9.

58. Garinis GA, Patrinos GP, Spanakis NE, Menounos PG. DNA hypermethylation: when tumour suppressor genes go silent. Hum Genet 2002;111:115-27.

59. Kohno T, Yokota J. How many tumor suppressor genes are involved in human lung carcinogenesis? Carcinogenesis 1999;20:1403-10.

60. Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet 2000;25:338-42.

61. Iorio MV, Ferracin M, Liu CG, et al. MicroRNA gene expression deregulation in human breast cancer. Cancer Res 2005;65:7065-70.

62. Statistics CCSsACoC. Canadian Cancer Statistics 20152015 2015.

63. Canada LC. The Faces of Lung Cancer: Fighting Disease, Fighting Disparity 2014. Toronto, Ontario2014 2014.

64. Raponi M, Zhang Y, Yu J, et al. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer Res 2006;66:7466-72.

65. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clin Chest Med 2011;32:605-44.

66. Shields PG. Molecular epidemiology of smoking and lung cancer. Oncogene 2002;21:6870-6.

67. Molina JR YP, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clinic Proceedings 2008;83:584-94.

68. Guo NL, Tosun K, Horn K. Impact and interactions between smoking and traditional prognostic factors in lung cancer progression. Lung Cancer 2009;66:386-92.

69. Kenfield SA, Wei EK, Stampfer MJ, Rosner BA, Colditz GA. Comparison of aspects of smoking among the four histological types of lung cancer. Tob Control 2008;17:198-204.

70. de Perrot M, Licker M, Bouchardy C, Usel M, Robert J, Spiliopoulos A. Sex differences in presentation, management, and prognosis of patients with non-small cell lung carcinoma. J Thorac Cardiovasc Surg 2000;119:21-6.

71. Toyooka S, Maruyama R, Toyooka KO, et al. Smoke exposure, histologic type and geography-related differences in the methylation profiles of non-small cell lung cancer. Int J Cancer 2003;103:153-60.

72. Huang T, Chen X, Hong Q, et al. Meta-analyses of gene methylation and smoking behavior in non-small cell lung cancer patients. Sci Rep 2015;5:8897.

73. Small Cell Lung Cancer. 2015. (Accessed May 15, 2015, at http://www.cancer.ca/en/cancer-information/cancer-type/lung/lung-cancer/small-cell-lung-cancer/?region=on.)

74. Govindan R, Page N, Morgensztern D, et al. Changing epidemiology of small-cell lung cancer in the United States over the last 30 years: analysis of the surveillance, epidemiologic, and end results database. J Clin Oncol 2006;24:4539-44.

75. Jackman DM, Johnson BE. Small-cell lung cancer. Lancet 2005;366:1385-96.

76. Travis W.D. BE, Muller-Hermelink H.K., Harris C.C. World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart. Lyon, France: IARC Press; 2004.

77. Peifer M, Fernandez-Cuesta L, Sos ML, et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. Nat Genet 2012;44:1104-10.

78. Imielinski M, Berger AH, Hammerman PS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 2012;150:1107-20.

79. Petersen I, Bujard M, Petersen S, et al. Patterns of chromosomal imbalances in adenocarcinoma and squamous cell carcinoma of the lung. Cancer Res 1997;57:2331-5.

80. Khuder SA, Mutgi AB. Effect of smoking cessation on major histologic types of lung cancer. Chest 2001;120:1577-83.

81. TCGA. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511:543-50.

82. TCGA. Comprehensive genomic characterization of squamous cell lung cancers. Nature 2012;489:519-25.

83. Simon GR, Wagner H. Small cell lung cancer. Chest 2003;123:259S-71S.

84. Janne PA, Freidlin B, Saxman S, et al. Twenty-five years of clinical research for patients with limited-stage small cell lung carcinoma in North America. Cancer 2002;95:1528-38.

85. Chute JP, Chen T, Feigal E, Simon R, Johnson BE. Twenty years of phase III trials for patients with extensive-stage small-cell lung cancer: perceptible progress. J Clin Oncol 1999;17:1794-801.

86. Barletta JA, Yeap BY, Chirieac LR. Prognostic significance of grading in lung adenocarcinoma. Cancer 2010;116:659-69.

87. Zabarovsky ER, Lerman MI, Minna JD. Tumor suppressor genes on chromosome 3p involved in the pathogenesis of lung and other cancers. Oncogene 2002;21:6915-35.

88. Smit EF, de Vries EG, Timmer-Bosscha H, et al. In vitro response of human small-cell lung-cancer cell lines to chemotherapeutic drugs; no correlation with clinical data. Int J Cancer 1992;51:72-8.

89. Gresner P, Gromadzinska J, Wasowicz W. Reference genes for gene expression studies on non-small cell lung cancer. Acta Biochim Pol 2009;56:307-16.

90. Liu DW, Chen ST, Liu HP. Choice of endogenous control for gene expression in nonsmall cell lung cancer. Eur Respir J 2005;26:1002-8.

91. Saviozzi S, Cordero F, Lo Iacono M, Novello S, Scagliotti GV, Calogero RA. Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. BMC Cancer 2006;6:200.

92. Soes S, Sorensen BS, Alsner J, et al. Identification of accurate reference genes for RT-qPCR analysis of formalin-fixed paraffin-embedded tissue from primary non-small cell lung cancers and brain and lymph node metastases. Lung Cancer 2013;81:180-6.

93. Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun 2004;313:856-62.

94. Wang G, Xu Z, Wang R, et al. Genes associated with MUC5AC expression in small airway epithelium of human smokers and non-smokers. BMC Med Genomics 2012;5:21.

95. Graff JW, Powers LS, Dickson AM, et al. Cigarette smoking decreases global microRNA expression in human alveolar macrophages. PLoS One 2012;7:e44066.

96. Girard L, Zochbauer-Muller S, Virmani AK, Gazdar AF, Minna JD. Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering. Cancer Res 2000;60:4894-906.

97. Bayer TS, Booth LN, Knudsen SM, Ellington AD. Arginine-rich motifs present multiple interfaces for specific binding by RNA. RNA 2005;11:1848-57.

98. Heid CA, Stevens J, Livak KJ, Williams PM. Real time quantitative PCR. Genome Res 1996;6:986-94.

99. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 2008;3:1101-8.

100. D'haene B, Vandesompele J, Hellemans J. Accurate and objective copy number profiling using real-time quantitative PCR. Methods 2010;50:262-70.

101. Yun JJ, Heisler LE, Hwang, II, et al. Genomic DNA functions as a universal external standard in quantitative real-time PCR. Nucleic Acids Res 2006;34:e85.

102. Adikaram KK, Hussein MA, Effenberger M, Becker T. Outlier detection method in linear regression based on sum of arithmetic progression. ScientificWorldJournal 2014;2014:821623.

103. Loiselle JJ, Sutherland LC. Differential downregulation of Rbm5 and Rbm10 during skeletal and cardiac differentiation. In Vitro Cell Dev Biol Anim 2014;50:331-9.

104. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 2001;29:e45.

105. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol 2007;8:R19.

106. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (1). Restor Dent Endod 2012;37:245-8.

107. Eggers KM, Lindahl B. Application of Cardiac Troponin in Cardiovascular Diseases Other Than Acute Coronary Syndrome. Clin Chem 2017;63:223-35.

108. Hart A. Mann-Whitney test is not just a test of medians: differences in spread can be important. BMJ 2001;323:391-3.

109. Rosner B, Glynn RJ, Lee ML. The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics 2006;62:185-92.

110. Raymond M, Rousset F. An Exact Test for Population Differentiation. Evolution 1995;49:1280-3.

111. Chan Y, Walmsley RP. Learning and understanding the Kruskal-Wallis one-way analysis-of-variance-by-ranks test for differences among three or more independent groups. Phys Ther 1997;77:1755-62.

112. Bustin SA, Benes V, Garson JA, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem 2009;55:611-22.

113. de Kok JB, Roelofs RW, Giesendorf BA, et al. Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. Lab Invest 2005;85:154-9.

114. Pfaffl MW. Relative Quantification. In: Dorak T, ed. Real-time PCR. New York: Taylor and Francis Group; 2006:63-82.

115. Souaze F, Ntodou-Thome A, Tran CY, Rostene W, Forgez P. Quantitative RT-PCR: limits and accuracy. Biotechniques 1996;21:280-5.

116. Ruijter JM, Ramakers C, Hoogaars WM, et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res 2009;37:e45.

117. Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M. A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. Methods 2010;50:S1-5.

118. Masilamani TJ, Loiselle JJ, Sutherland LC. Assessment of reference genes for real-time quantitative PCR gene expression normalization during C2C12 and H9c2 skeletal muscle differentiation. Mol Biotechnol 2014;56:329-39.

119. Larionov A, Krause A, Miller W. A standard curve based method for relative real time PCR data processing. BMC Bioinformatics 2005;6:62.

120. Weston A, Willey JC, Modali R, et al. Differential DNA sequence deletions from chromosomes 3, 11, 13, and 17 in squamous-cell carcinoma, large-cell carcinoma, and adenocarcinoma of the human lung. Proc Natl Acad Sci U S A 1989;86:5099-103.

121. Alkan C, Kidd JM, Marques-Bonet T, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet 2009;41:1061-7.

122. De Preter K, Speleman F, Combaret V, et al. Quantification of MYCN, DDX1, and NAG gene copy number in neuroblastoma using a real-time quantitative PCR assay. Mod Pathol 2002;15:159-66.

123. Coghlin CL, Smith LJ, Bakar S, et al. Quantitative analysis of tumor in bronchial biopsy specimens. J Thorac Oncol 2010;5:448-52.

124. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. Nat Genet 2007;39:S37-42.

125. Fend F, Raffeld M. Laser capture microdissection in pathology. J Clin Pathol 2000;53:666-72.

126. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. Nature 2013;501:355-64.

127. Wistuba II, Mao L, Gazdar AF. Smoking molecular damage in bronchial epithelium. Oncogene 2002;21:7298-306.

128. Park IW, Wistuba, II, Maitra A, et al. Multiple clonal abnormalities in the bronchial epithelium of patients with lung cancer. J Natl Cancer Inst 1999;91:1863-8.

129. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39:175-91.

130. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009;41:1149-60.

131. Halasova E, Adamkov M, Matakova T, Kavcova E, Poliacek I, Singliar A. Lung cancer incidence and survival in chromium exposed individuals with respect to expression of anti-

apoptotic protein survivin and tumor suppressor P53 protein. Eur J Med Res 2010;15 Suppl 2:55-9.

132. Wistuba, II, Berry J, Behrens C, et al. Molecular changes in the bronchial epithelium of patients with small cell lung cancer. Clin Cancer Res 2000;6:2604-10.

133. Wistuba, II, Behrens C, Virmani AK, et al. High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. Cancer Res 2000;60:1949-60.

134. Sebestyen E, Singh B, Minana B, et al. Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. Genome Res 2016;26:732-44.

135. Kurdyukov S, Bullock M. DNA Methylation Analysis: Choosing the Right Method. Biology (Basel) 2016;5.

136. Sato M, Shames DS, Gazdar AF, Minna JD. A translational view of the molecular pathogenesis of lung cancer. J Thorac Oncol 2007;2:327-43.

137. Hesson LB, Cooper WN, Latif F. Evaluation of the 3p21.3 tumour-suppressor gene cluster. Oncogene 2007;26:7283-301.

138. Agathanggelou A, Honorio S, Macartney DP, et al. Methylation associated inactivation of RASSF1A from region 3p21.3 in lung, breast and ovarian tumours. Oncogene 2001;20:1509-18.

139. Yang ZG, Ma XD, He ZH, Guo YX. miR-483-5p promotes prostate cancer cell proliferation and invasion by targeting RBM5. Int Braz J Urol 2017;43.

140. Rintala-Maki ND, Goard CA, Langdon CE, et al. Expression of RBM5-related factors in primary breast tissue. J Cell Biochem 2007;100:1440-58.

141. Sutherland LC, Anderson CL, Williams GT. Zinc has no effect on IL-3-mediated apoptosis of BAF-3 cells but enhances CD95-mediated apoptosis of jurkat cells. J Immunol Methods 2000;234:43-50.

142. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. Nat Biotechnol 2003;21:255-61.

143. Hornbeck PV, Kornhauser JM, Tkachev S, et al. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res 2012;40:D261-70.

144. Adams DJ, van der WL, Mayeda A, Stamm S, Morris BJ, Rasko JE. ZNF265--a novel spliceosomal protein able to induce alternative splicing. J Cell Biol 2001;154:25-32.

145. Hilgarth RS, Sarge KD. Detection of sumoylated proteins. Methods Mol Biol 2005;301:329-38.

146. Tatham MH, Rodriguez MS, Xirodimas DP, Hay RT. Detection of protein SUMOylation in vivo. Nat Protoc 2009;4:1363-71.

Appendix

A.1 Introduction

It has previously been documented that RBM5 protein is downregulated in approximately 70% of NSCLCs [27,40]. Since protein samples were obtained from patients in the HSN cohort, RBM5 protein expression was examined for further confirmation that RBM5 was downregulated in NSCLC tumours. It was previously confirmed that RBM5 mRNA expression significantly decreased in tumours from the HSN cohort, and so it was anticipated that the protein expression would also be significantly decreased in tumours. Concerning the protein, the primary objective was to confirm that RBM5 protein was downregulated in NSCLC tumours. This was achieved by Western blot analysis of 132 patient samples, consisting of paired non-tumour and tumour protein from 60 patients diagnosed with ADC or SqCC, and six patients diagnosed with mixed NSCLC. Statistical analyses were carried out on the patients diagnosed with ADC and SqCC only.

A.2 Materials and Methods

A.2.1 Protein samples

Protein was extracted from fresh frozen tissue as described previously, using the Qiagen Allprep DNA/RNA/Protein Mini kit (Qiagen, Mississauga, ON, Canada). Protein was quantified using the Bio-Rad DC Protein Assay Kit (Bio-Rad Laboratories, Hercules, CA, U.S.A.). A total of 132 samples consisting of paired non-tumour and tumour protein from 66 patients were used in this study. Of these 66 patients, 42 patients were diagnosed with ADC, 18 were diagnosed with SqCC, and six were diagnosed with mixed NSCLC.

A.2.2 Western Blot Analysis

Protein samples were prepared by diluting 50 μg of protein in a 1:2 ratio with Western loading buffer consisting of 0.06 M Tris-HCL pH 6.8, 25% glycerol, 2% SDS, 0.01% bromophenol blue, and 5% β-mercaptoethanol. Diluted samples were then heated for 5 minutes at 95°C. The protein samples and a Precision Plus Protein Ladder (Bio-Rad) were then loaded onto a polyacrylamide gel containing a 4% stacking gel (0.125 M Tris, 0.1% SDS, 4% acrylamide, 0.05% APS and TEMED) and a 7% resolving gel (0.375 M Tris, 0.4% SDS, 7% acrylamide, 0.1% APS and TEMED). Samples were electrophoresed in running buffer (0.025 M Tris, 0.19 M glycine and 0.1% SDS) at 100 V for approximately 1.5 hours, until the 37 kDa marker was at the bottom of the resolving gel.

Once electrophoresis was complete, the protein was then transferred onto a PVDF membrane (GE Health Care, Mississauga, ON, Canada) using a wet transfer apparatus containing transfer buffer (25 nM Tris, 192 mM glycine and 20% methanol) for 75 minutes at 350 mA. The transfer of the protein to the PVDF membrane was confirmed with Ponceau S (Bioshop). The PVDF membrane was subsequently rinsed twice with TBS-T (20 mM Tris, 0.5 M NaCl and 0.1% Tween-20), and then washed three times in TBS-T on a shaker at room temperature for 15 minutes, 5 minutes and 5 minutes, respectively. The membrane was then blocked in 5% non-fat dry milk (Carnation Evaporated Milk, Smuckers, Markham, ON, Canada) diluted with TBS-T on a shaker for 45 minutes at room temperature. After blocking, the membrane was rinsed twice with TBS-T, and then washed with TBS-T three times on a shaker at room temperature for 15 minutes, 5 minutes and 5 minutes, respectively. The membrane was then placed in plastic Kapak bag (Ampac, Cincinnati, OH, U.S.A.) containing 3% non-fat dry milk diluted in TBS-T with primary antibody. The antibodies used included rabbit anti-human-RBM5 LUCA-15 UK

(1:3500, non-commercially available) and rabbit anti-human β-Actin (ACTB) (1:10,000, NB600-532, Novus Biologicals, Oakville, ON, Canada). The sealed Kapak bag containing the membrane and antibody was then placed at 4°C on a rotator overnight.

Following the overnight incubation, the membrane was rinsed twice with TBS-T and washed three times in TBS-T on the shaker at room temperature for 15 minutes, 5 minutes and 5 minutes, respectively. The membrane was then placed in a 3% non-fat dry milk solution diluted with TBS-T, containing a goat anti-rabbit HRP-conjugated secondary antibody (1:10,000, sc-2004, Santa Cruz Biotechnologies Inc., Dallas, TX, U.S.A.). This solution was placed on a shaker for 1 hour at room temperature. Detection of antibodies was achieved using Amersham ECL Western Blotting Detection Reagents (GE Healthcare, Mississauga, ON, Canada) and Amersham Hyperfilm ECL (GE Healthcare). Quantification of RBM5 expression was performed using AlphaEase FC software (Alpha Innotech). All RBM5 expression values were normalized to ACTB. Duplicate Western blots were performed for each patient sample.

A.2.3 Peptide Block

A peptide block was used to examine LUCA-15 UK antibody specificity. Three peptides were used in the initial peptide block experiments: UK peptide, LSLN peptide and LSLC peptide. The location of the peptides on the full-length RBM5 protein is displayed in Figure A1. Peptides were diluted to 0.5 μg/μl in 1X PBS (Life Technologies, Burlington, ON, Canada) and then stored at -20°C. For the peptide block, the anti-human RBM5 LUCA-15 UK antibody was combined with 30 μg of peptide in a microfuge tube containing 1X PBS (Life Technologies), with a total volume of 500 μl. The microfuge tubes containing the antibody and the peptide were incubated at 4°C overnight, on a rotator. Prior to incubating the PVDF membranes in primary antibody, the antibody and peptide solution was combined with TBS-T, making a total volume of

6 ml, and mixed with 0.18 g non-fat dry milk powder to achieve a 3% milk solution. For the peptide block, 30 µg of peptide were used because concentrations of 5 µg/ml (in the final milk solution) were recommended by antibody manufacturers in online peptide blocking protocols (i.e. AnaSpec and Neuromics). PVDF membranes were then placed in a Kapak bag and combined with antibody, as described in A.2.2. After performing the peptide block experiment, it was noticed that the LSLC peptide, while not specifically blocking the LUCA-15 UK antibody, reduced the background associated with the LUCA-15 UK antibody. Because of this, a peptide block was performed prior to using the LUCA-15 UK antibody when analysing RBM5 protein expression in patient samples, using the method described above.

Figure A1 Location of RBM5 peptides on full-length RBM5 protein. This schematic diagram was constructed to demonstrate the regions of full-length RBM5 protein from which UK peptide, LSLN peptide and LSLC peptide originated.

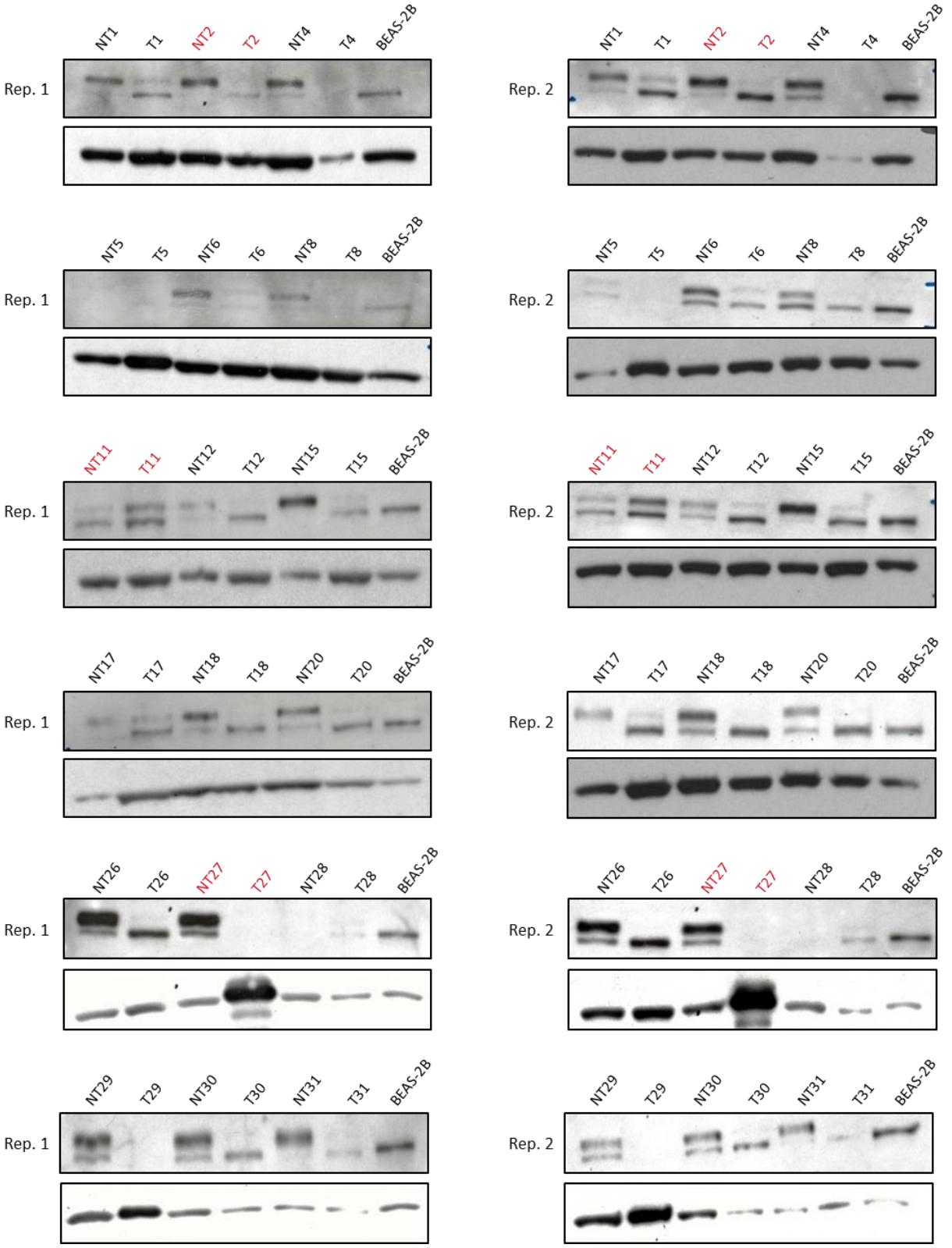
A.3 Results

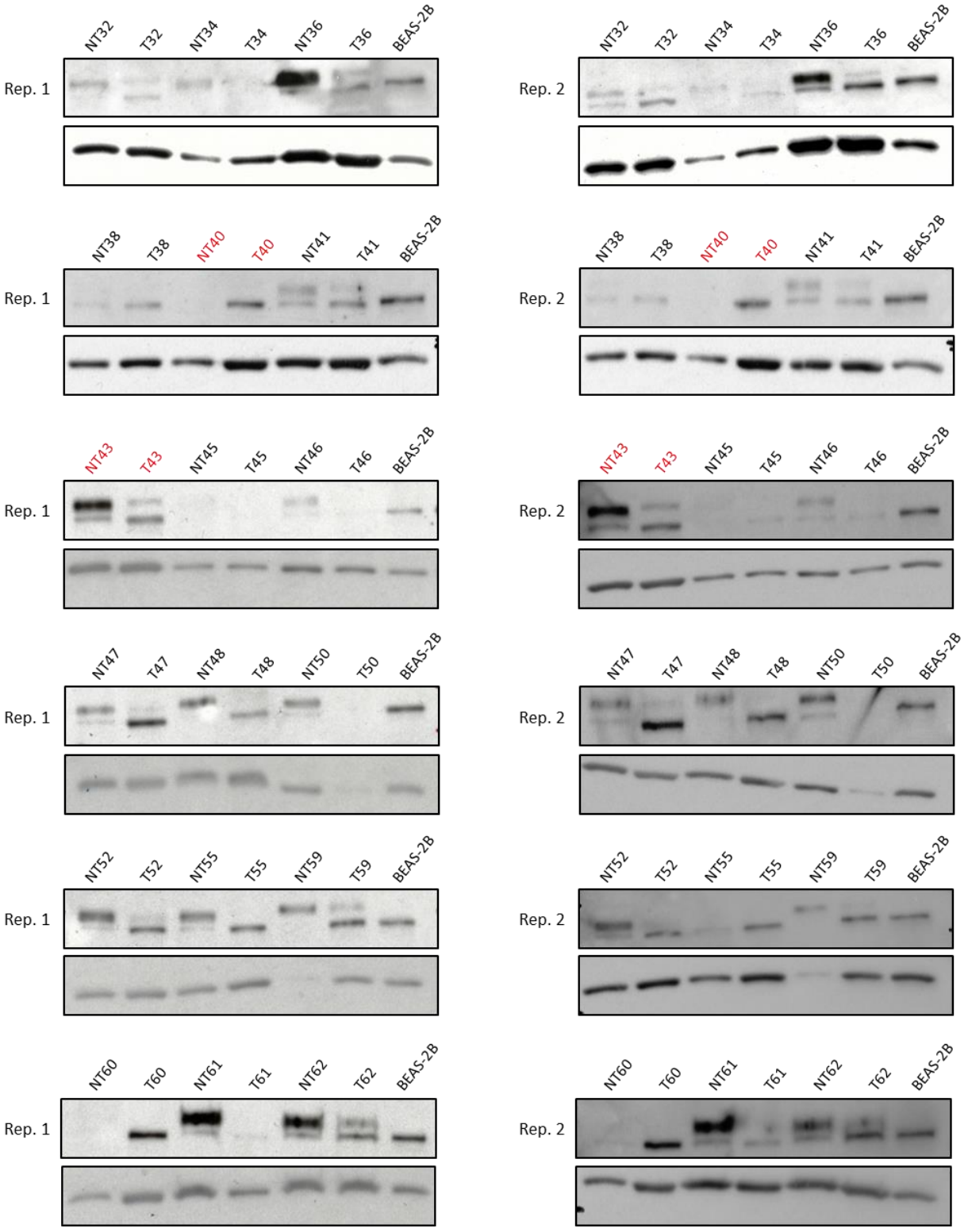
A.3.1 Two RBM5 bands with reciprocal expression were observed in non-tumour and tumour protein

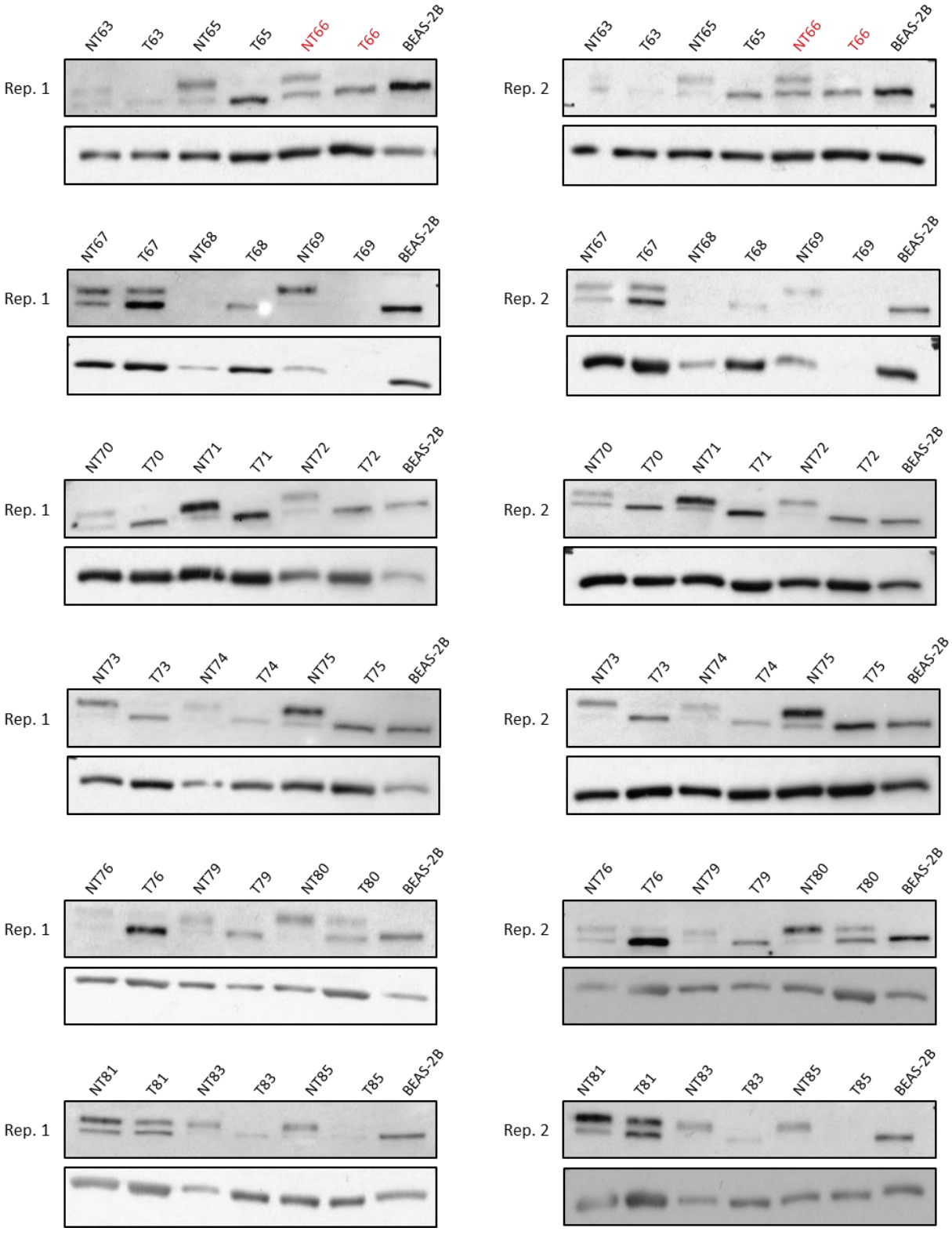
In initial Western blots with patient protein samples, one RBM5 protein band around 113 kDa was anticipated, as only one RBM5 band was previously observed in various cell lines and in primary breast samples using the LUCA-15 UK antibody [16,37,140,141]. Unexpectedly, two prominent bands at approximately 113 kDa (RBM5) and 124 kDa, were observed. Expression was reciprocal in nature, when comparing non-tumour and tumour protein (Figure A2).

Although the LUCA-15 UK antibody has been shown to detect a number of non-specific bands, a non-specific band at 124 kDa has not been observed in various cell lines or in breast tissue [16,37,140]. Interestingly, expression of the 113 kDa and 124 kDa bands were reciprocal in nature when comparing non-tumour and tumour tissue. In non-tumour tissue, the upper band was more highly expressed than the lower band, while, in tumour tissue, the lower band was more highly expressed than the upper band. The reciprocity of the two bands, and the fact that a non-specific band at 124 kDa has never been observed, suggested that the 124 kDa band was a specific post-translationally modified version of RBM5 protein that was specific to lung tissue.

In order to determine if the upper band was specific, a peptide block was performed using three distinct peptides, each specific to a different segment of the full-length RBM5 protein. The UK peptide was the peptide that the LUCA-15 UK antibody was raised against, and consisted of the first 15 amino acids of the full-length RBM5 protein, while the LSLN and LSLC peptides were not specific targets of the LUCA-15 UK antibody, as they consisted of amino acids 67-83 and 735-748, respectively. It was anticipated that blocking the LUCA-15 UK antibody with the UK peptide would result in elimination of specific RBM5 bands on the membrane, as the UK peptide


NT1
T1
NT2
T2
NT4
T4
BEAS-2B
Rep. 1
Rep. 2
NT5
T5
NT6
T6
NT8
T8
NT11
T11
NT12
T12
NT15
T15
NT17
T17
NT18
T18
NT20
T20
NT26
T26
NT27
T27
NT28
T28
NT29
T29
NT30
T30
NT31
T31


NT32
T32
NT34
T34
NT36
T36
BEAS-2B
Rep. 1
Rep. 2
NT38
T38
NT40
T40
NT41
T41
NT43
T43
NT45
T45
NT46
T46
NT47
T47
NT48
T48
NT50
T50
NT52
T52
NT55
T55
NT59
T59
NT60
T60
NT61
T61
NT62
T62


Rep. 1
NT63
T63
NT65
T65
NT66
T66
BEAS-2B
Rep. 2
NT63
T63
NT65
T65
NT66
T66
BEAS-2B
Rep. 1
NT67
T67
NT68
T68
NT69
T69
BEAS-2B
Rep. 2
NT67
T67
NT68
T68
NT69
T69
BEAS-2B
Rep. 1
NT70
T70
NT71
T71
NT72
T72
BEAS-2B
Rep. 2
NT70
T70
NT71
T71
NT72
T72
BEAS-2B
Rep. 1
NT73
T73
NT74
T74
NT75
T75
BEAS-2B
Rep. 2
NT73
T73
NT74
T74
NT75
T75
BEAS-2B
Rep. 1
NT76
T76
NT79
T79
NT80
T80
BEAS-2B
Rep. 2
NT76
T76
NT79
T79
NT80
T80
BEAS-2B
Rep. 1
NT81
T81
NT83
T83
NT85
T85
BEAS-2B
Rep. 2
NT81
T81
NT83
T83
NT85
T85
BEAS-2B

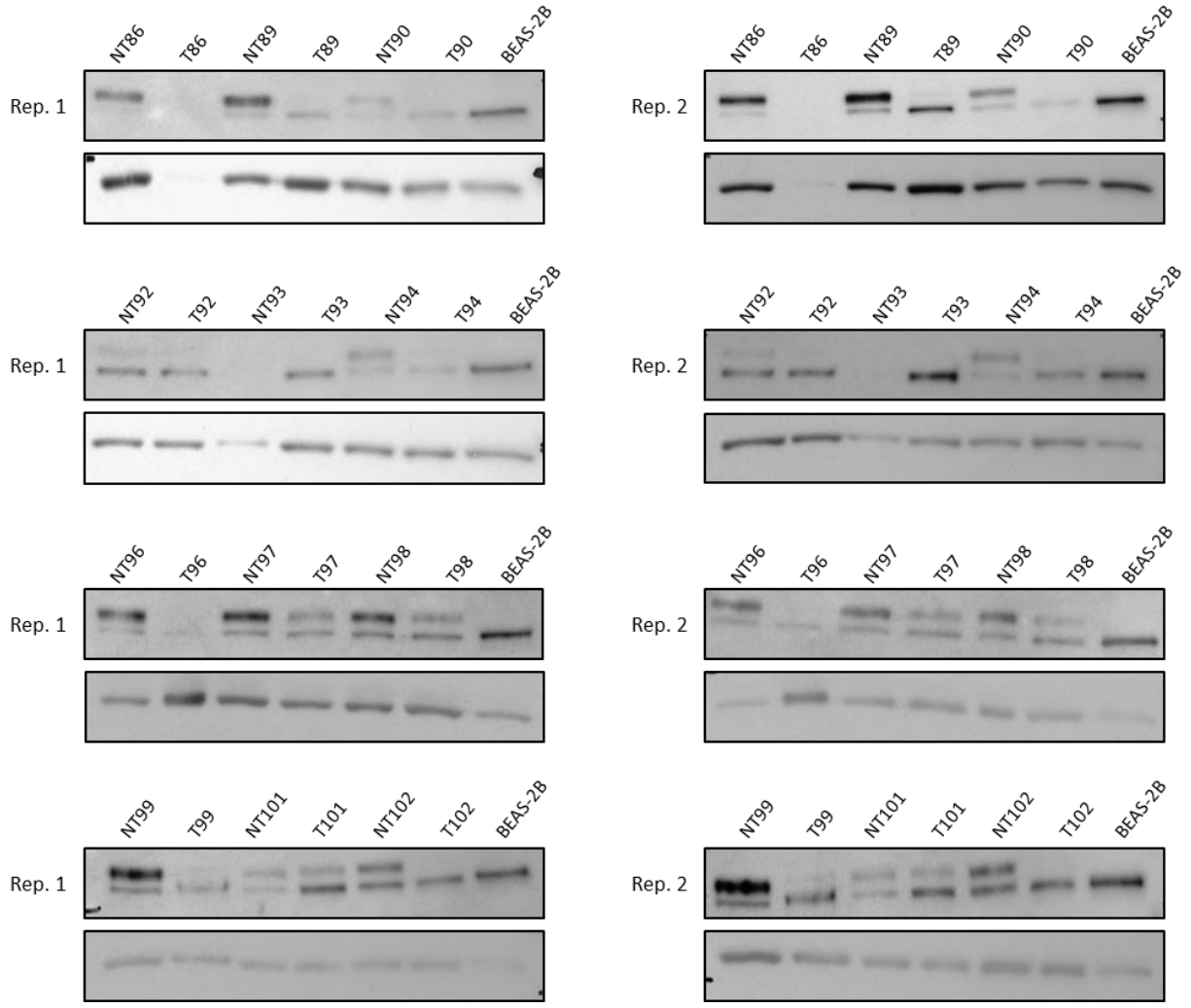


Figure A2 Western blots containing patient non-tumour and tumour samples that were probed with the LUCA-15 UK antibody. Western blot replicates containing patient non-tumour and tumour protein probed with the LUCA-15 UK antibody, demonstrating the 113 kDa and 124 kDa bands. ACTB was used as the loading control for all patient samples. Upper panel: RBM5. Lower panel: ACTB. Red letters = mixed NSCLC samples that were not included in statistical analyses.

would saturate the antibody, preventing the antibody from binding to specific targets. Alternately, the non-specific LSLN and LSLC peptides were expected to eliminate non-specific binding on the membrane, allowing the antibody to bind more specifically to true RBM5 proteins. Since it was suspected that the 124 kDa band was, in fact, a post-translationally modified version of the RBM5 protein, it was expected that both the 124 kDa and 113 kDa bands would not appear when blocking the antibody with the UK peptide, but remain unaffected when the antibody was blocked with the LSLN and LSLC peptides. Results of the peptide block are displayed in Figure A3.

When using the LUCA-15 UK antibody on its own, both bands of interest were observed at 113 kDa and 124 kDa, respectively, along with a number of other bands below 100 kDa (Figure A3 A). As anticipated, when the LUCA-15 UK antibody was blocked with UK peptide, both 113 kDa and 124 kDa bands disappeared, while non-specific bands below 100 kDa remained, with the exception of the band located at 40 kDa (Figure A3 B). The 40 kDa band was observed previously in Jurkat cells [37]. The authors of the paper that previously identified the 40 kDa band stated that the protein did not correspond to any mRNA variants, and may, therefore, represent a proteolytic cleavage product derived from the full-length protein [37]. Expression of the 40 kDa band was not analyzed in this study. After blocking with LSLN and LSLC peptides, both 113 kDa and 124 kDa bands were visible once again (Figure A3 C and A3 D). Taking the results of the peptide blocks as a whole, it was concluded that the 124 kDa upper band was, indeed, RBM5 protein, and hypothesized that it was post-translationally modified. It was also realized that the LSLC peptide reduced the background associated with the LUCA-15 UK antibody. As such, the

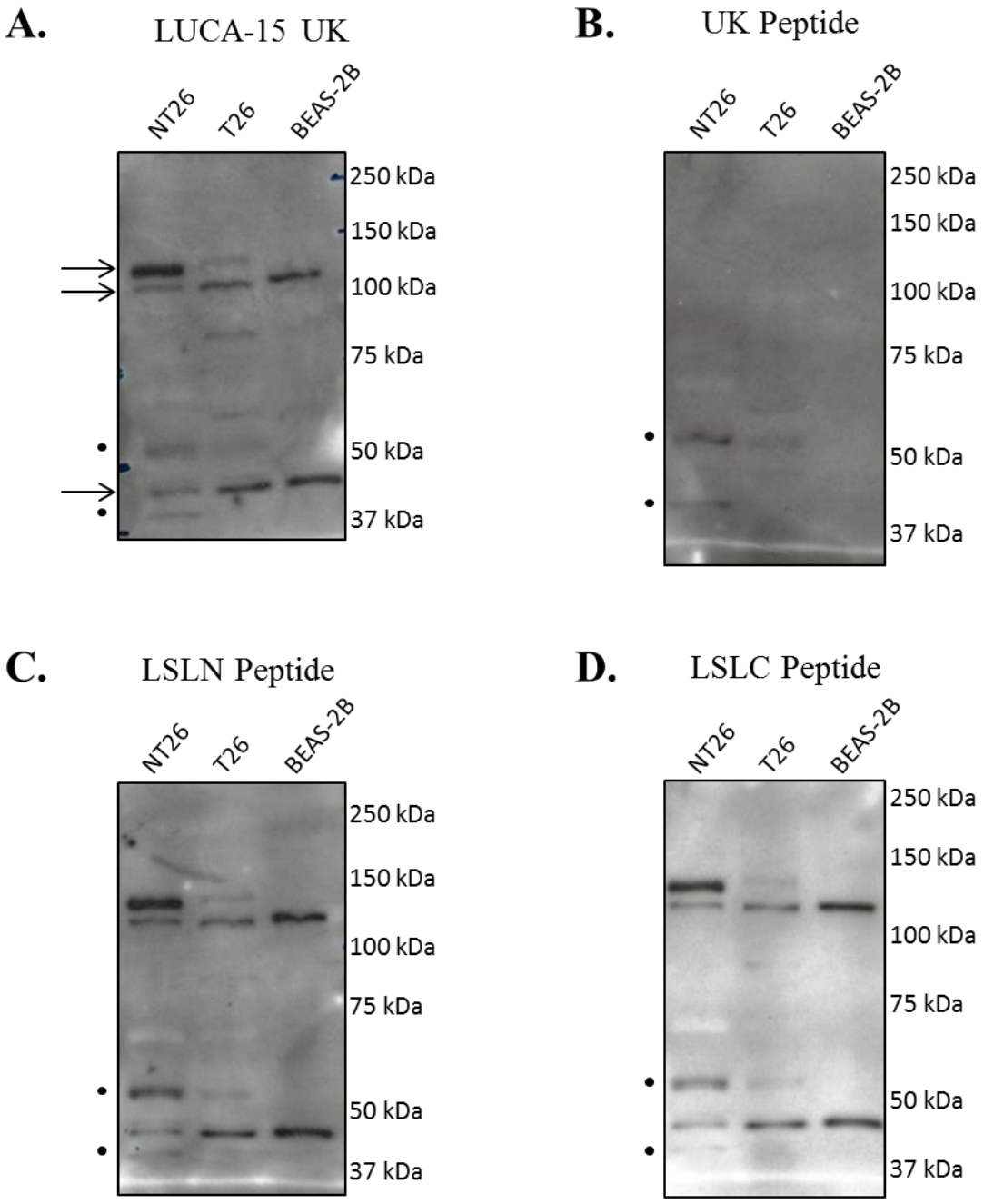


Figure A3 LUCA-15UK peptide block analyses. Four identical PVDF membranes containing the same protein samples were probed with (A) 1:3500 LUCA-15 UK, (B) 1:3500 LUCA-15 UK pre-incubated with 30 µg UK peptide, (C) 1:3500 LUCA-15 UK pre-incubated with 30 µg LSLN peptide, and (D) 1:3500 LUCA-15 UK pre-incubated with 30 µg LSLC peptide. NT26: non-tumour lung tissue from patient 26; T26: tumour lung tissue from patient 26; BEAS-2B: lung cell line. Marker: Precision Plus Protein Ladder (Bio-Rad). Arrows indicated specific LUCA-15 UK bands that were blocked by the UK peptide. Dots indicate non-specific bands detected by the LUCA-15 UK peptide.

LUCA-15 UK antibody was pre-incubated with LSLC peptide prior to use on PVDF membranes containing patient samples.

Duplicate Western blots were performed with paired non-tumour and tumour protein samples from 60 patients, including 42 diagnosed with ADC and 18 diagnosed with SqCC, respectively. Expression of each RBM5 band was quantified by densitometry and normalized to ACTB expression. Expression of all RBM5 bands were quantified, and the average expression of each band was calculated for duplicate samples. Average normalized RBM5 expression values were then graphed. Outliers were identified using the Tukey method, and excluded prior to statistical analyses. The normalized RBM5 expression values in non-tumour and tumour data sets did not follow a normal distribution, as per the Shapiro-Wilk test, meaning that non-parametric analyses were ideal for subsequent statistical tests. Since the non-tumour and tumour samples were paired, the non-parametric Friedman test was used to determine if there was a statistically significant difference in mean expression of the various RBM5 bands in paired samples. Results are displayed in Figure A4. Combining the data from patients diagnosed with ADC and SqCC, it was clear that the reciprocal expression patterns of the RBM5 bands were conserved across the majority of patient samples (Figure A4 A). The upper band was significantly more expressed than the lower band in non-tumour tissue, while the lower band was significantly more expressed than the upper band in tumour tissue. Comparing non-tumour tissue to tumour tissue, the expression of the upper band significantly decreased in tumour tissue, while the expression of the lower band significantly increased in tumour tissue. Examining the patients diagnosed with ADC and SqCC separately, this expression pattern was conserved (Figure A4 B and A4 C).

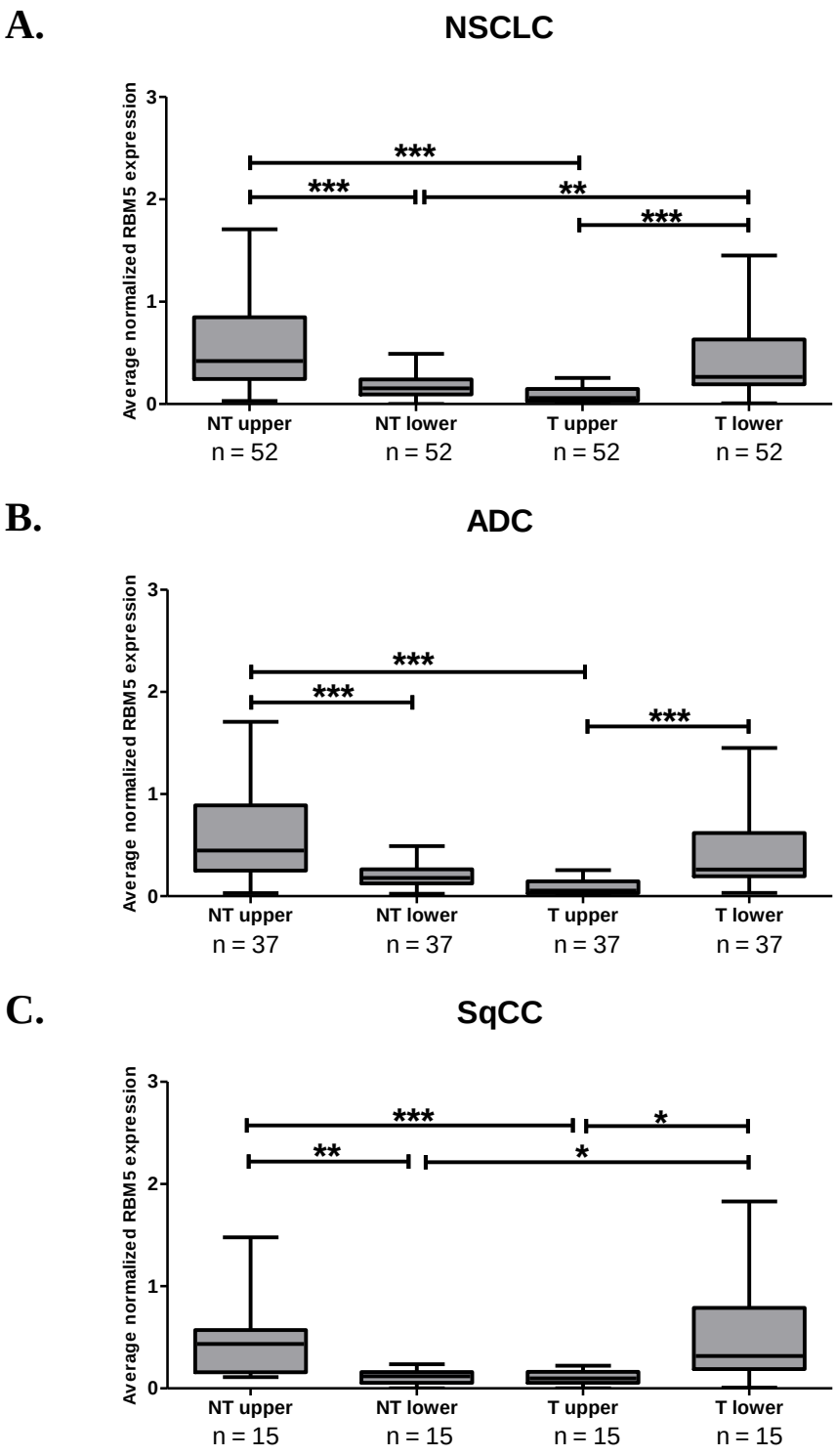


Figure A4 Expression analyses of 113 kDa and 124 kDa RBM5 bands. A Freidman test was used to determine if there were significant differences in expression of 113 kDa (lower) and 124 kDa (upper) bands in the non-tumour and tumour tissue of patients diagnosed with (A) NSCLC, (B) ADC, and (C) SqCC. *p<0.05, **p<0.01, ***p<0.001. NT = non-tumour, T = tumour.

A.3.2 Total RBM5 expression decreased in NSCLC tumours

Since the original objective of this study was to determine if RBM5 protein was significantly reduced in NSCLC tumours, the difference in mean total RBM5 expression was compared between non-tumour and tumour tissue, taking the total expression of both of the 113 kDa and 124 kDa RBM5 bands into account. For each patient sample, average normalized total expression values were calculated from the duplicate Western blots. Outliers were identified using the Tukey method, and excluded prior to statistical analyses. The average normalized total RBM5 expression values did not follow a normal distribution according to the Shapiro-Wilk test. Since the non-tumour and tumour samples were obtained from the same patients, the paired Wilcoxon signed-rank test was used to determine if the difference in mean normalized total RBM5 expression was significant between non-tumour and tumour tissue. As expected, RBM5 expression was significantly decreased in NSCLC (Figure A5 A). Separating ADC and SqCC, total RBM5 expression significantly decreased in ADC tumours, compared to normal adjacent lung tissue (Figure A5 B). Total RBM5 expression in SqCC tumours, however, was not significantly decreased (Figure A5 C).

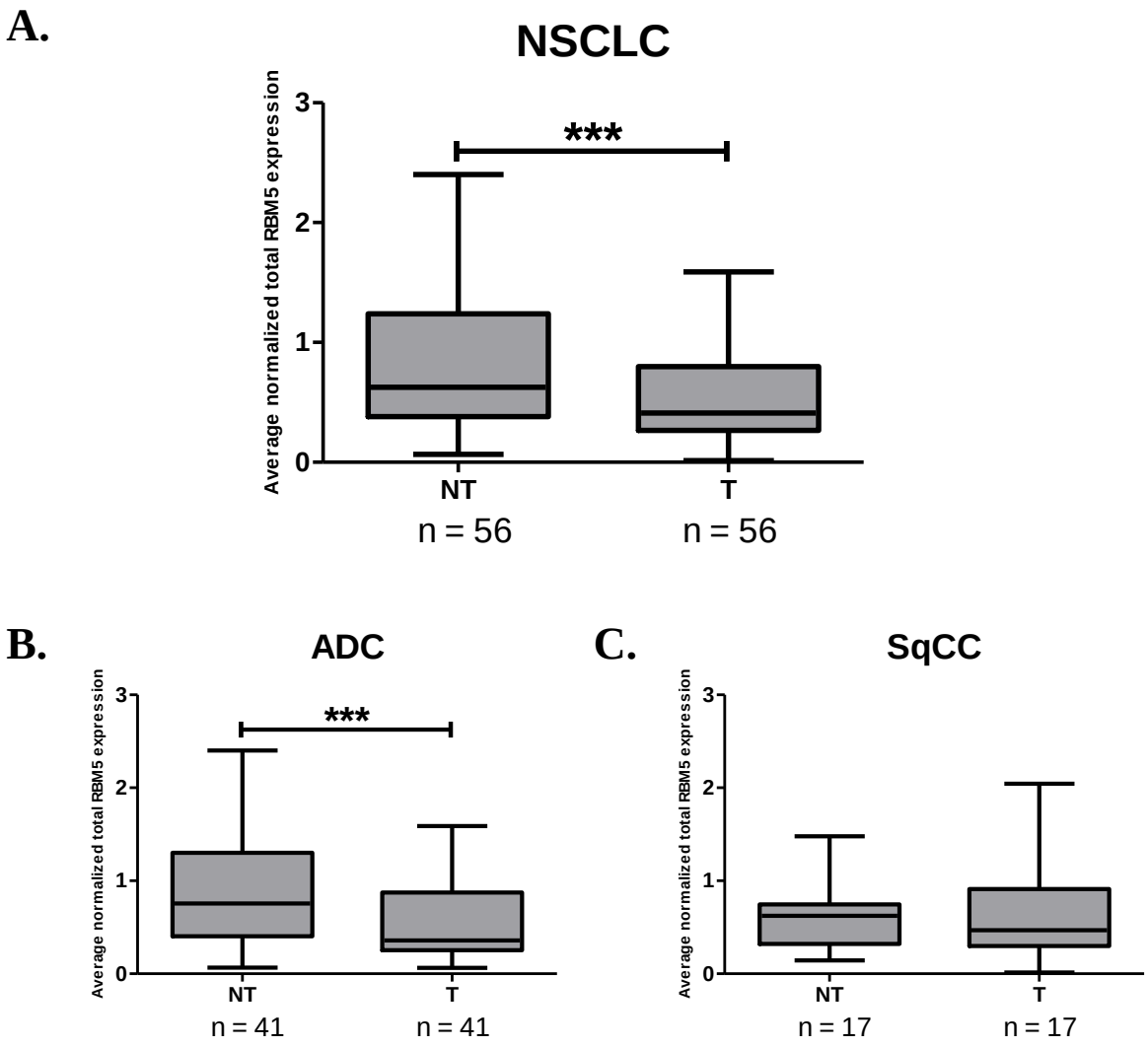


Figure A5 Expression analyses of total RBM5 protein expression. A Wilcoxon signed-rank test was used to examine significant differences in the mean total RBM5 expression values between non-tumour and tumour tissue in patients diagnosed with (A) NSCLC, (B) ADC, and (C) SqCC. ***$p<0.001$.

A.4 Discussion

Since it was confirmed that RBM5 mRNA was significantly downregulated in NSCLC tumours, and the paired protein samples were available from the HSN cohort, it was possible to confirm the extent of RBM5 downregulation in NSCLC. Concerning total RBM5 protein expression, it was confirmed that RBM5 expression was significantly downregulated in NSCLC tumours, especially in ADC tumours, compared to adjacent normal lung tissue. Interestingly, the LUCA-15 UK antibody detected a specific RBM5 protein at 124 kDa that had not previously been observed in other studies, in addition to the expected RBM5 protein observed at 113 kDa. The upper band was thought to be a post-translationally modified RBM5 protein that is specific to lung tissue. The expression of the 113 kDa and 124 kDa bands were reciprocal in nature, with the upper band being more highly expressed in the non-tumour tissue, and the lower band being more highly expressed in the tumour tissue. The reciprocity of the two specific RBM5 proteins alluded to a novel mechanism of RBM5 protein regulation in lung tissue.

After all of the HSN NSCLC protein samples were analysed, a Western blot that was performed previously in the Sutherland Lab, using the LUCA-15 UK antibody, was examined to see if the 124 kDa band was present in patients diagnosed with SCLC. The non-tumour and SCLC tumour tissue samples were obtained from the Ontario Tumour Bank. Paired non-tumour and tumour samples were received from two patients from Hamilton and Ottawa, whereas only tumour sample was obtained from one patient from Kingston. Looking at Figure A6, the 124 kDa RBM5 band appeared in the Hamilton non-tumour sample, and appeared faintly in the Ottawa non-tumour sample. In regards to the tumour samples, the 124 kDa band did not appear in any of the tumour protein samples. This result confirmed that the presence of the 124 kDa band was not limited to the HSN primary tissue specimens.

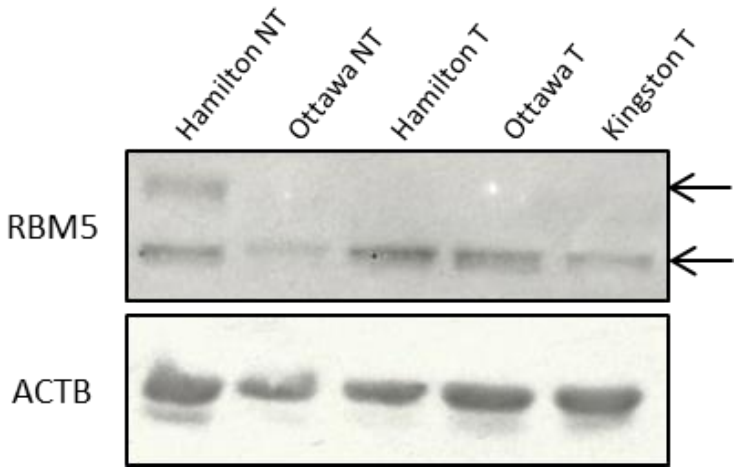


Figure A6 Confirmation of 113 kDa and 124 kDa RBM5 bands in Western blots from patient samples not included within the HSN cohort. A single Western blot containing Ontario Tumour Bank SCLC protein, and adjacent non-tumour protein, was probed with LUCA-15 UK antibody (upper panel). ACTB was used as the loading control (lower panel). Arrows indicate the 113 kDa and 124 kDa RBM5 bands, respectively.

It was hypothesized that the 124 kDa RBM5 protein was a version of RBM5 that had undergone extensive post-translational modification. Phosphorylation of RBM5 protein has been observed previously, but the change in molecular weight was only 5 kDa, approximately [17]. The difference of 5 kDa already suggests extensive phosphorylation, as the addition of only one phosporyl group is expected to alter the molecular weight of a protein by 80 Da [142]. The fact that a difference of approximately 11 kDa was observed suggests that, if RBM5 is post-translationally modified, the amount of modification is extensive. Using online resources, such as PhospoSite Plus®, that report experimentally observed post-translational modification to amino acid residues, we found that, in addition to phosphorylation, the RBM5 protein can also undergo acetylation, methylation and sumoylation [143]. Looking at the RBM5 protein sequence, it also appears as

though RBM5 contains both myristoylation and glycosylation sequence motifs, however, neither of these post-translational modifications have been proven experimentally. According to a review of common post-translational modifications, modifications including myristoylation, acetylation, methylation and glycosylation are only expected to alter the protein molecular weight by 210 Da, 42 Da, 14 Da and >800 Da, respectively [144]. Sumoylation, on the other hand, involves the covalent attachment of a SUMO protein, which can alter the molecular weight by approximately 15 kDa [145].

In order to confirm or disprove the presence of any of these post-translational modifications, it would be necessary to treat protein lysates with enzymes or chemicals that would remove the various modifications, and then perform a Western blot analysis of the protein samples. If the position of the upper band shifts downward after specific treatments, and after being probed with LUCA-15 UK, we can determine which modifications are present. Sumoylated proteins can be detected by isolating RBM5 protein by immunoprecipitation, performing a Western blot containing the immunoprecipitated protein sample, and then probing with a SUMO antibody [146]. Performing these experiments would be the next step in determining which modifications are present. Pinpointing which modifications are present may provide functional information regarding novel RBM5 regulation in normal and tumour lung tissue.

A.5 Conclusion

It was previously confirmed that RBM5 mRNA was significantly downregulated in the NSCLC tumours of the HSN cohort. To confirm the extent of RBM5 downregulation in this cohort, RBM5 expression was examined in non-tumour and tumour protein. As expected, RBM5 protein was significantly downregulated in NSCLC tumours. What was unexpected, however, was the presence of two RBM5 proteins with reciprocal expression in non-tumour and tumour tissue,

believed to be the result of post-translational modification of RBM5 protein. Further analyses must be completed in order to confirm which modifications are present, and how these modifications impact the regulation of RBM5 protein in lung tumours.