The role of Grb2 splicing in colorectal cancer development

Caroline Seiler

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds School of Molecular and Cellular Biology

September 2019

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgments

I would like to express my sincere gratitude to the people who have helped me along my journey to achieving this PhD. Firstly to Dr Darren Tomlinson, for allowing me to spend two summers working in his lab during my undergraduate degree, as this sparked my desire to pursue a career in science research and more specifically to focus on cancer research. Secondly to Professor Alan Berry, for giving me the opportunity to join the Wellcome Trust PhD scheme and for his constant support throughout the process. I am grateful to Professor John Ladbury for allowing me to carry out my PhD in his lab, for giving me the freedom to work independently and grow as a scientist and for supporting me to develop my own ideas. I would also like to thank Dr Chris Jones, for being a great friend in the lab and for all his guidance and advice with clinical aspects of the project. Special thanks go to the other PhD students in the lab, Janne, Ellie, Dova, Amy and Sophie that have made my time in the lab fun and full of banter.

Special thanks go to my partner Ryan Rhodes, for all his support in particular during the last year of my PhD, and for believing in me more than I believe in myself. Lastly but not least I would like to thank my parents, for their constant support, advice and guidance throughout my life.

This thesis is dedicated to the loving memory of my Grandma Gwen, my Aunty Maralyn and both my Granddad's who sadly lost their battle to cancer. I sincerely hope that this research contributes in a small way towards the bigger goal of understanding cancer and developing better medicines for patients. I hope to fulfil a career in research dedicated to this goal and look forward to a future where cancer is a disease of the past.

Abstract

Colorectal cancer (CRC) is one of the four most common cancers globally and is a major cause of death. Hyper-activation of the Ras-Erk signalling cascade is a key molecular event in colorectal cancer development. In 45-50% of patients this occurs through activating mutations in the Ras and Raf proteins. However, in the remaining 50% of patients it is less clear how Ras signalling becomes up-regulated; one potential mechanism is changes in *GRB2* splicing.

Grb2 is an adaptor protein that mediates activation of the Ras-Erk signalling cascade. It does so by recruiting the Ras guanine nucleotide exchange factor Sos to activated receptor tyrosine kinases (RTKs) by binding to the receptor through its SH2 domain and Sos through its SH3 domains. Alternate splicing of *GRB2* produces a splice variant; Grb3-3 that lacks exon 4 resulting in the production a protein isoform with a deletion in the SH2 domain that renders it non-functional.

In this study I show using a splice sensitive PCR assay that *GRB2* splicing is altered in a subset of CRC patients resulting in reduced levels of the truncated isoform Grb3-3. This correlated with elevated Ras activation in the absence of genetic mutations. Using a combination of in vitro and in cell analysis I show that Grb3-3 has antagonistic functions to Grb2 in the Ras signalling cascade. Grb3-3 inhibits Ras activation by binding to Sos and inhibiting Grb2 binding and membrane recruitment. Bioinformatic analysis and in vitro (pull-down) analysis identified hnRNPC as the splicing factor that regulates *GRB2* splicing. hnRNPC binds to *GRB2* alternate exon 4 and suppresses exon inclusion; enhancing Grb3-3 production. hnRNPC expression is reduced in CRC highlighting a novel mechanism for up-regulation of Ras signalling in the absence of genetic mutations.

Table of contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 RNA splicing

During gene expression, double stranded DNA is transcribed into single stranded RNA, known as pre-mRNA. Pre-mRNA is composed of both protein coding exons typically a few hundred bases in length separated by much longer non-coding introns usually several thousand nucleotides in length. The pre-mRNA then undergoes a maturation process that involves polyadenylation of the 3' end and RNA capping of the 5' end. In eukaryotes most transcripts undergo an additional maturation step that involves the removal of non-coding introns; this process is known as RNA splicing.

RNA splicing was first hypothesised in the 1960s and 1970s from experiments using radiolabelled uridine nucleotides. These experiments showed that a large fraction of nuclear RNA was degraded very rapidly after synthesis, and only a small fraction was exported from the cytoplasm [1]. The short-lived nuclear RNA was up to several tens of thousands of bases in length; much longer than the cytoplasmic RNA. This lead to the initial hypothesise that cytoplasmic RNA was produced by cleavage of much larger precursor RNA. Further knowledge on RNA splicing came from studying gene expression in adenoviruses [2, 3]. The Sharp lab used electron microscopy to study RNA:DNA hybrids that were formed during gene expression. They observed that the RNA segment of the adenovirus genome coding for the hexon polypeptide contained sequences from non-contiguous sites in the viral genome, produced by the joining of shorter segments of RNA now known as exons and the removal of intervening sequences (introns) from the longer nuclear RNA to generate the mature RNA [3]. Shortly after the discovery of RNA splicing in adenoviruses and advances in molecular biology techniques it was

observed that the vast majority of human genes also contained introns that were removed from the final mRNA by splicing [4]. The process of splicing is conserved across species from yeast to humans. Although the prevalence of splicing varies, for example, in yeast it is much less than in humans where 95% of the genome undergoes RNA splicing.

1.1.1 The biochemical splicing mechanism

The primary signals that determine where splicing occurs are encoded in the pre-mRNA sequences. These include the 5' splice site (donor site), the branch point sequence, the polypyrimidine tract and the 3' splice site (acceptor site), these sequences act to define intron-exon boundaries [5]. The 5' splice site is located at the exon/intron junction, the exon terminates in nucleotides AG and the intron begins with the conserved nucleotides GU (Figure 1.0). Mutations in either of the conserved GU nucleotides completely abolishes splicing [6]. The branch point is an adenine within the sequence YNYURAY where N can be any nucleotide, R represents purine and Y is a pyrimidine. The branch point is located within the intron 19-32 nucleotides downstream of the 5' splice site. Downstream of the branch point is the polypyrimidine tract, which is most commonly a repetition of 11 uridines but has also been reported to be a repetition of cytosines. The final sequence is the 3' splice site, this lies at the intron/exon junction. At this junction the intron terminates in the sequence AG and the neighbouring exon begins with a guanine nucleotide [7].

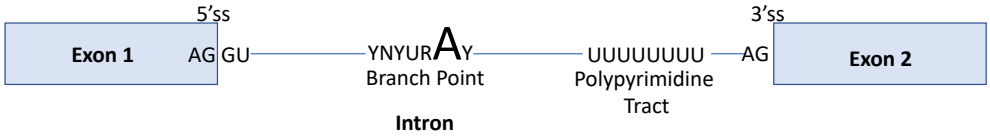


Figure 1.0 Schematic representation of the conserved sequence elements required for the splicing process. pre-mRNA containing two exons shown as blue rectangles and the intervening intron shown as a blue line. The 5' splice site is located at the exon 1 intron junction, the exon terminates in AG and the intron begins with GU. The intron contains the branch point sequence and the polypyrimidine tract. The capital A in the branch point sequence represents the adenine responsible for nucleophilic attack of the 5' splice site. The 3' splice site is located at the boundary between the intron and exon 2, the intron terminates in the conserved nucleotides AG.

Splicing of pre-mRNA molecules occurs via 2 consecutive transesterification reactions, as shown in Figure 1.1 [8-10]. Splicing of tRNA molecules is unique and does not occur by transesterification. The first reaction occurs at the 5' splice site, where exon 1 is cleaved through nucleophilic attack by the 2' hydroxyl of the adenine nucleotide within the branch point sequence. This generates a 5' exon fragment and a lariat intermediate that contains the intron and exon 2. The second transesterification reaction occurs at the 3' splice site, where exon 2 is cleaved through nucleophilic attack by the 3' hydroxyl of exon 1. The two exons then ligate and the intron is released in the form of a lariat.

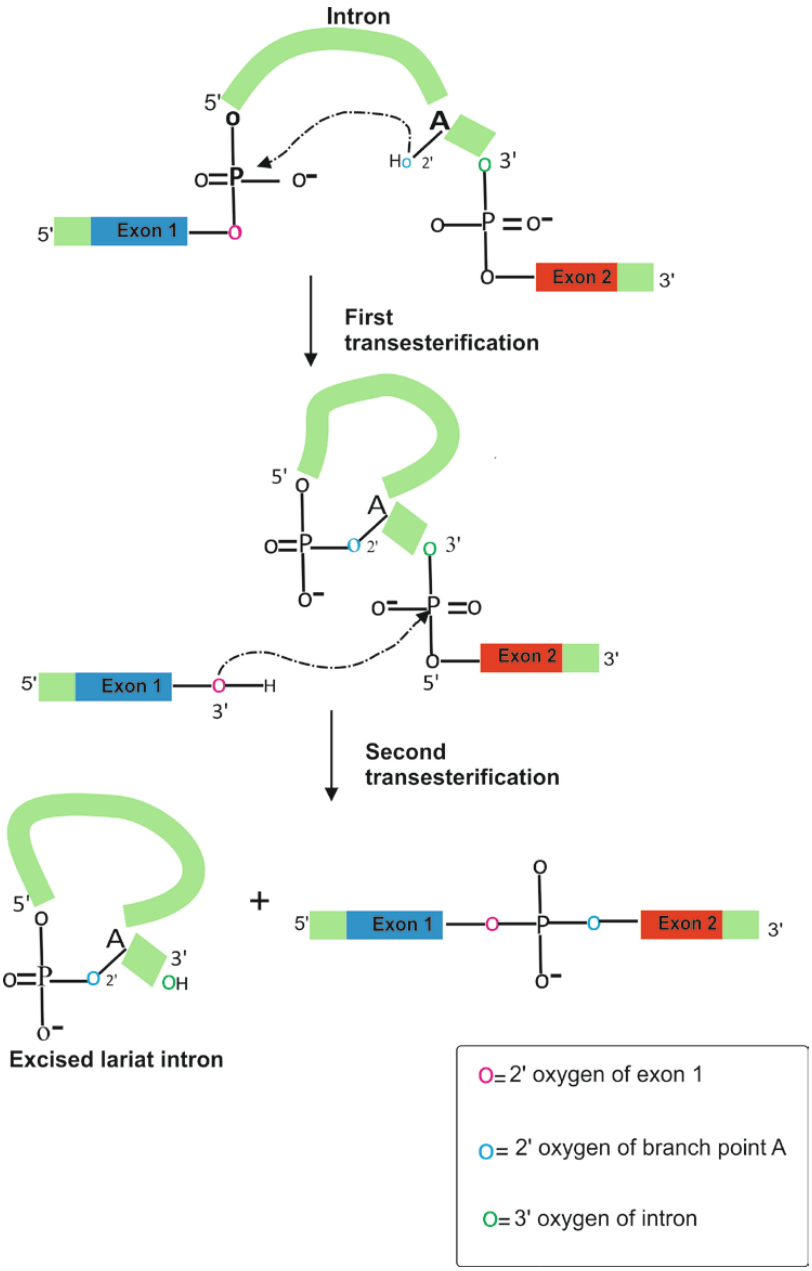


Figure 1.1 Schematic representation of the biochemical splicing reaction. The adenine in the branch point sequence of the intervening intron attacks the guanine nucleotide at the 5' splice site of exon 1 during the first transesterification reaction. In the second transesterification reaction the 3' splice site is cleaved by the hydroxyl group of exon 1.

1.1.2 The spliceosome

RNA splicing is catalysed by the spliceosome; a highly dynamic macromolecular complex composed of several small nuclear RNAs (snRNAs) and approximately 80 associated proteins [11, 12]. The spliceosome is assembled on the pre-mRNA prior to each splicing reaction [13]. In eukaryotes the spliceosome is composed of 5 snRNAs denoted U1, U2, U4, U5 and U6. When these small RNAs are combined with the associated proteins they form RNA-protein complexes called small nuclear ribonucleo proteins (snRNPs) [14].

1.1.3 Spliceosome assembly

Spliceosome formation involves the stepwise assembly and disassembly of different snRNPs on the pre-mRNA (Figure 1.2) [15]. In humans, four distinct spliceosomal complexes form in a sequential order: complexes E, A B and C [16]. Complex E, A and B contain the unspliced pre-mRNA, complex C contains the products following the first transesterification reaction (the 5' exon and the lariat-3' exon) [11].

Assembly of the spliceosome begins with the recognition of the 5' splice site by binding of U1 snRNP to the conserved GU nucleotide at the start of the intron in an ATP independent manner. In addition splicing factor 1 (SF1) binds to the adenine nucleotide within the branch point sequence in the intron. SF1 then interacts with the U2 axillary factor 1 and 2 (U2AF1 and U2AF2). U2AF1 and 2 bind to the 3' splice site and the polypyrimidine tract respectively to form complex E.

The formation of the A complex also known as the pre-spliceosome occurs through the recruitment of the U2 snRNP to the branch point site through

interactions with the U2 auxiliary factor (U2AF) resulting in the displacement of SF1 in an ATP dependent manner.

Following the interaction of U1 and U2 snRNPs with the pre-mRNA U4, U5 and U6 snRNPs form a trimer which gets recruited to the pre-mRNA. U5 snRNP binds the 5' splice site and U6 snRNP binds to U2 snRNP at the branchpoint forming complex B, or the pre-catalytic spliceosome, ready for the first transesterification reaction. Although this complex contains all the snRNPs required for splicing, it lacks a catalytic centre so is unable to catalyse the splicing reaction. Activation of the pre-catalytic spliceosome involves rearrangements of RNA-RNA interactions between the pre-mRNA and the snRNPs which results in the release of U1 and U4 snRNP. U6 snRNP binds to the 5' splice site through its conserved ACAGAG motif resulting in the release of U1 and the disruption of the U4/U6 interaction causing the release of U4 snRNP. The U2 and U6 snRNPs base pair with one another and with the branch point sequence and 5' splice site respectively. These rearrangements result in the formation of the catalytically active B complex also known as the B* complex. The B* complex promotes the first transesterification reaction to generate the free 5' exon and the lariat 3' exon intermediate.

Following this step the final spliceosome complex is formed, complex C, which catalyses the second transesterification reaction. U5 snRNP is important in this step as it holds the 5' and 3' splice sites in close proximity. After the second transesterification reaction the spliceosome releases the mRNA, the lariat is degraded and the snRNPs are recycled for subsequent splicing events.

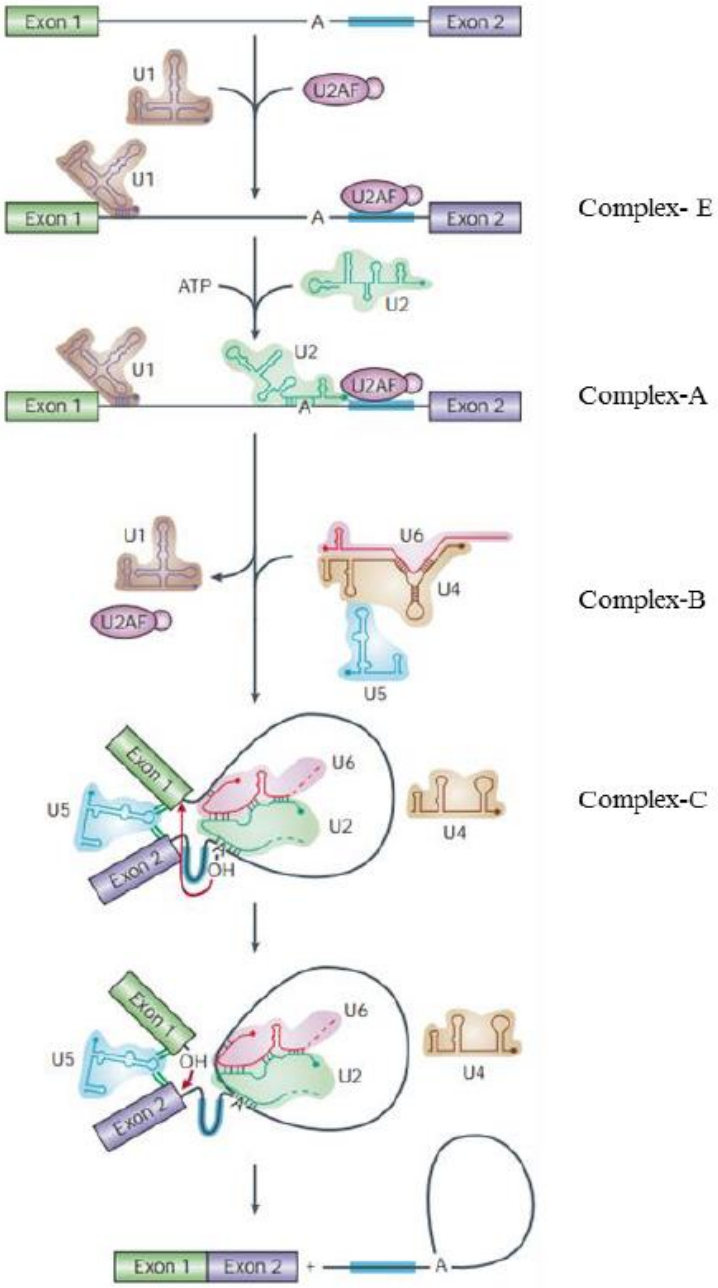


Figure 1.2 Stepwise assembly of the spliceosome. Complex E forms by U1 binding to the 5' splice site and U2AF binding to the intervening intron. Complex A forms by binding of U2 to the branch point sequence and U2AF. Complex B forms by the binding of the U4,5,6 trimer and the release of U1 and U2AF. Complex C is the final spliceosome complex which catalyses the transesterification reactions. Figure adapted from Schroeder et al.

1.2 Regulation of RNA splicing

RNA splicing must be tightly regulated as many human diseases are caused by errors in splicing [17, 18]. Correct splicing by the spliceosome relies on the accurate definition of regions of RNA that are introns and regions that are exons. This can be challenging as the non-protein coding introns are much longer than the exons. Although the correct definition of intron-exon boundaries is mediated by the 5' and 3' splice sites, the polypyrimidine tract and branch point these sequence elements alone are insufficient for correct RNA splicing to occur. Only 50% of the information required for splicing is at the splice site, additional cis-acting regulatory sequences are required for correct splicing.

1.2.1 Splicing regulatory sequences

These cis-acting sequences are usually 4-10 nucleotides in length and form the binding site for trans-acting splicing factors (RNA binding proteins), which either promote or supress assembly of the spliceosome at particular exon/intron junctions. The regulatory sequences can be defined based on their location: exonic splicing enhancers (ESE) or silencers (ESS) are located within the exon and intronic splicing enhancers (ISE) or silencers (ISS) occur within the intron (Figure 1.3) [19].

1.2.1.1 Exonic splicing enhancers

ESEs are short DNA sequences usually 6 bases in length. ESEs function to enhance the assembly of the spliceosome at particular splice sites. They are usually enriched at weaker splice sites and in exons which are flanked either side by large introns, as it is more difficult to identify the splice site within these longer sequences. ESEs are typically located 50-100 base pairs from the start or the end of the exon, in close proximity to exon intron boundaries. ESEs most commonly

form the binding sites for serine arginine (SR) rich proteins. SR proteins are a large family of splicing factors that usually function as splicing enhancers (discussed in more detail in Section 1.2.2).

The SR proteins predominantly recognise purine rich sequences, for example the consensus binding site for the serine-arginine rich splicing factor 1 (SRSF1) is over 80% purine nucleotides. ESEs are therefore often also enriched in purines although other AC rich and pyrimidine rich ESEs have been identified. The pure purine hexamer GAAGAA is one of the strongest ESEs. Mutations in ESE motifs contribute to aberrant splicing and have been identified in some genetic disorders and cancers. For example disruption of an exon splicing enhancer motif in exon 3 of the MLH1 gene causes hereditary nonpolyposis colorectal cancer [20].

1.2.1.2 Exonic splicing silencers

Exonic splicing silencers (ESSs) are much less well characterised than the ESEs. They are usually 4-18 nucleotide sequences and function to inhibit assembly of the spliceosome at the adjacent splice sites. ESSs contribute to regulation of both constitutive and alternate splicing (discussed in section 1.3). They have four general roles to inhibit exon inclusion (promote exon skipping), inhibit intron retention and to regulate the use of alternate 5' or 3' splice sites.

Wang et al identified 141 ESSs in a cell-based screen [21]. They generated a splicing mini gene containing 3 exons, the first and third exon encoded the complete mRNA for GFP, and the second exon was a test exon that is normally included to form an mRNA that does not encode a functional protein. They inserted a library of randomly generated decamers into the second exon. If the decamer contained an ESSs then the exon would be skipped and green fluorescence would be produced. The decamers that contained ESSs could be

sorted into 7 groups based on similarity. Two of these groups resembled the binding site for heterogenous nuclear ribonucleoprotein (hnRNP) A1 and a third group resembled the binding site for hnRNP H. The hnRNPs are a large family of RNA binding proteins that typically function as splicing suppressors. The ESSs identified from this screen were typically enriched in thymine and guanine nucleotides [21].

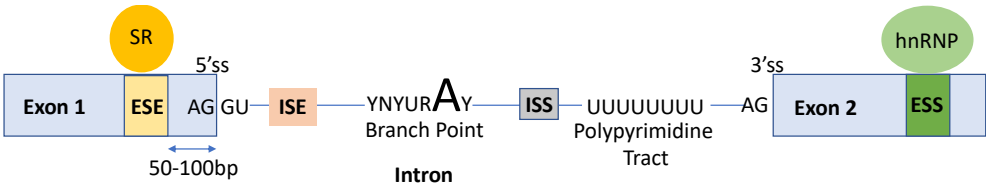


Figure 1.3 Schematic representation of the canonical and cis-regulatory RNA sequence elements required for correct splicing. The conserved sequence elements are shown as in Figure 1.0. Additional cis-acting regulatory sequence elements are also highlighted as coloured boxes. These include splicing enhancer and silencer sites, which can occur within the exon (ESE, ESS) or the intron (ISE, ISS). These sequence elements form the binding sites for trans-acting splicing regulatory proteins. The SR proteins typically function as enhancers and the hnRNPs typically function as repressors.

1.2.2 Splicing Regulatory Proteins

The trans-acting RNA binding proteins that recognise cis-regulatory sequences are known as splicing factors. SpliceAid-F is a database that contains information on all the splicing factor proteins. To date 71 splicing factors have been identified and experimentally validated [22-24]. Two major families of splicing factor proteins discussed in more detail below are the serine-arginine rich proteins (SR proteins) and the heterogenous nuclear ribonucleoproteins (hnRNPs). These two families of proteins regulate the recruitment of the core splicing machinery to different splice sites, which is important in both constitutive and alternate splicing.

1.2.2.1 The serine-arginine rich proteins (SR proteins)

The serine-arginine rich proteins are a large superfamily of structurally-related RNA binding proteins. They are involved in multiple steps of pre-mRNA processing, including RNA splicing, export, stability and translation. The SR proteins were discovered in *D. melanogaster* in the early 1990s and have since been identified in several human cell lines. Humans possess 12 relatively well characterised SR proteins, shown in Figure 1.4. The first SR proteins to be discovered in humans were SRSF1 and SRSF2, these proteins play essential roles in constitutive splicing by enhancing binding of U1 snRNP to the 5' splice site and U2 to the 3' splice site.

All SR proteins share two main structural features: at least one N-terminal RNA-recognition motif (RRM) and a C-terminal arginine-serine rich domain (RS domain). The RS domain is also found in other types of RNA binding protein domains including zinc finger domains and RNA helicases, which are also known as SR related proteins. The RS domain protein mediates protein-protein

interactions with other SR proteins. It also mediates nuclear import of the SR proteins through binding the SR protein nuclear import receptor: transporting-SR. The RNA recognition motif (RRM) mediates sequence specific binding to the RNA. Several SR proteins have more than one RRM; usually one will bind with a high specificity to the RNA consensus whereas the other will be much weaker.

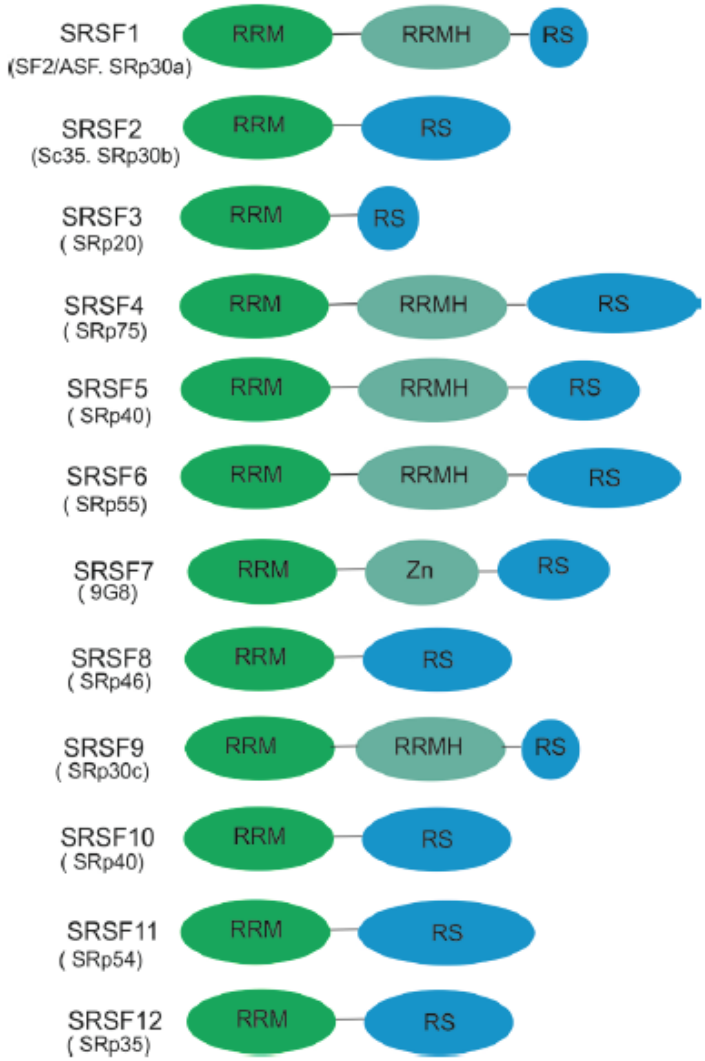


Figure 1.4 The SR protein families. Schematic representation of the 12 SR proteins identified in humans and their domain structure, RRM is the RNA recognition motif, some of the SR proteins have a second homologous RRM, RS is the arginine/serine-rich domain, and Zn is the zinc binding domain. Figure taken from Shepard and Hertel [25]).

The SR proteins typically function as splicing enhancers, promoting exon inclusion. Their role in RNA splicing is dependent on regulatory phosphorylation and dephosphorylation cycles. Phosphorylation of the RS domain is required for efficient splice site recognition whereas dephosphorylation is required to mediate catalysis of the splicing reaction. Treatment of cells with phosphatase inhibitors prevents splicing occurring. The requirement of phosphorylation of the SR proteins for facilitating spliceosome assembly is to prevent non-specific interactions of the highly positively charged RS domain with negatively charged RNA. Several kinases responsible for phosphorylating the SR proteins have been identified including the SR protein kinase 1 (SRPK1), Clk/Sty kinase and cdc2p34 [26, 27].

There are three different models for the mechanism of action of the SR proteins. The most widely accepted model is that the SR proteins bind to ESEs and recruit the spliceosomal machinery through protein-protein interactions mediated by the RS domain [28, 29]. Multiple SR proteins may bind to an ESE and have an additive effect on splice site selection. Another model is that the SR proteins bring the 5' and 3' splice sites closer by mediating protein-protein interactions across the introns through the interaction with U1 snRNP at the 5' splice site and U2AF at the 3' splice site mediated by the RS domain. A final model is that the SR proteins antagonise the inhibitory effect of splicing suppressors bound at ESS sites.

1.2.2.2 Heterogeneous ribonucleoproteins (hnRNP)

The hnRNP proteins are another family of RNA binding proteins that regulate splicing. At least 20 major hnRNP proteins have been identified in human cells that are named alphabetically from A to U, summarised in Table 1 [30]. The hnRNPs are a diverse group of proteins ranging in size from 34-120 kDa. Most

hnRNPs have a nuclear localisation signal and are predominantly localised to the nucleus, however some are able to shuttle between the nucleus and cytoplasm, suggesting these hnRNPs may also play a role in nuclear export of RNA [31]. The hnRNP family have four unique RNA binding domains (RBD); an RNA recognition motif (RRM) also found in the SR proteins, the quasi-RRM, an RGG box and a KH domain (Figure 1.5). The most common RBD is the RNA recognition motif, found in all but five hnRNP family members. In addition to the RNA binding motifs they possess auxiliary domains such as glycine, proline or acidic rich domains which are often involved in protein-protein interactions (Figure 1.5) [30].

In contrast to the SR proteins, hnRNPs typically act to supress exon inclusion. This is achieved predominantly through binding to exonic splicing silencer (ESS) elements or through interference with protein interactions that facilitate exon inclusion.

hnRNP	Molecular Weight (kDa)	RNA binding domains	Binding Sequence	Function	Link to disease
A1	34	2X RRM, Gly rich RGG	UAGGGA/U	Splicing, mRNA stability, translational regulation	ALS, Cancer
A2/B1	36/38	2X RRM, Gly rich RGG	UUAGGG	Splicing, mRNA stability	ALS, Alzheimer's disease, cancer
C1/C2	41/43	RRM, Acid rich	Poly U	Splicing, Translational regulation, Transcript sorting	Alzheimer's disease, Fragile X syndrome, cancer
D	44-48	2x RRM	AU rich	mRNA decay, Telomere maintenance	-
E1/E2/E3/E4	39	3x KH	Poly C	Translational regulation, Transcriptional regulation, mRNA stability, splicing	Cancer
F	53	3x qRRM, 2x Gly rich	UUAGG	Splicing, Telomere maintenance	ALS, Cancer
G	43	RRM, Glyc rich	CC(A/C)	Splicing	SMA
H	56	3x qRRM, 2x Gly rich	UUAGG	Splicing	ALS, cancer
I (PTB)	59	4x RRM	UCUU(C)	Splicing, mRNA stability, Transcriptional Regulation	-
K	62	3x KH, other	Poly C	Translational regulation, Transcriptional regulation, mRNA stability, Splicing	ALS, Cancer
L	68	4x RRM, Gly rich	CA repeats	Splicing, mRNA stability	-
M	77	3x RRM	Poly G/U	Splicing	SMA, Cancer
P (Fus)	72	2x Gly rich, RRM, 2x RGG	GGUG	Splicing	ALS
Q1/Q2/Q3	55-70	3X RRM, acid rich	UCUAUC	Splicing, Translational regulation	SMA
R	71	3X RRM, acid rich, RGG, other	UCUAUC	Transcriptional regulation, Translational regulation	SMA
U	120	Acid rich, other, Gly rich, RGG	GGACUGCRRUCGC	Splicing, Transcriptional regulation	-

Table 1.0 Summary of the heterogeneous nuclear ribonucleoproteins (hnRNPs) found in human cells. Table includes details of the molecular properties including molecular weight, binding domain, binding sequence, function and association with disease. Table adapted from Geuens et al [30].

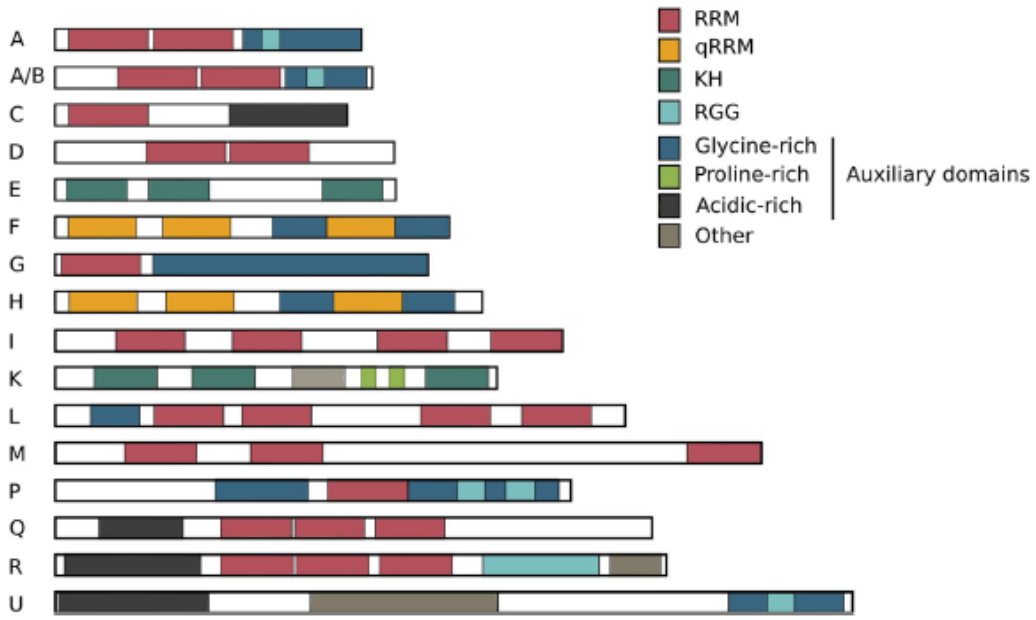


Figure 1.5 Schematic representation of the hnRNP family domain structure. All hnRNPs possess an RNA binding domain which can be an RNA recognition motif (RRM), quasi-RRM, a K-Homology domain or a RNA binding domain consisting of Arg-Gly-Glyc repeats (RGG). They also all possess auxillary domains which are required for protein-protein interactions with other splicing regulatory proteins. Figure taken from Geuens et al [30].

1.3 Alternate splicing

In addition to the removal of non-protein coding introns during RNA splicing multi-exonic genes can undergo alternate splicing (AS). Alternate splicing describes the inclusion of different combinations of exons and or introns in the mature mRNA transcript, therefore enabling a single gene to encode multiple different mRNA products. AS is an evolutionarily conserved process and a major source of transcriptomic and proteomic diversity. It is highly prevalent in humans; over 95% of multi-exon genes undergo this process. On average a single gene produces 7 different transcripts [32]. Specific alternate splicing patterns are seen in different tissues [33], at different developmental stages and more recently between normal and disease states [34, 35]. Mapping of alternatively spliced regions onto known protein structures suggests that the majority of alternate exons encode coiled or loop regions on the surface of the protein. These regions typically form binding interfaces and therefore alternate splicing can influence protein-protein interaction networks [36].

1.3.1 Mechanisms of alternate splicing

There are five mechanisms for alternate splicing as depicted in Figure 1.6 [37]. Firstly; and most commonly exon skipping. This occurs when an exon is not recognised as being an exon and is subsequently excluded from the mature mRNA. These exons are known as alternate, as opposed to constitutive exons that are always included in the final transcript. The second form of alternate splicing is intron retention, this occurs when an intron is not correctly spliced and therefore is included in the mature mRNA. Additionally, alternate 5’ and 3’ splice sites can be recognised by the spliceosome which can lead to the partial inclusion of exons. The final splicing mechanism is mutually exclusive exons, this occurs when there is an exclusive selection between alternate exons, meaning the

inclusion of one causes the exclusion of the other. The production of alternatively spliced mRNAs is regulated by trans-acting splicing factors in the same way as constitutive splicing as described in section 1.2 [38].

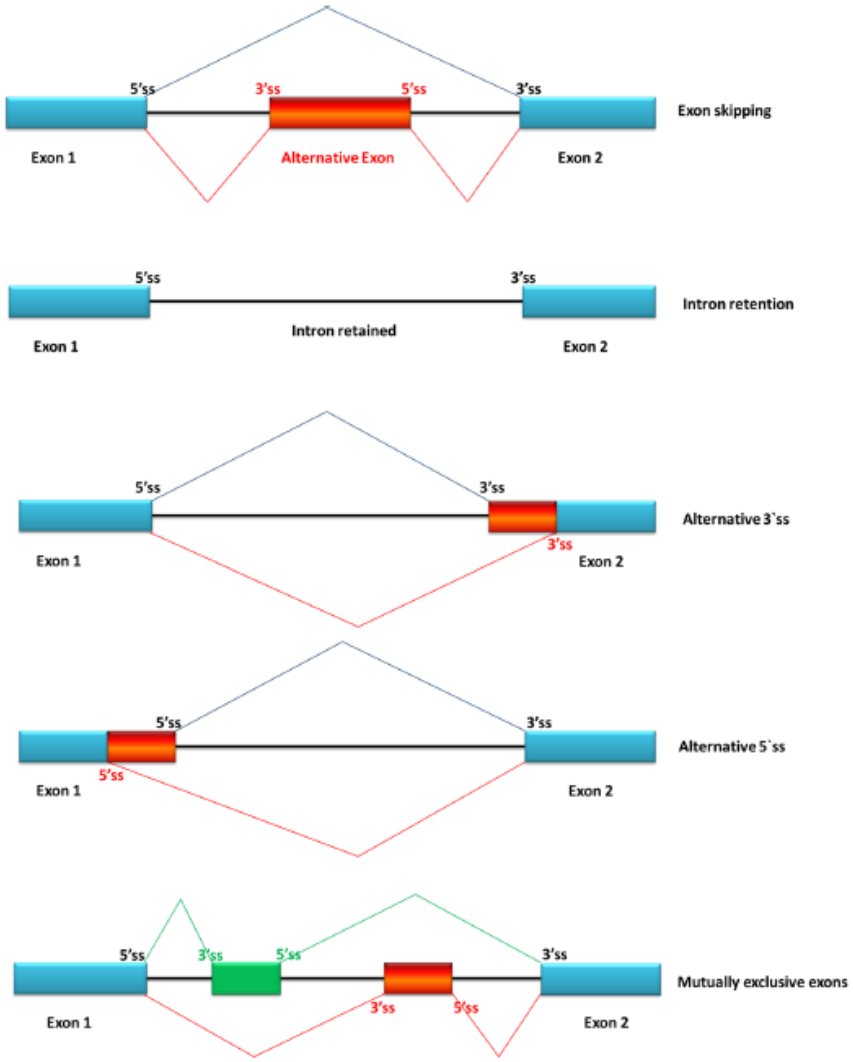


Figure 1.6 Mechanisms of alternate splicing. The five common mechanisms for alternate splicing are schematically represented. Constitute exons are represented as blue rectangles, alternate exons as red or green rectangles and introns as black lines. The different splicing events are highlighted using the red, blue and green coloured lines.

1.3.2 Alternate splicing and cancer

The process of alternate splicing is frequently disrupted in cancer in order to favour the production of isoforms that facilitate tumour development, progression and therapeutic resistance [39-42]. All areas of tumour biology are affected by changes in AS, in particular proliferation, apoptosis and angiogenesis [35, 43, 44]. For example, changes in *VEGF-A* splicing in cancer favour the production of a pro-angiogenic isoform that contains a longer version of exon 8 due to the use of an alternate 3' splice site. Additionally alternate splicing of the *BCL2* gene produces two isoforms; BCL-XL and BCL-XS. BCL-XL prevents apoptosis; this isoform is up-regulated in cancer leading to evasion of apoptosis (Figure 1.7) [35]. Genome wide analyses have identified thousands of alternate splicing events disrupted in cancer. For example in a study on chronic lymphocytic leukaemia 2000 genes had differential splicing patterns relative to the non-cancerous normal cells [45]. Table 1.1 summarises some of the alternate splicing events identified that are altered during tumour development and progression [46]. The challenge now, however is to understand the functional consequence of these alternate splicing events.

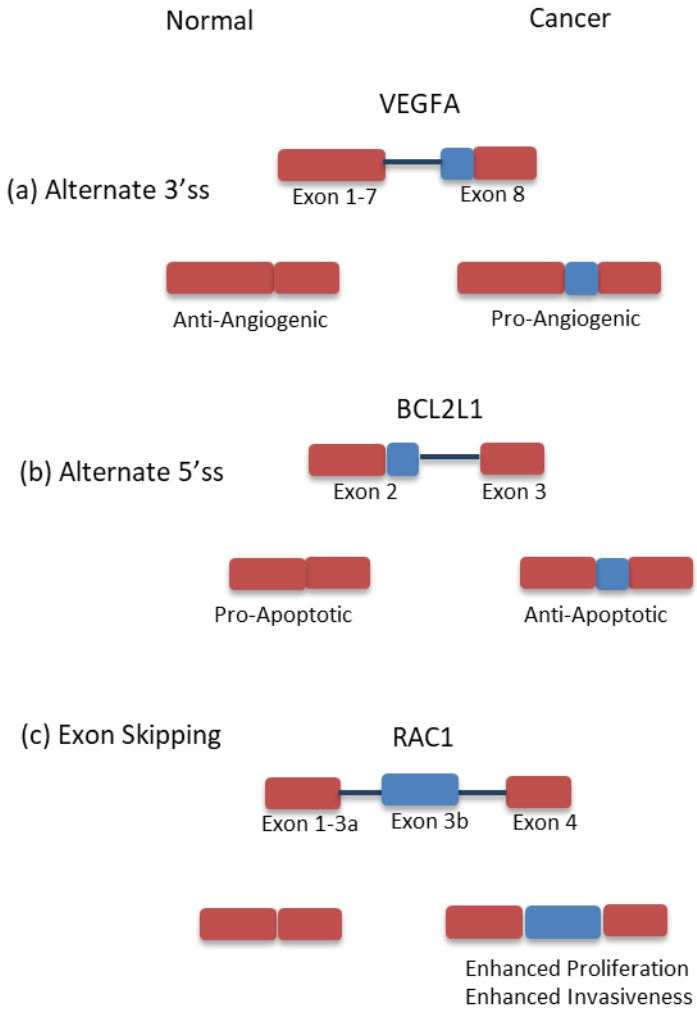


Figure 1.7 Examples of alternate splicing events that lead to the acquisition of the hallmarks of cancer. (a) Use of an alternate 3' splice site in VEGFA leads to the production of an isoform that promotes angiogenesis. (b) Use of an alternate 5' splice site in the BCL2 gene leads to the production of an anti-apoptotic isoform as oppose to a pro-apoptotic isoform. (c) The inclusion of exon 3b in the Rac1 gene produces an isoform that induces cell proliferation and invasiveness. Figure adapted from Sveen et al 2016 [35].

Gene	Description	AS Event	Role in Cancer
BIN1	This gene encodes several isoforms of a nucleocytoplasmic adaptor protein, one of which was initially identified as a Myc-interacting protein with features of a tumor suppressor.	Alternative exclusion/inclusion of the cassette exon 7. This shorter variant is also called IId and S1/R3-6 and binds dynamin, synaptojanin, and clathrin.	Caspase-independent apoptotic activation is impaired when aberrant isoforms are expressed.
CASP8	Member of the caspase family, which may interact with Fas-interacting protein FADD.	Alternative exclusion/inclusion of the cassette exon 4.	This protein is involved in apoptosis induced by Fas and various apoptotic stimuli.
ENAH	Response factor to mitogenic stimuli, such as EGF that triggers MAPK activation.	Alternative exclusion/inclusion of the cassette exon 12.	Functional role of hMena + 11a in breast cancer cell proliferation.
ERBB2	Member of the epidermal growth factor (EGF) receptor family of receptor tyrosine kinases that lacks the binding domain for growth factors. However, it can heterodimerize with other ligand-bound EGF receptor family members, enhancing kinase-mediated activation of downstream signaling pathways, such as those involving MAPK and IP3K.	Alternative start site, where isoform a corresponds to the full length, while isoform b shows a shorter N-terminus. Alternative exclusion/inclusion of the cassette exon 5.	Amplification and/or overexpression of this gene has been reported in numerous cancers, including breast and ovarian tumors.
FGFR1	Member of a family of trans-membranous receptors that possess an extracellular domain composed of three Ig-like domains, a single transmembrane helix-domain and an intracellular domain with tyrosine kinase activity.	Alternative site at the 5′ UTR, that generates the use of an alternative promoter, the inclusion of an alternate exon, the use of an alternate translation start site, and uses an alternate in-frame splice site, lacking two internal segments.	Key roles in proliferation, differentiation, and tumorigenesis.
PLAUR	uPAR was originally identified on the monocyte-like human cell line U937 as the membrane receptor for the serine protease urokinase-type plasminogen activator (uPA).	Alternative exclusion/inclusion of the cassette exon 5, which partially covers domain II of the receptor. The short splice variant lacking exon 5 has prognostic relevance in breast cancer.	Implicated in cancer invasion and metastasis.
S100A 4	The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs.	Variant 1 has an alternate 5′ UTR, compared to variant 2. Both variants 1 and 2 encode the same isoform.	Involved in the regulation of several cellular processes such as cell cycle progression and differentiation.
SYK	Member of the family of non-receptor type Tyr protein kinases. Widely expressed in hematopoietic cells and involved in coupling activated immunoreceptors to downstream signaling events that mediate diverse cellular responses.	The short isoform lacks the alternative in-frame exon 7, resulting in isoform Syk (S). The longest isoform Syk (L) corresponds to the full-length transcript.	Involved in proliferation, differentiation, and phagocytosis, it is considered a modulator of epithelial cell growth and a potential tumor suppressor in human breast carcinomas.

Table 1.1 Summary of known alternate splicing events that play a role in tumour progression. The gene, the specific alternate splicing event and its role in tumour development are summarised in this Table, taken from Montiel et al 2018 [46].

There are several different mechanisms that can lead to altered splicing. One mechanism is through the acquisition of mutations within the cis-acting splicing regulatory sequences (the exonic/intronic splicing enhancers and silencers elements). Up to 50% of disease causing mutations are known to occur within these sequence elements [18]. These mutations can inhibit or alter the binding of the trans-acting splicing regulatory proteins to these sites. Table 1.2 highlights some of the known mutations in the splicing regulatory sequences that occur in different types of cancer, including how these mutations affect alternate splicing [47]. One example includes the *KIT* gene that encodes for a receptor tyrosine kinase (RTK) that is constitutively activated in gastrointestinal stromal tumours. Deletion of the region containing the 3' splice site of intron 10 results in the production of an isoform that lacks a regulatory region of the kinase required for auto-inhibition and therefore driving constitutive activation of the receptor. Another example, comes from splice variants of the *CDH17* gene, that encodes the liver intestine cadherin involved in cell-cell adhesion, isoforms that lack exon 7 through the acquisition of an exonic splicing silencer site are strongly associated with poor prognosis of liver hepatocellular carcinoma and high incidence of re-occurrence.

Gene (function)	Disease	Mutated Sequence	Molecular Consequence
KIT (oncogene)	Gastrointestinal stromal tumour	Deletion of 30 or 34 residues including the 3; splice site in intron 10	Aberrant splicing leading to constitutive protein activation
CDH17 (Liver Intestine Cadherin)	Hepatocellular carcinoma	35A>G point mutation which disrupts the branch point site in intron 6	Generation of an ISS resulting in exon 7 skipping
CDH17 (Liver Intestine Cadherin)	Hepatocellular carcinoma	Point mutation of codon 651 in exon 6	Generation of an ISS resulting in exon 7 skipping
KLF6 (tumour suppressor)	Prostate cancer	27G>A point mutation in intron 1	Generation of an intronic splicing enhancer site for the SRp40 protein leading to the production of a splice variant which acts as a dominant negative regulator
HAS1 (hyaluronan synthase)	Multiple myeloma	Exon 3 point mutation C7760T	Exon 4 skipping
BRCA1 (tumour suppressor)	Breast and ovarian cancer	Exon 18 point mutation G5199T (E1694X)	Exonic splicing enhancer disruption resulting in skipping of exon 18

Table 1.2 Summary of mutations in splicing regulatory sequences that play a role in tumour progression. Table shows the gene that is mutated including the function of that gene in brackets, the tumour that the gene is associated with, the specific mutation in the splicing regulatory sequence and how the mutations affect alternate splicing (adapted from Srebrow et al 2006 [47]).

Another mechanism that can drive changes in alternate splicing is altered expression of the trans-acting splicing regulatory proteins. Altered expression of splicing factor proteins has been observed in many different types of cancers [35, 48-51]. It is important to understand the specific splicing events that are regulated by these splicing factor proteins in the various cancers. For example, the splicing factor SF2/ASF is up regulated in lung, colon and breast cancer [5, 52, 53]. SF2/ASF induces the skipping of exon 11 in the gene encoding for the RTK RON. The alternately spliced receptor which lacks exon 11 is constitutively active and drives cell dissociation, motility and matrix invasion [54]. Table 1.3 summarises some of the splicing factors known to have altered expression in cancer and the specific alternate splicing event they regulate [5].

Several of the splicing factor proteins have been shown to have proto-oncogenic properties [5, 48]. SRSF1, a member of the SR family of splicing factor proteins was the first [52]. SRSF1 is up regulated in a number of cancers and over-expression in mice induces tumour formation [53]. Over-expression or knockdown of SRSF1 in human cell lines results in altered splicing of tumour suppressor genes and oncogenes. SRSF6 is another SR family member that is known to have oncogenic properties in both lung and colon cancers [49].

The hnRNPs have also been implicated in cancer and have been documented to have both pro- and anti-tumourigenic capacities. For instance, up regulation of hnRNP A1, A2 and E leads to increased exclusion of exon 9 in the pyruvate kinase muscle (PKM) gene which plays a role in glioma formation through promoting aerobic glycolysis [55]. Conversely hnRNPLL supresses metastasis by regulating CD44 exon 6 inclusion. In colorectal cancer, hnRNPLL expression is down-regulated [56].

Splicing factor	Downstream dysregulated isoforms that are functionally linked to cancer
ESRP1 and ESRP2	Promotes an epithelial splicing program to regulate EMT
hnRNP A1	Contributes to aerobic glycolysis in cancer by promoting the expression of specific isoforms of pyruvate kinase (PKM2 isoform) and MAX (the Delta MAX isoform)
hnRNP A2	Contributes to aerobic glycolysis in cancer by promoting the expression of a specific isoform of pyruvate kinase (PKM2 isoform). Increases breast cancer cell invasion by promoting the expression of a specific isoform of TP53INP2
hnRNP A2/B1	Acts as an oncogenic driver in glioblastoma by regulating splicing of several tumor suppressors and oncogenes, including RON
hnRNP H	Contributes to survival of gliomas as well as invasion by promoting the expression of specific isoforms of IG20 and RON, respectively
hnRNP K	Serves as a tumor suppressor in leukemia. Deletion is associated with aberrant p21 and C/EBP expression (although mechanistic links to splicing and gene expression are unclear)
hnRNP M	Contributes to EMT in breast cancer and increases metastasis in mice by promoting the expression of a specific isoform of CD44 (CD44s)
PRPF6	Promotes cell proliferation in colon cancer by altering splicing of genes associated with growth regulation, including the kinase gene ZAK
PTB (PTBP1)	Contributes to aerobic glycolysis in cancer by promoting the expression of a specific isoform of pyruvate kinase (PKM2 isoform)47,48. Also promotes the expression of an isoform of the deubiquitinating enzyme-encoding gene USP5 that has been shown to promote glioma cell growth and mobility (although the mechanism underlying this phenotypic association is not resolved).
QKI	Acts as a tumor suppressor by regulating alternative splicing of NUMB in lung cancer cells
RBFOX2	Promotes a mesenchymal splicing program to regulate EMT
RBM4	Acts as a tumor suppressor by promoting the pro-apoptotic isoform BCL-XS of BCL2L1 and opposing the pro-tumorigenic effects of SRSF1 on mTOR activation

RBM5, RBM6 and RBM10	RBM5 modulates apoptosis by regulating alternative splicing of CASP2 and FAS9. RBM5 or RMB6 depletion has an opposite effect to RBM10 depletion, as these factors antagonistically regulate the alternative splicing of NUMB
SRSF1	Promotes an isoform of the kinase MNK2 that promotes eIF4E phosphorylation independently of MAP kinase signaling. In the context of breast cancer, SRSF1 overexpression promotes alternative splicing of BIM and BIN1 to promote the expression of isoforms that lack pro-apoptotic functions.
SRSF3	Regulates alternative splicing of TP53 such that SRSF3 loss promotes expression of p53β, an isoform of p53 that promotes p53-mediated senescence.
SRSF6	Promotes expression of isoforms of the extracellular matrix protein tenascin C that are characteristic of invasive and metastatic skin cancer45, contributing to epithelial cell hyperplasia.
SRSF10	Promotes cell proliferation and colony formation in vitro and increases tumorigenic capacity of colon cancer cells in mice by inducing expression of a specific isoform of BCLAF1 (BCLAF1-L).

Table 1.3 Summary of splicing factors with altered expression in cancer. Summary of the splicing factors that are up or down regulated in different types of cancer and the splicing event that they regulate (Taken from Dvinge et al 2016 [5]).

As well as altered expression of the splicing factor proteins in cancer splicing factor genes have long been known to be rich in somatic mutations [57]. Mutations in splicing factor proteins alters their ability to bind to cis-regulatory sequences which can lead to aberrant splicing. SF3BP1 is one of the most commonly mutated splicing factor genes; it is mutated in 10-15% of cases of chronic lymphocytic leukaemia [58]. U2AF1 is another splicing factor that is frequently mutated; its mutation leads to altered splicing of the proto-oncogene *CTNNB1* in lung adenocarcinoma [59]. Mutations in the core splicing machinery are however much less common. This may be because splicing is so prevalent that mutations in the spliceosomal components are fatal to the cell.

1.3.3 Therapeutically targeting alternate splicing

The emerging role of alternate splicing in tumour development and progression makes it an attractive target for the development of novel anti-cancer therapeutics. Present strategies to target aberrant splicing are discussed below; these include targeting the specific splice variants up-regulated in cancer, to targeting the splicing regulatory proteins that regulate the splicing events [60].

1.3.3.1 Targeting oncogenic splicing events

Antisense oligonucleotides (ASOs) are short oligonucleotides, usually 15-25 bases in length that are complementary to a specific RNA transcript. They bind to the complementary mRNA and block translation [61, 62]. ASOs can be used to down regulate transcripts that are up regulated in cancer.

Splice switching oligonucleotides (SSOs) have also been developed which can shift splicing events to favour the production of isoforms with anti-tumour properties. SSOs function by hybridising to canonical splicing sites or the cis-regulatory sequences [63]. This interferes with the binding and function of the trans-acting

splicing regulatory proteins, and thus alters alternate splicing. Due to their high specificity ASOs and SSOs are a promising tool to modify splicing or RNA expression toward therapeutic purposes [64].

SSOs have been used to target *BCL-X* splicing in glioma cells [65]. The *BCL-X* gene is alternatively spliced to produce two transcripts through the use of an alternate 5' splice site: a longer transcript bcl-xl and a shorter transcript Bcl-xs. The two protein isoforms have antagonistic functions; Bcl-xl is anti-apoptotic whereas Bcl-xs promotes apoptosis; in cancer expression of the anti-apoptotic isoform prevails. An RNA oligonucleotide has been designed to hybridize and inhibit the use of the alternate 5' splice site in the Bcl-x exon 2, this restores the normal ratio of Bcl-xL and Bcl-Xs, and sensitises the cells to undergoing apoptosis in response to chemotherapeutic treatment [66].

Another example of using splice switching oligonucleotides to correct aberrant splicing is the *MDM4* gene. This gene mediates inactivation of the tumour suppressor protein p53 and is up-regulated in many cancers including melanoma [67]. The *MDM4* gene produces two transcripts: MDM4, the full-length transcript; and MDM4-S, a shorter transcript that lacks exon 6. In normal cells the shorter transcript is targeted by non-sense-mediated decay meaning there are usually low levels of the MDM4 protein [68]. However, in many cancers the MDM4 protein levels are up-regulated through increased production of the full-length transcript containing exon 6, which does not undergo non-sense mediated decay. SRSF3 is the splicing regulatory protein that enhances inclusion of exon 6. An RNA oligonucleotide was developed that binds and inhibits the SRSF3 binding site within MDM4 exon 6. This subsequently inhibited exon 6 inclusion, and therefore reduced the MDM4 protein level, which resulted in activation of p53 and

subsequent inhibition of cell proliferation and colony formation in both melanoma cell lines and in vivo mice models [69].

The use of both ASOs and SSOs therapeutically is reliant on delivery methods to introduce the oligonucleotides into the cells. Cell penetrating peptides have been used to mediate delivery including penetratin and transportan [70].

1.3.3.2 Targeting the splicing regulatory proteins

As well as targeting the specific splicing events linked to disease the trans-acting regulatory proteins can be therapeutically targeted. The SR proteins require phosphorylation in order to recognise and bind to enhancer elements in pre-mRNA, therefore one strategy to inhibit the function of the SR proteins is to target the kinases responsible for their phosphorylation. SRPK, CLK and DYRK are examples of kinases known to phosphorylate the SR family of splicing regulatory proteins. The tyrosine kinase inhibitors TG003 and SRPIN340 reduce phosphorylation of the SR proteins, which results in reduced SRSF2, Fas and VEGF transcript levels in cancer cell lines [71, 72]. Cpd compounds are another family of tyrosine kinase inhibitors that have increased specificity to the SRPK and CLK families of protein kinases [73, 74]. This therapeutic mechanism is less selective than targeting an aberrant splicing event as the SR protein kinases will phosphorylate multiple SR proteins and the SR proteins regulate splicing of multiple different genes.

1.3.3.3 Targeting the spliceosome

A number of natural products have been discovered which inhibit protein components of the spliceosome [75]. For example FR901464 is a natural product that inhibits splicing factor 3B (SF3B) with IC_{50} in the low nanomolar range in various solid tumour cell lines. This has been shown to have potent anti-

proliferative affects in several cancer cell lines and in mouse xenograft models. Spliceostatin A (SSA) is a derivative of FR901464 with anti-proliferative and anti-tumour activities through affecting splicing of genes encoding cell cycle regulators such as cyclin A2 and Aurora A kinase [76]. SSA inhibits spliceosome assembly by slowing the transition from the A complex to the B complex. Isoginketin is another natural compound isolated from dried leaves of Gingko Biloba which interferes with spliceosome assembly by stabilising the U4/U5/U6 trimer also preventing the transition from complex A to B [77].

1.3.3.4 Diagnostic and prognostic potential of alternate splicing

In addition to oncogenic alternate splicing events being potential targets for the development of novel anti-cancer therapies, they can also be used in the diagnosis and prognosis of cancer [78, 79]. For example, unique splicing patterns have been shown to distinguish between normal and tumour tissue in 9 different types of cancer [80]. Additionally alternate splicing signatures can differentiate between different tumour subtypes, for example triple negative, non-triple negative and HER2 positive breast cancers all have unique splicing patterns. Furthermore, alternatively spliced transcripts can serve as prognostic indicators. For example, in colorectal cancer low expression of CD44*v6* isoform that lacks exon 6 is associated with improved survival, whereas high levels of exon 6 inclusion are associated with worse survival. In addition, CD44 splicing may also be predictive of response to anti-CD44 therapy in several solid tumours [81].

1.4 Growth factor receptor bound protein 2

Growth factor receptor bound protein 2 (Grb2) is a small 25 kDa adaptor protein involved in signal transduction. It was first discovered in 1992 through a screen for proteins that bound to the tyrosine phosphorylated carboxy-terminus of the epidermal growth factor receptor (EGFR) [82]. It is widely expressed and essential to multiple cellular functions including cell survival, proliferation and differentiation. No known naturally occurring mutations in the human *GRB2* gene have been reported suggesting that its function is critical for survival. Inhibition of Grb2 results in developmental defects but has also been shown to block cellular transformation in various cell types.

1.4.1 Alternate splicing of *GRB2*

The *GRB2* gene is located on chromosome 17 in humans and is homologous to the Sem-5 gene in *C. elegans* and the Drk gene in *D. melanogaster* [83]. It is composed of 6 exons, 5 coding and 1 non-coding as depicted in Figure 1.8. Exons 2-6 code for the 217-residue adaptor protein Grb2 which consists of a central SH2 domain flanked either side by SH3 domains. This structure is reminiscent of the proto-oncogene Crk. Exons 2 and 3 code for the 58 residue N-terminal SH3 domain, exon 4 and 5 code for residues 60-158 which form the central SH2 domain and exon 6 codes for residues 164-217 which form the C-terminal SH3 domain. These domains are capable of binding to other proteins, but do not have any intrinsic enzymatic activity. Grb2 therefore contributes to signal transduction through mediating the formation of signalling complexes [84].

The SH2 domain of Grb2 binds phosphotyrosine residues through a conserved positively charged phosphotyrosine binding pocket which forms electrostatic interactions with the negatively charged phosphate group. A secondary binding

site within the SH2 domain interacts with residues immediately C-terminal to the phosphotyrosine residue which allows modest substrate selectivity amongst different SH2 domain-containing proteins [85]. The Grb2 SH2 domain has a preference for binding to the consensus sequence pYXNX, where N represents asparagine and X can be any amino acid, usually hydrophobic [86]. This consensus sequence is found at the C-terminus of many receptor tyrosine kinases (RTKs) including EGFR, platelet derived growth factor receptor (PDGFR) and hepatocyte growth factor receptor Met, as well as non-RTKs such as focal adhesion kinase (FAK) and insulin receptor substrate 1 (IRS1). The N and C terminal SH3 domains in Grb2 bind to proline-rich sequences, with a preference for the sequence P-X-X-P-X-R where X can be any amino acid, R is arginine and P represents proline [87]. However, Grb2 is also able to bind to proteins that lack this consensus sequence. For example, the C terminal SH3 domain of Grb2 has been reported to bind Gab1 and SLP-76 through recognition of a the non-canonical motif: P-X-X-X-R-X-X-K-P [88].

Alternate splicing of *GRB2* produces a variant that lacks exon 4 (Figure 1.8). The splice variant encodes a 176 residue protein called Grb3-3 which contains a 40 residue deletion in the central SH2 domain. Grb3-3 is widely expressed amongst different tissues. It is usually at 10-50x lower levels than Grb2 although under certain conditions such as negative selection of T cells it is expressed at equal levels to Grb2. The truncated SH2 domain in Grb3-3, referred to as $SH2\Delta_{40}$ is non-functional and is unable to bind phosphotyrosine residues such as those found at the C-terminus of EGFR. The two SH3 domains are still able to recognise proline-rich motifs [89].

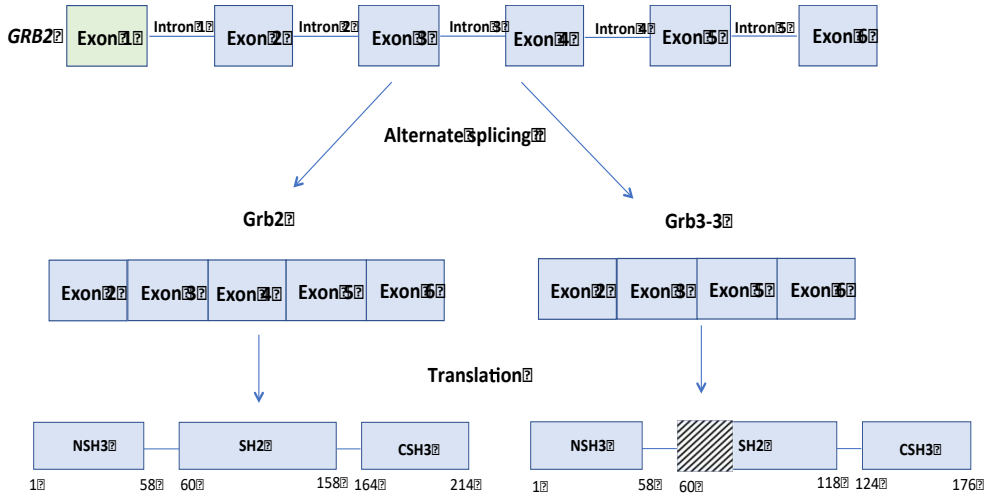


Figure 1.8 Alternate splicing of *GRB2*. Row 1 shows the *GRB2* gene is composed of 6 exons. Protein coding exons are shown as blue rectangles, non-protein coding exons are shown as green rectangles, the exons are separated by non-protein coding introns which are shown as blue lines. Row 2 shows the two splice variants produced by alternate splicing of *GRB2* which differ only by the inclusion of exon 4. Row 3 shows the polypeptides encoded by these splice variants, the region of the SH2 domain in Grb3-3 shown as a black and white diagonal box represents the first 40 residues that are deleted in this isoform.

1.4.2 The role of Grb2 in Ras-Erk signalling

The most well characterised role of the full-length adaptor protein Grb2 is in linking activated RTKs to the Ras-Erk signalling cascade (Figure 1.9) [90]. The Ras-Erk signalling pathway is perhaps the best-characterised signal transduction pathway owing largely to its prominent role in cancer. It functions to transduce signals from the extracellular environment to the nucleus resulting in the induction of the expression of specific genes that control cell proliferation, cell survival, cell differentiation and apoptosis [91].

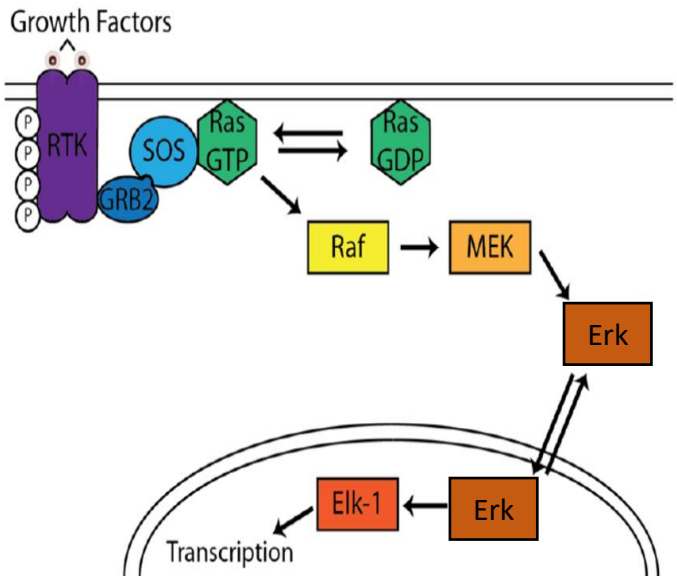


Figure 1.9 The Ras-Erk signalling cascade. Schematic representation of the Ras-Erk signalling cascade, signalling molecules are shown as coloured shapes, arrows show the direction of activation, double arrows represent reversible reactions.

The Ras-Erk signalling cascade is activated downstream of RTKs in response to growth factors [92]. Humans possess 58 RTKs categorised into twenty subfamilies [93]. All RTKs consist of an extracellular ligand binding region, a transmembrane domain, an intracellular kinase domain and a cytoplasmic tail. The majority of RTKs exist at the plasma membrane as monomers, with the exception of the insulin receptor and fibroblast growth factor receptor 2 (FGFR2) which form dimers in the absence of ligand [94]. Upon engagement of the extracellular domain with their specific ligand the receptors undergo homo- or hetero-dimerization. Dimerization up-regulates the intrinsic kinase activity of the receptor which induces trans-auto-phosphorylation of tyrosine residues in the cytosolic region. These phosphotyrosine residues form the docking sites for cytosolic proteins that contain phosphotyrosine binding (PTB) domains or Src homology 2 (SH2) domains. These proteins function to transduce the signal from the activated receptor by mediating the formation of signalling complexes [92].

Grb2 is the major adaptor protein that transduces the signal from activated RTKs to the Ras-Erk signalling cascade [95]. Grb2 interacts with the activated receptors via its SH2 domain and is able to interact with other signalling proteins via its two SH3 domains [90, 96-98]. The N-terminal SH3 domain of Grb2 interacts with the Ras guanine nucleotide exchange factor Son of sevenless (Sos) [99]. The interaction between Grb2 and Sos has been extensively characterised; the N terminal SH3 domain but not the C-terminal SH3 domain interacts with multiple proline-rich motifs in the last 150 residues of the C-terminal tail of Sos [100-102]. This interaction was generally viewed to be constitutive however has now been shown to be inducible by growth factors such as EGF. In the absence of growth factors Grb2 exists as a concentration dependent dimer and the dimeric form of Grb2 has a lower affinity for Sos. Upon growth factor stimulation Grb2 becomes

phosphorylated at tyrosine 160 (Y160) by the activated RTKs. Y160 lies at the dimer interface and phosphorylation induces Grb2 monomerization. The monomeric form is then able to bind to Sos resulting in the formation of a trimeric RTK-Grb2-Sos complex at the plasma membrane. The formation of this complex results in the recruitment of Sos to the cell surface which brings it into close proximity of its membrane anchored substrate Ras [103].

The Sos proteins belong to the larger family of Ras guanine exchange factors (GEFs), which facilitate Ras activation by promoting the dissociation of GDP to allow association of the more abundant GTP which activates Ras [104, 105]. Ras GEFs are required to mediate GDP dissociation because the interaction between Ras and GDP is extremely high affinity; occuring in the low nanonmolar/picomolar range such that in the absence of a catalyst the rate of GDP dissociation would be extremely slow with a half-life of several hours [106, 107]. The Sos proteins are large proteins composed of multiple domains including HD, DH and PH domains that are found at the N-terminus which is connected by a helical linker to the catalytic domain composed of REM and CDC25 domains followed by a C-terminal proline-rich tail (Figure 1.10) [104]. They interact with the Ras proteins through their catalytic domain, which has two Ras binding sites, an allosteric site and a catalytic site [108]. GDP bound (inactive) Ras binds in the catalytic site of Sos, inducing a conformational change in the nucleotide binding region (switch 2) of Ras which reduces the affinity of nucleotide binding resulting in GDP release [105, 106]. As GTP is present in the cell at 10-fold higher concentrations over GDP, Ras becomes GTP bound. GTP Ras is then able to interact with the allosteric site of Sos which results in a 15 fold increase in affinity for binding of GDP Ras to the catalytic site and an 8-fold increase in the rate of Sos catalysed nucleotide exchange, thus creating a positive feedback loop enhancing Sos-mediated Ras

activation at the plasma membrane [109, 110]. GTP bound Ras then undergoes dimerization, which is thought to be mediated by the C-terminus of Ras. GTP Ras dimers recruit Raf to the membrane activating Raf and initiating the Raf-Mek-Erk signalling cascade (Figure 1.9) [111, 112]. Once activated, Erk translocates into the nucleus and induces the expression of genes involved in cell cycle progression, survival and differentiation [113].

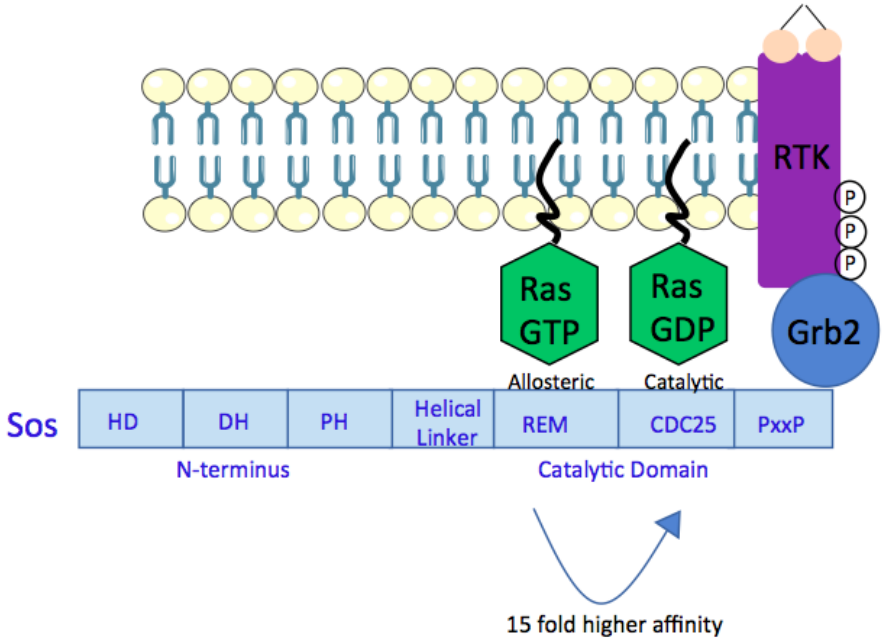


Figure 1.10 Schematic representation of Sos mediated Ras activation. Grb2 (represented as a blue circle) binds to the activated RTK shown in purple and the proline-rich c-terminus of Sos, bringing it in close proximity of Ras. A single Sos protein is able to bind two Ras proteins; GDP Ras binds at the catalytic site where it gets converted to GTP Ras. GTP Ras can then bind at the allosteric site which results in a 15-fold increase in affinity for GDP Ras at the catalytic site.

1.4.3 The role of Grb3-3 in Ras-Erk signalling

The role of the truncated adaptor protein Grb3-3 in the Ras-Erk signalling cascade is less well characterised. Initial studies in chinease hamster lung fibroblasts cells that overexpress human EGFR (ER22 cells) suggest that it may have antagonistic functions to Grb2 in this signalling pathway. Overexpression of Grb3-3 in ER22 cells inhibits EGF dependent transactivation of a Ras responsive element. This inhibition can be rescued by overexpression of Grb2. However, there have been no studies on the role of Grb3-3 in Ras-Erk signalling conducted in human cells. In vitro experiments show Grb3-3 is able to bind to a proline-rich peptide derived from human Sos1 and that the interaction is mediated by the SH3 domains since mutation of a glycine residue at position 162 to arginine (G162R) abolishes this interaction. Grb3-3, however is unable to bind to tyrosine phosphorylated EGFR due to the truncation in the SH2 domain. The Grb3-3 mutant G162R that is unable to bind to Sos, does not affect EGF induced expression of Ras responsive genes in ER22 cells suggesting that its ability to bind to Sos is required for the inhibitory affect. The significance of *GRB2* splicing in regulating the Ras-Erk signalling cascade is unknown and a detailed understanding of the role of Grb3-3 in this pathway in human cells is lacking [89].

1.4.4 The Ras-Erk signalling cascade in cancer

The Ras-Erk signalling pathway is the most commonly deregulated signalling pathway in cancer; it is up regulated in approximately one-third of all human cancers including colorectal, breast and liver cancer [114-117]. Most cancer associated lesions in Erk signalling affect the early steps of the pathway including overexpression or activating mutations in the RTKs, sustained autocrine or paracrine production of growth factors or mutations in the Ras and Raf genes.

The fact that deregulation of Erk signalling occurs at multiple levels highlights its importance for tumourgenesis.

There is ample evidence to suggest that Grb2 is an essential mediator of oncogenic Ras-Erk signalling. For example, gastric cancer, one of the most common malignancies worldwide is driven by Grb2 mediated human epidermal growth factor receptor 2 (HER2) signalling. Grb2 is overexpressed in 48% (553/1,143) of primary gastric tumours and 59% (155/261) of lymph node metastases. Overexpression of Grb2 is associated with worse overall survival [118]. Overexpression of Grb2 and Sos has also been suggested to be sufficient to up regulate Ras-Erk signalling in bladder cancer cell lines [119]. Grb2 also plays a role in Leukaemias positive for the Philadelphia chromosome through the interaction with the BCR-ABL oncoprotein. Grb2 binds to the BCR-ABL kinase through its SH2 domain binding to a phosphorylated tyrosine at position 177 in the B-cell receptor (BCR). This interaction is required for activation of the Ras signalling pathway and transformation of primary bone marrow cultures. Mutation of Y177 abolishes Grb2 binding and inhibits the transforming capabilities of BCR-ABL [120]. Finally up regulation of the signalling active monomeric form of Grb2 that binds to Sos has been highlighted in prostate and colorectal cancer [103].

1.4.5 Grb2 inhibitors

Given the prevalence of Ras-Erk signalling in cancer pharmacological inhibition is considered to be the “Holy Grail” of cancer therapy. The critical role of Grb2 in linking upstream receptor tyrosine kinases to the Ras signalling pathway makes it an attractive therapeutic target [121]. Loss-of-function mutations in *D. melanogaster* and *C. elegan* homologues of Grb2 inhibit Ras signalling

downstream of the EGFR homologues suggesting that the selective inhibition of Grb2 is a viable strategy to inhibit oncogenic Ras signalling. Several mechanisms have been explored to inhibit Grb2 that include down regulation of Grb2 expression, or selective inhibition of the SH2 and SH3 domains, discussed in more detail below. However, so far no Grb2 inhibitors have advanced into the clinic.

Anti-sense oligonucleotides complementary to Grb2 mRNA have been developed, which target it for degradation. Liposome mediated delivery of these anti-sense oligos to breast cancer cells with high EGFR expression down regulates Grb2 expression resulting in reduced Erk activation and cell proliferation in a dose dependent manner [122]. Down regulation of Grb2 using anti-sense oligonucleotides has also been shown to have promising anti-cancer properties in the treatment of leukaemia [123]. However, an existing challenge with the adoption of anti-sense oligonucleotides in clinical practice remains their poor bioavailability due to degradation by cellular nucleases.

The interaction between Grb2 and Sos mediated by the N-terminal SH3 domain is critical to its role in Ras-Erk signalling. Dissociation of the Grb2-Sos complex through Erk mediated phosphorylation of Sos functions as a negative regulatory mechanism in cells to switch of Erk signalling [124-126]. Therefore an alternate therapeutic strategy to antisense oligos is to inhibit Grb2-Sos complex formation. Yu et al [127] developed proline-rich reactive peptides that covalently bind to the N-terminal SH3 domain of Grb2 close to its Sos binding site. These peptides inhibited migration of SKBR3 breast cancer cells and partially inhibited Erk activation, presumably through preventing Sos binding, although this effect has not been directly shown [127].

Peptide inhibitors have also been developed against the Grb2 SH2 domain. The peptides mimic the consensus binding sequence for Grb2 SH2 domain found in its substrate proteins including activated RTKs. Binding of the peptides to the Grb2 SH2 domain competitively inhibits phosphotyrosine-binding [128]. The minimal peptide sequence required for Grb2 SH2 domain binding is pYXN (where pY is phosphotyrosine, X is any amino acid and N is asparagine), two peptide mimetics CGP78850 and CGP85793 have been developed based on this minimal peptide region shown in Figure 1.11. CGP78850 binds with 0.04 μM affinity to Grb2 SH2 domain which is over 1000 fold higher than its affinity for other SH2 domain containing proteins such as Fyn, Shp2 and Shc, demonstrating selectivity for Grb2. CGP85793 is a pro-drug of CGP78850 which has increased cell penetration and efficacy. Both phosphopeptides reduce EGFR-Grb2 complex formation and Ras activation in breast cancer cells which results in inhibition of anchorage independent cell growth [129, 130]. Another phospho tyrosine mimetic is C90, which inhibits growth factor induced cell motility, invasion and angiogenesis in cell culture and in animal models [131, 132].

Although the Grb2 inhibitors described above have not yet been advanced into clinical practice they clearly demonstrate that inhibition of Grb2 is an effective strategy to down-regulate Ras activation and Erk signalling. Given that this pathway is hyper-activated in a wide array of cancers, the selective inhibition of Grb2 could be used for the treatment of many cancers.

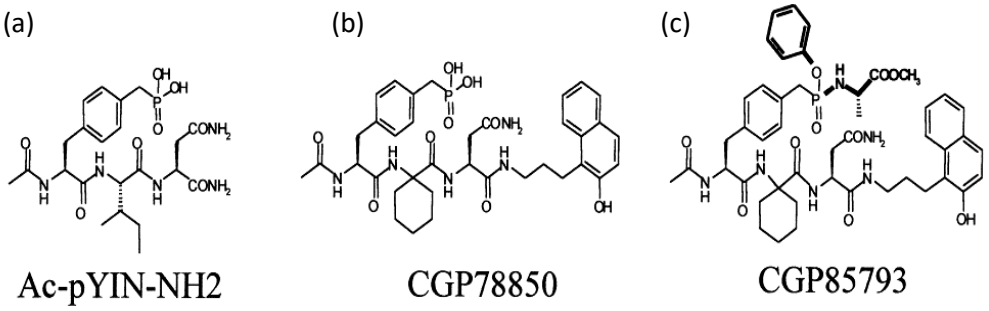


Figure 1.11 Peptide inhibitors of Grb2 SH2 domain. Chemical structure of (a) the minimal phosphotyrosine sequence recognised by Grb2 SH2 domain (b and c) phospho-peptide mimetics designed based on these sequence. Figure taken from [129].

1.5 Colorectal Cancer

Colorectal cancer (CRC) is one of the most common malignancies worldwide, with 1 million new cases per year. It is the second leading cause of cancer-related death in women and third in men. In 2016 in the UK alone there were 16,384 deaths from CRC, accounting for 10% of all cancer-related deaths in the UK [133]. The incidence of CRC increases significantly with age, with almost 9 out of 10 cases occur in people aged over 60. CRC is uncommon before the age of 40; age-specific incidence rates rise steeply from around the age range 50-54, and are highest between the ages of 85 and 89. Several environmental factors influence the risk of CRC. Diets high in fat, in particular animal fat, are major risk factors. Smoking and regular alcohol consumption are also associated with increased risk. The incidence of CRC varies geographically, with highest incidence rates in Australia and New Zealand [134].

1.5.1 The anatomy of the colon

Colorectal cancers develop in the colon or rectum which, form part of the lower digestive system. The colon (also known as the large intestine) is about 1.5 metres long and is composed of four sections; the ascending colon (found on the right side of the abdomen), the transverse colon, the descending colon (found of the left side of the abdomen) and finally the sigmoid colon, which leads to the rectum and then anus. The majority of colorectal cancers occur in the sigmoid colon and the rectum.

The colon forms a hollow tube, inside this tube is referred to as the lumen, surrounding the lumen are four tissue layers as shown in Figure 1.12. The innermost layer known as the mucosa is lined with tube like structures known as colonic crypts. Surrounding the mucosa is a layer of blood vessels, nerves and

connective tissue that form the submucosa. The muscularis layer surrounds the submucosa and is composed largely of smooth muscle cells that contract to move the large intestine. The outermost layer of the membrane is known as the serosa, which is a thick layer of epithelial tissue.

The colonic crypts that line the mucosa are highly organised monolayers of self-renewing epithelial cells. At or near the base of the crypt is a stem cell population and their progeny are highly proliferating cells that occupy the lower third of the crypt. As the cells move up the crypt they differentiate forming mature epithelial cells. Colorectal cancers originate in these colonic crypts and then invade the subsequent tissue layers of the bowel wall.

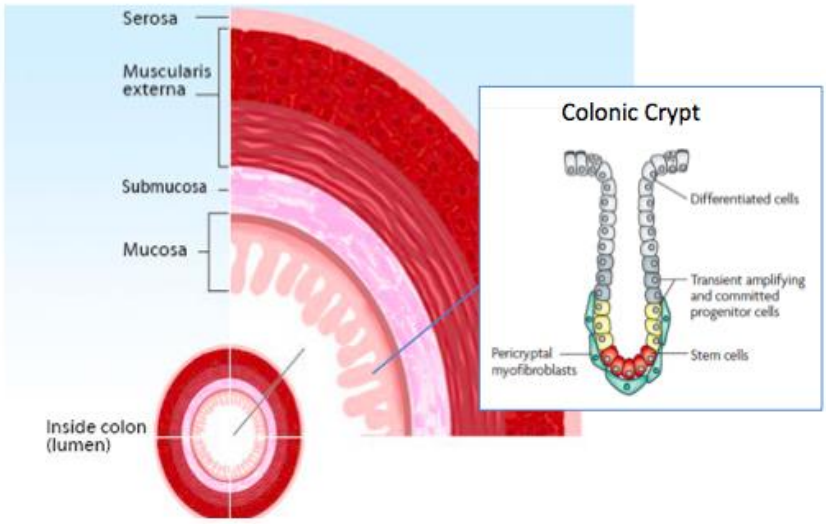


Figure 1.12 The anatomy of the colon. The colon forms a hollow tube the inside of which is known as the lumen. The membrane surrounding the lumen is magnified showing the four tissue layers. Further magnification of a single colonic crypt from the mucosal layer is shown. The differentiated cells at the top of the crypt are shown in grey; highly proliferating cells in the lower third of the crypt are shown in yellow and the stem cell population in red.

1.5.2 Oncogenic Ras signalling in colorectal cancer

Hyper-activation of the Ras-Erk signalling cascade (described in Section 1.4.4) is a key step in colorectal cancer development [135-137]. Elevated Ras signalling in colorectal cancer can occur through different mechanisms. One mechanism is through increased activation of EGFR [138-142]. This RTK lies upstream of Ras and is activated by several ligands, including the epidermal growth factor (EGF), tumour growth factor (TGF) alpha, amphiregulin and epiregulin [143]. In colorectal cancer expression of both EGFR and its ligands are up-regulated [144]. The EGFR gene is found on chromosome 7 and increased copy numbers of this gene occur through chromosome 7 polysomy in colorectal cancer [145]. Over-expression of EGFR is associated with a poor prognosis. Additionally, activating mutations in the tyrosine kinase domain of EGFR occur in a subset of colorectal cancer patients [146]. The tyrosine kinase domain is encoded by exons 18-23 of the *EGFR* gene and mutations in these exons result in constitutive kinase activity and also resistance to targeted EGFR inhibitors (discussed later). In a study involving 58 patients, 13 (22.4%) possessed point mutations in exon 20 of the *EGFR* gene. Mutations in EGFR exon 20 were most common in patients at an earlier tumour stage, and without lymph node metastasis. Other EGFR mutations that have been identified in colorectal cancer include E749K, E726G and A767T which are found in exon 19 and 20 [146].

The Ras proteins downstream of EGFR are also hyperactive in CRC. In 35-45% of patients this occurs through the presence of activating mutations [147-149]. Humans have three Ras genes, H, N and K Ras, mutations in these genes occur with different frequencies in CRC. Mutations in K-Ras are most common, 80% of K-Ras mutations in CRC affect codon 12 and 20% affect codon 13. These mutations impair hydrolysis of GTP rendering Ras in a constitutively active GTP

bound state. The point mutation from glycine at position 12 to aspartic acid (G12D) is the most common amino acid change. K-Ras mutations have been detected in patients with early stage disease, suggesting this may be an early event during tumorigenesis. Mutations in K-Ras are associated with a worse prognosis and causes resistance to EGFR inhibitors [140, 150]. Much lower rates of N-Ras mutations have been seen (up to 3% of cases). The most common N-Ras mutations are G12D in exon 2 and Q61K in exon3. N-Ras mutations are more common in female patients than male patients, 75% and 25% respectively. So far no activating mutations in H-Ras have been observed in CRC [149].

Finally, mutations in the *RAF* genes also occur in colorectal cancer, although these are less common than *EGFR* and *RAS* mutations. Humans possess three *RAF* genes, *ARAF*, *BRAF* and *CRAF*. Approximately 10% of CRC patients have *BRAF* mutations [151]. The Raf proteins are downstream of Ras in the Erk signalling cascade, and mutations in Ras and Raf are mutually exclusive. In the absence of stimulation the Raf proteins exist in an auto-inhibited state where the N-terminal domain binds and inhibits the catalytic domain. Activation of Raf requires release of this auto-inhibition and phosphorylation of serine, threonine or tyrosine residues within the catalytic domain [111, 152-156]. *BRAF* mutations facilitate activation via two mechanisms; by mimicking phosphorylation of activatory sites within Raf through mutations to negatively charged residues, and by releasing auto-inhibition by the N-terminal domain through truncations in this domains. The most common BRAF mutation in CRC (80% of all *BRAF* mutations) is valine to glutamic acid at residue 600 (V600E). V600 lies within the activation segment of the kinase domain in exon 15, mutation to glutamic acid increases the kinase activity of B-Raf.

1.5.3 Therapeutically targeting Ras-signalling in CRC

Given the clear role of the Ras-Erk signalling pathway in colorectal cancer targeted inhibition of this pathway represents a viable therapeutic strategy for the treatment of colorectal cancer. Two monoclonal antibodies cetuximab and panitumumab have been developed that inhibit EGFR induced activation of the Ras-Erk signalling cascade [144]. Cetuximab is a chimeric monoclonal antibody whereas panitumumab is fully humanised and induces less of an immune response. Both antibodies bind to the extracellular domain of the receptor and block ligand binding and downstream signalling. These therapies however, are only suitable for a sub-population of patients that lack mutations in *EGFR, RAS* and *RAF* genes. Mutations in these genes lead to intrinsic resistance to treatment with anti-EGFR antibodies. The response to anti-EGFR therapies in the Ras-wild type patient cohort is around 35%.

Variable responses have been documented for different clinical trials. For example in the CRYSTAL trial both progression-free survival (PFS) and overall survival (OS) were increased in patients treated with cytotoxic chemotherapy in combination with cetuximab versus those managed with chemotherapy alone (PFS 9.9 months vs 8.4 months, OS 23.5 months vs 20 months respectively). However in the COIN and NODIX VII trial there was no evidence that cetuximab added a significant benefit to chemotherapy alone. Furthermore in some studies the use of anti-EGFR therapy has been reported to contribute to the acquisition of Ras mutations. In one analysis of 10 wild type Ras patients treated with anti-EGFR therapy, six acquired Ras mutations following treatment leading to resistance [157].

Another class of EGFR inhibitors have been developed that target the tyrosine kinase domain as oppose to the extracellular domain. These small molecule tyrosine kinase inhibitors include gefitinib and erlotinib. Whilst known to be beneficial in the treatment of other cancers such as non-small cell lung cancer, the use of these inhibitors in CRC induced severe side effects, such as life threatening neutropenia, and patients had to stop treatment due to toxicity, therefore their benefit in treatment of colorectal cancer is yet to be established.

Anti-EGFR therapies represent a major step forward in the treatment of metastatic colorectal cancer. However, their success only occurs in a restricted population of patients and even patients that initially respond to treatment eventually gain resistance [157]. Therefore, there is still a need for the development of novel therapeutic strategies that target Ras-Erk signalling in colorectal cancer. Despite ample evidence for the role of Grb2 in mediating oncogenic Ras-Erk signalling in various cancers (discussed in section 1.4.4) and several reports that inhibition of Grb2 down regulates pathway activation (discussed in section 1.4.5) the role of Grb2 in colorectal cancer development has not yet been investigated. Inhibition of Grb2 may be a novel therapeutic strategy for CRC.

1.5.4 Alternate splicing in colorectal cancer

Alternate splicing is also emerging as a key step in colorectal cancer development [40, 158]. Several genome wide analyses have identified changes in alternate splicing in colorectal tumour tissue relative to the adjacent normal tissue. Over 9000 genes show altered alternate splicing patterns and exon skipping is the most commonly altered splicing event. Additionally changes in alternate splicing between the primary tumour and distant metastases have been observed

suggesting that alternate splicing also plays a role in colorectal cancer progression. Characterisation of the differentially spliced genes revealed that many coded for actin binding and Ras GTPase binding proteins. The major biological processes affected by alternate splicing in colorectal cancer are muscle development, organ development, cytoskeleton and actin organisation [159]. Vladimirovna et al [39] also identified 7 genes related to energy metabolism (OGDH, COL6A3, ICAM1, PHPT1,PPP2R5D, SLC29A1, and TRIB3) whose splicing was altered in colorectal cancer. Changes in metabolism occur across many different types of cancer in order to increase aerobic glycolysis under hypoxic conditions. Additionally several alternate splicing events that regulate the Notch signalling pathway are altered in CRC such as Rac1 exon 3b inclusion, NUMB exon 9 inclusion and PKM exon 10 inclusion which mediate up regulation of Notch signalling [158].

Furthermore altered expression of the trans acting splicing factors has also been reported in CRC. SRSF6 is frequently up regulated in CRC and is associated with poor prognosis. SRSF6 regulates alternate splicing of ZO1 exon 23 and enhances the production of the isoform that lacks exon 23; this isoform promotes migration and invasion in cells and animal models. [49]. hnRNPLL has also been identified as a metastasis suppressor in CRC through regulating CD44 exon 6 splicing, hnRNPLL expression is down regulated during the epithelial mesenchymal transition in CRC [56].

1.6 Hypothesis and Aims

The aim of this project is to investigate the role of *GRB2* splicing in colorectal cancer. Colorectal cancer is a major cause of cancer related deaths and is known to be driven, at least in part, by increased EGFR induced activation of the Ras-Erk signalling cascade and by aberrant splicing. Grb2 is the major adaptor protein responsible for linking the epidermal growth factor receptor to the Ras-Erk signalling pathway and targeted Grb2 inhibitors have been shown to down regulate Ras activation and signalling. Alternate splicing of *GRB2* produces a truncated isoform that has been reported to inhibit EGF induced expression of Ras responsive genes, and therefore may function as a naturally occurring antagonist of Grb2 in cells. Despite this the significance of this splicing event in CRC development has not been investigated. Therefore, the overall aim of this thesis is to understand the role of *GRB2* splicing in Ras activation in colorectal cancer. This aim has been subdivided into individual aims that are detailed below:

Aim 1: To develop a splice-sensitive PCR assay to investigate *GRB2* splicing in colorectal cancer.

Aim 2: To determine the role of the Grb2 isoform Grb3-3 in the Ras-Erk signalling cascade in human cell lines.

Aim 3: To identify the splicing factors regulating this splicing event and determine their role in colorectal cancer development.

Chapter 2: Materials and Methods

2.1 Molecular Biology: DNA manipulation

2.1.1 Molecular Cloning

The recombinant plasmids used in this study (listed in Table 2.1 and 2.2) for bacterial and mammalian protein expression were produced via molecular cloning. Primers were designed to amplify the open reading frame of interest using PCR. Restriction sites were added to the 5' and 3' end of the forwards and reverse primers respectively. PCR products and the plasmid vector were digested with the appropriate restriction enzymes and analysed by agarose gel electrophoresis. The digested products were gel-extracted and used to prepare DNA ligations, which were transformed into DH5 alpha competent cells for DNA production.

Plasmid	Use	Source
pET28a	His-tagged bacterial expression vector	Novagen
pET28a Grb2	Bacterial expression of full length His-tagged Grb2 & to produce the standard curve for the splice sensitive PCR assay	Dr Chi-Chuan Lin
pET28a Grb3-3	Bacterial expression of full length His-tagged Grb3-3 & to produce the standard curve for the splice sensitive PCR assay	This study
pET30-2-GAPDH	To produce the standard curve for the splice sensitive PCR assay	Addgene
pET28a Grb3-3 NSH3SH2Δ_{40}	Bacterial expression of truncated His-tagged Grb3-3	This study
pET28a Grb3-3 SH2Δ_{40}	Bacterial expression of truncated His-tagged Grb3-3	This study
pET28a Grb3-3 SH2Δ_{40}CSH3	Bacterial expression of truncated His-tagged Grb3-3	This study
pET28a Soscat	Bacterial expression of His-tagged Sos catalytic domain residues 564-1059	Professor Alex Breeze
pMAL-c5x	Bacterial expression of maltose binding protein tagged proteins	NEB
MBP Grb3-3 NSH3SH2Δ_{40}	Bacterial expression of truncated MBP-tagged Grb3-3, offers superior expression over pET28b constructs	This study
MBP Grb3-3 SH2Δ_{40}	Bacterial expression of truncated MBP-tagged Grb3-3, offers superior expression over pET28b constructs	This study

MBP Grb3-3 SH2Δ_{40}CSH3	Bacterial expression of truncated MBP-tagged Grb3-3, offers superior expression over pET28b constructs	This study
pGEX2T	Bacterial expression of Glutathione-S-transferase (GST)	GE
pGEX4T Sos1-8	Bacterial expression of GST fused Sos proline-rich peptides	This study

Table 2.1 Bacterial expression plasmids. Bacterial plasmids utilized in this study for recombinant protein expression and to produce standard curves for splice sensitive PCR.

Plasmid	Use	Source
pcDNA3.1	Mammalian expression of red fluorescent protein (RFP)	Dr Chi Chuan Lin
pcDNA3.1 Grb3-3	Mammalian expression of RFP fused Grb3-3	This study
pcDNA3.1 Grb2	Mammalian expression of RFP fused Grb2	This Study
pcDNA3.1 Grb2 R82L	Mammalian expression of RFP fused R82L mutant Grb2	Amy Stainthorpe
pcDNA3.1 Grb3-3 NSH3SH2Δ_{40}	Mammalian expression of RFP fused truncated Grb3-3	This study
pcDNA3.1 Grb3-3 SH2Δ_{40}	Mammalian expression of RFP fused truncated Grb3-3	This study
pcDNA3.1 Grb3-3 SH2Δ_{40}CSH3	Mammalian expression of RFP fused truncated Grb3-3	This study
EGFP N2	Mammalian expression of green fluorescent protein	ClonTech
EGFP N2 Sos	Mammalian expression of GFP fused Sos1	Dr Anika Schuller
pcDNA-Myc-Strep	Mammalian expression of strep vector	Dr Chi Chuan Lin
pcDNA-Myc-Strep Grb3-3	Mammalian expression of strep tagged Grb3-3	This study
pcDNA-Myc-Strep Grb2	Mammalian expression of strep tagged Grb2	This study
pcDNA3.1 hnRNPC Myc	Mammalian expression of myc tagged hnRNPC	Addgene

Table 2.2 Mammalian expression plasmids. Mammalian plasmids utilized in this study for exogenous expression of proteins in cell lines.

2.1.2 TA Cloning

TA cloning was used to enable sequencing of the PCR products generated in the splice sensitive PCR assay. This was done using Life Technologies TA cloning kit (K2070-20). Unlike molecular cloning, TA cloning does not require restriction digestion of the PCR products because the DNA polymerase used during PCR generates a product with 5' and 3' adenine overhangs. The PCR product is ligated into PCR-II vector, which also has 5', and 3' thymine overhangs allowing recombination to occur as shown schematically in figure 2.1. The ligation reaction was set up in a total volume of 10 µl using the reagents shown in table 2.3 and incubated at room temperature for 15 minutes before transforming into one shot Topo10 cells (see Section 2.1.9 for detailed transformation protocol) and plated onto LB agar plates containing 50 mg/ml Kanamycin, 40 mg/ml X-Gal and 100 mM IPTG. The PCR-II plasmid contains a LacZ gene, which codes for β-galactosidase this enzyme cleaves X-Gal to produce a blue product resulting in the formation of blue colonies. Successful TA cloning disrupts the LacZ gene preventing cleavage of X-Gal and therefore the cells remain white. White colonies were picked and used to produce bacterial cultures for DNA isolation and sequencing (see Sections 2.1.10-2.1.13 for more detail).

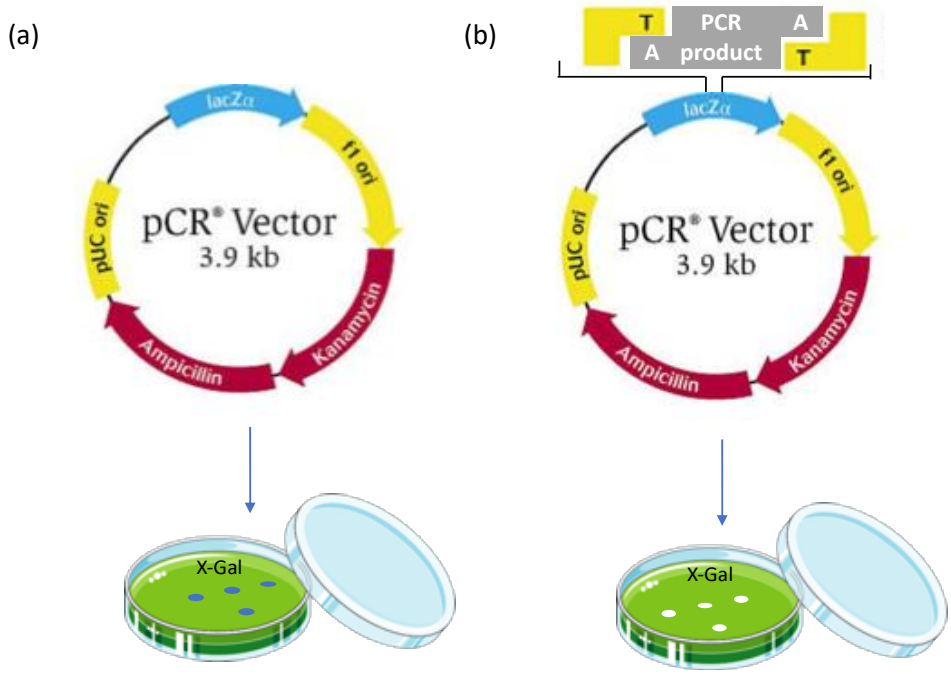


Figure 2.1 Schematic representation of TA cloning. PCR-II vector map is shown, including antibiotic resistance genes in red and LacZ gene in blue, in (a) the laz-Z gene encodes beta galactosidase resulting in the production of blue colonies when grown in the presence of x-gal in (b) successful insertion of the PCR product disrupts the LacZ gene resulting in the production of white colonies. Figure adapted from Life Technologies.

Reagent	Volume
Fresh PCR product	1 µl
5X Express link T4 DNA ligase buffer (final concentration 1X)	2 µl
PCR II vector (50ng)	2 µl
Express Link T4 DNA ligase (5 units)	1 µl
Nuclease-free Water	4 µl

Table 2.3 TA Ligation reaction set up. Components required to set up the ligation reaction during TA cloning of PCR products into PCR II vector.

2.1.3 Polymerase chain reaction (PCR)

Polymerase chain reaction was used to amplify genes of interest. Oligonucleotide primers were synthesised by integrated DNA technologies (IDT) re-suspended in nuclease free H_20 and stored at -20°C throughout the study. PCR amplification was performed in a 20 µl reaction volume; the reaction mixture for which is shown in Table 2.4. PCR samples were ran alongside negative controls which lacked the template DNA in a thermocycler, PCR conditions are shown in table 2.4

Reagent	1x Reaction Volume
Vent Polymerase	0.2 µl
10 µM dNTP's mix	1 µl
10x DNA polymerase buffer	Final 1x concentration
Forwards Primer	1 µM
Reverse Primer	1 µM
Template DNA	10 ng
Nuclease free H_20	to a total volume of 20 µl

Table 2.4 Polymerase chain reaction mixture set up. Components and volumes required to set up PCR for molecular cloning

Reaction Step	Temperature	Duration
Initial template denaturation (1 cycle)	95°C	5 minutes
Denaturation (35 cycles)	95°C	1 minute
Annealing step (35 cycles)	Temperature varied dependent on primers	1 minute
Elongation (35 cycles)	72°C	½ minute
Final elongation	72°C	10 minutes

Table 2.5 PCR cycling conditions. Table summarising the cycling conditions to produce PCR products for molecular cloning.

2.1.4 Purified of PCR products

Following PCR amplification PCR products were purified using the QIAquick PCR purification kit (28104) from Qiagen. The PCR sample was mixed with 100 µl of buffer PB1 and applied to a QIAquick spin column that binds DNA and centrifuged for 1 minute at 10,000 x g. The flow through was discarded and the column washed with 750 µl of buffer PE and centrifuged as above. The flow through was discarded and the column centrifuged again to remove any residual buffer PE. For DNA elution 30 µl of buffer EB was added to the column, incubated at room temperature for 1 minute before centrifuging. The purified PCR products were processed according to section 2.1.5

2.1.5 Restriction enzyme digestion of plasmid DNA

Restriction enzymes were purchased from New England Biolabs (NEB). Restriction digests were performed in a 50 µl reaction volume. This included 2 µg of plasmid DNA or 30 µl of purified PCR product. To this 2 µl of each restriction enzyme, 5 µl of restriction enzyme buffer (varied depending on the enzyme) and water was added to a total volume of 50 µl. Restriction digests were incubated overnight at 37°C. DNA fragments were separated using gel electrophoresis and extracted and purified from the agarose gel using the QIAquick Gel Extraction Kit as described in more detail in Section 2.1.6 & 2.1.7.

2.1.6 Agarose gel electrophoresis

Agarose Gel electrophoresis was used for size dependent separation of DNA molecules. Agarose gels were prepared by dissolving agarose in TAE buffer (40 mM Tris-acetate, 1 mM ethylenediaminetetraacetic acid (EDTA) pH 8.0). 0.7% agarose gels (0.7 g in 100 ml of TAE) were used to separate DNA fragments >10 kbp, 1% agarose gels (1 g agarose in 100 ml of TAE) were used to separate DNA

fragments between 1 and 10 kbp in size and 2% agarose gels (2 g of agarose in 100 ml of TAE) were used to separate DNA molecules less than 1 kbp in length. Sybr Safe was added to the agarose at a final concentration of 0.6 mg/ml to allow DNA visualisation. Gels were set in a gel cassette at room temperature. DNA samples were prepared using loading buffer (0.05% bromophenol blue, 40% sucrose, 1M EDTA pH 8.0, 0.5% SDS) and ran alongside 1 kbp or 100 bp ladders used as molecular weight standards. Gels were run at a constant voltage of 90V until the loading dye had reached the bottom of the gel. The resolved bands were visualised using ultraviolet (UV) light.

2.1.7 Agarose gel extraction

DNA fragments were extracted and purified from the agarose gel using Qiagen's QIAquick Gel Extraction Kit (28704). Bands of interest were cut from the agarose gel and the gel slice was added to a 1.5 ml Eppendorf. Gel slices were dissolved in buffer QG at 50°C for 10 minutes, the volume of buffer used was varied depending on the weight of the gel slice (100 µl of buffer per 100 mg). The solution was vortexed every 2-3 minutes throughout the incubation. Once the gel slice had completely dissolved isopropanol was added in an equal volume to buffer QG and the sample was applied to a QIAquick spin column. The column was centrifuged at 10,000 x g for 1 minute before washing with 500 µl of buffer QG and repeating the centrifugation step to remove any remaining agarose. The column was then washed with 750 µl of buffer PE by centrifugation, then centrifuged an additional time in the absence of buffer to remove any remaining buffer PE. Finally, DNA was eluted from the column using 50 µl of buffer EB and centrifuged at 21,000 x g for 1 minute.

2.1.8 DNA ligation

DNA ligations were set up in 10 µl volumes according to Table 2.6, mixed well by pipetting and incubated overnight at room temperature.

Reagent	1X Volume
Vector DNA	20 ng
Insert DNA	60 ng
T4 DNA ligation buffer containing 1 mM ATP	Final 1X concentration
T4 DNA ligase	400U
Nuclease free H_20	To a total volume of 10 µl

Table 2.6 Reaction components required to set up the DNA ligation reaction.

2.1.9 Bacterial transformation

Recombinant plasmids produced by molecular cloning were transformed into DH5 alpha component cells for DNA production or BL21 cells for protein production. Plasmids produced by TA cloning were transformed into one shot Topo10 cells. 50 µl of competent cells were thawed on ice then mixed with 1-100 ng of plasmid DNA and incubated on ice for 30 minutes. Cells were subjected to heat shock by incubation at 42°C for 45 seconds then incubated on ice for 2 minutes. Cells were recovered by addition of 950 µl of SOC media and incubated at 37°C shaking at 250rpm for 1 hour. Following incubation, the cells were gently pelleted by centrifugation for 30 seconds at 3000 x g. The supernatant was removed and the cells were re-suspended in 50 µl of SOC media and plated onto solid LB agar plates with the appropriate selection antibiotic.

2.1.10 Bacterial cultures

Bacterial cultures were used for plasmid DNA preparation. Cultures were grown from single colonies of competent cells grown and selected on solid LB agar plates with the appropriate selection antibiotic. A single colony was used to inoculate 10ml of LB broth and grown overnight at 37°C shaking at 250rpm in the presence of the appropriate selection antibiotic. Cultures were handled under sterile conditions in the presence of a flame. Bacterial cultures containing the plasmid were maintained as glycerol stocks in 25% glycerol at -80°C.

2.1.11 DNA extraction from bacterial cells

Plasmid DNA was isolated from bacterial cultures using the QIAprep spin Miniprep kit (Qiagen) according to the manufacturer's instructions. Bacterial cultures were pelleted by centrifugation at 5000 x g for 20 minutes. Culture media was removed and the pelleted cells were lysed in 250 µl of alkaline buffer P1 containing RNaseH, the lysates were neutralized and adjusted to high salt binding conditions using 250 µl of buffer P2 and 350 µl of buffer N3 and cleared by centrifuging at 13,000 x g for 10 minutes. The cleared lysate was loaded onto a silica column, which adsorbs DNA in the presence of high salts and centrifuged for 1 minute. Unbound sample was discarded and the column was washed with 500 µl of buffer PB and centrifuged for 1 minute. Buffer PB was discarded and the column was washed with 750 µl of buffer PE and centrifuged. The flow through was removed and the column was centrifuged an additional to remove any remaining buffer PE. Finally, plasmid DNA was eluted using 50 µl of buffer EB and centrifuged for 1 minute at 13,000 x g.

2.1.12 DNA quantification

DNA concentration was determined using a Nanodrop spectrophotometer; the absorbance of 2 µl of DNA was measured at 260 nm in relation to the elution buffer, where an absorbance of 1.0 corresponds to 50 µg/ml of double stranded DNA. DNA purity was determined using the ratio between the absorbance at 260nm and 280nm. Ratios between 1.8 and 2.0 were assumed to be pure.

2.1.13 DNA sequencing

All new constructs produced by molecular cloning were sequenced by Genewiz.

15 µl of plasmid DNA was provided at a concentration of 100 ng/µl.

2.2 Molecular Biology: RNA manipulation

2.2.1 RNA extraction from mammalian cells

Total RNA was isolated from mammalian cells using the RNeasy kit (74104) from Qiagen. Cells were plated at 60-80% confluence in a 6 well plate prior to RNA isolation and treated according to the methods in Section 2.3 on cellular biology. Culture media was removed and the cells were washed with PBS then 350 µl of RLT buffer was added to the plate to lyse the cells. The cell lysate was removed from the plate using a cell scraper and added to 1.5ml microcentrifuge tubes. The lysate was homogenised by vortexing, mixed with an equal volume of 70% ethanol and transferred to an RNeasy spin column that binds RNA. Unbound RNA was removed by centrifuging the spin column for 30 seconds at 10,000 x g. The spin column was washed with 700 µl of RW1 buffer and centrifuged as above. Contaminating DNA was eliminated by incubation with 10 µl of DNaseI prepared in 70 µl of buffer RDD for 15 minutes at room temperature. The column was washed with 700 µl of RW1 buffer and centrifuged. The wash step was repeated a further 2 times using 500 µl of RPE buffer. The spin column was then placed in a

new 2ml collection tube and centrifuged for 1 minute at 22,000 x g to remove any remaining buffer. Bound RNA was eluted in 30-50 µl of nuclease free H_20. RNA was quantified using UV spectrophotometry at 260nm wavelength as described in section 2.1.12 and stored at -80°C throughout the study.

2.2.2 RNA extraction from tissues

Total RNA was isolated from Formalin fixed paraffin embedded tissue samples using the Qiagen RNeasy FFPE kit (73504). Six 5 µM thick tissue sections were used in each extraction. Using a scalpel the tissue sections were scraped into 1.5ml microcentrifuge tubes. 320 µl of deparaffinisation solution was added and the tube was vortexed vigorously for 10 seconds. Samples were then incubated at 56°C for 3 minutes to melt the paraffin. Following incubation 240 µl of buffer PKD was added and the sample was mixed by vortexing before centrifuging for 1 minute at 11,000 x g, 10 µl of proteinase K was added to the lower clear phase of the sample and mixed by pipetting before incubating at 56°C for 15 minutes then 80°C for 15 minutes. The lower phase was then transferred to a clean microcentrifuge tube, incubated on ice for 3 minutes then centrifuged for 15 minutes at 20,000 x g. The supernatant was transferred to a new tube and contaminating DNA was removed by incubation at room temperature for 15 minutes with 10 µl of DNase I prepared in 25 µl of DNase booster buffer. Following DNaseI treatment the sample was mixed with 500 µl of buffer RBC and 1200 µl of 100% ethanol and mixed by pipetting. The sample was then transferred to an RNeasy spin column and centrifuged at 10,000 x g for 30 seconds, this step was repeated until all the sample had passed through the column. The column was washed with 500 µl of buffer RPE and centrifuged as above, the flow through was discarded and this step was repeated but centrifuging for 2 minutes. The spin column was placed in a new 2ml collection

tube and centrifuged for 5 minutes at 20,000 x g to remove any remaining buffer. Total RNA was eluted by applying 30 µl of RNase free H_20 to the column and centrifuging at 20,000 x g for 1 minute. RNA extractions from formalin fixed paraffin embedded tissues were performed at St James's University hospital by Dr Kate Marks, the RNA was provided for analysis.

2.2.3 Reverse transcription

Total RNA from cell lines and tissue sections was converted to cDNA for gene expression analysis using the ThermoFisher superscript IV reverse transcriptase. 1.5 µg of RNA was transcribed in a total reaction volume of 20 µl using a final concentration of 2.5 µM Oligo dT primer and 10 mM dNTP. The mixture was heated at 65°C for 5 minutes and then chilled on ice. To each sample 1 unit of reverse transcriptase, 1 unit of RNase inhibitor, 10 mM DTT and reverse transcriptase buffer to a final concentration of 1x was added and the samples were incubated at 42°C for 50 minutes followed by 70°C for 15 minutes to inactivate the enzyme. cDNA samples were stored at -20°C throughout this study.

2.2.4 Quantitative real time PCR

Quantitative real time PCR was used to study the expression of genes of interest. Table 2.7 shows the reaction components used to set up the PCR assay, PCR assays were prepared in Corbett tubes (Qiagen 981103) using the pre-chilled Corbett tube rack.

Every polymerase chain reaction included three phases: 1) Denaturation phase to separate double stranded DNA to single stranded templates 2) annealing of gene specific primers and 3) elongation catalysed by the DNA polymerase. The temperature used for elongation was the optimum temperature for the DNA polymerase. The annealing temperature was adjusted dependent on the primers

used. 40 cycles were used for amplification of cDNA products. PCR cycling conditions are shown in table 2.8

Several different RT-qPCR assays were performed in this study; the primers used to amplify the different genes are shown in Table 2.9. PCR products were analysed by agarose gel electrophoresis (see Section 2.1.6) to check that a single pure product was produced at the correct size.

Reagent	**1x volume required**
SensiMix	10 µl
Primer mix	1 µl (10 µM each)
Nuclease free H_20	4 µl
cDNA	5 µl (diluted 1:100)

Table 2.7 Composition of reaction mix for RT-qPCR. Table summarises the volumes of each reaction component required for gene expression analysis

PCR step	**Conditions**
Initial Denaturation	95°C for 10 minutes
Cycle	40
Denaturation	95°C for 15 seconds
Annealing	Temperature dependent on primers used, see table below 30 seconds
Elongation	72°C for 20 seconds

Table 2.8 Quantitative RT PCR reaction conditions. Table summarising the cycling conditions used for gene expression analysis.

Gene of Interest	Specificity	Primer Sequence	Annealing Temp (°C)
Grb2	Grb2 transcripts only	F: 5'CCCAAGAACTACATAGAAATGAAACC 3' R: 5' CCGCTGTTTGCTAAGCATTT 3'	54
Grb3-3	Grb3-3 transcripts only	F: 5' AACCACATCCGTTTGGAAAC 3' R: 5' TTCTGGGGATCAAAGTCAAA 3'	54
GAPDH	All transcripts	F: 5' AGAAGGCTGGGGCTCATTTG 3' R: 5' TTCTCATGGTTCACACCCATG 3'	54
hnRNPC	All transcripts	F: 5' GCTGCTCTGTTCATAAGGGCTT3' R: 5' GCCAGCAATCATTCTGCCATCC 3'	60
18s	All transcripts	F: 5' CAGCCACCCGAGATTGAGCA 3' R: 5' TAGTAGCGACGGGCGGTGTG 3'	60

Table 2.9 Primers used in qRT-PCR assays. Table includes genes targeted with the primers, transcript specificity, sequence and annealing temperatures used.

2.2.5 RNA-protein pull downs

RNA-Protein pull downs were performed to identify splicing factors that bind to *GRB2* alternate exon 4. A brief overview of the method is shown in figure 2.2. As shown, RNA is labelled with desthiobiotin (step 1) and immobilized on streptavidin coated magnetic beads (step 2). The beads are then used to capture proteins from nuclear lysates (step 3) which are eluted and identified using either western blotting or mass spectrometry (step 4 and 5).

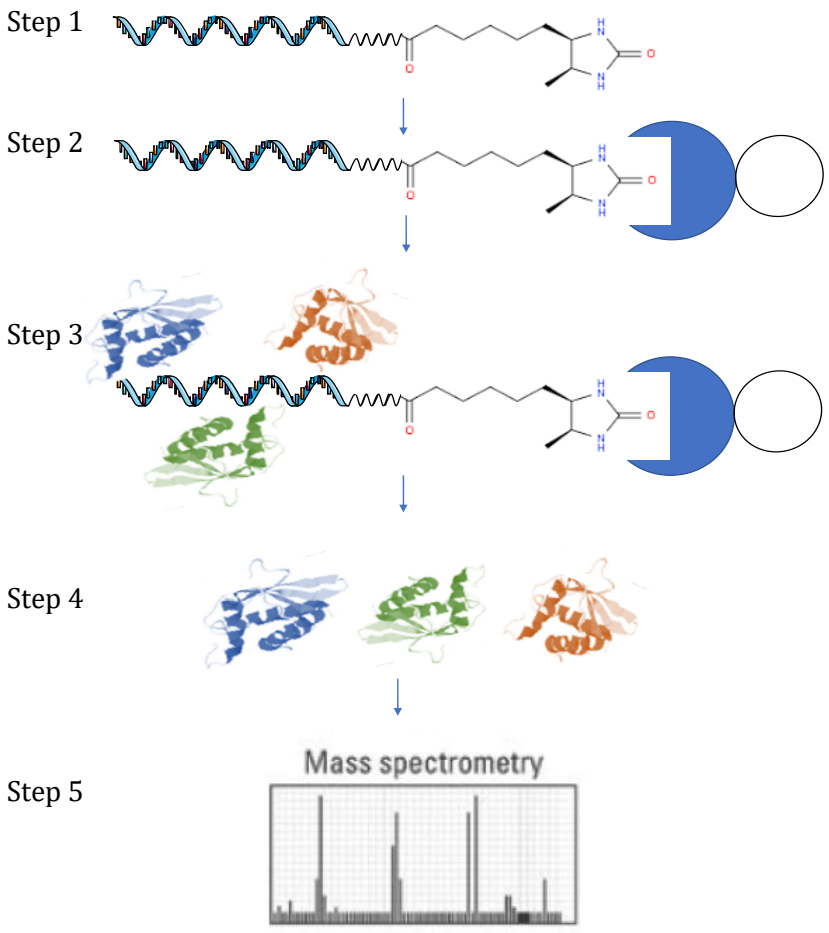


Figure 2.2 Overview of the method for RNA-protein pull downs.

2.2.5.1 RNA labelling

Synthetic RNA molecules spanning *GRB2* exon 4 were purchased from integrated DNA technologies reconstituted in water and stored at -80°C throughout the study. The sequences of the RNA oligos used for the pull down are shown in Figure 2.3. Synthetic RNA's were labelled using Pierce RNA 3' end Desthiobiotinylation Kit (20163). This kit uses RNA ligase to add a single desthiobiotinylated cytidine bisphosphate nucleotide to the 3' end of synthetic RNA oligos. This labelling method was chosen as it has minimal effect on RNA secondary structure which can be important for RNA protein interactions. The components for the ligase reaction are shown in Table 2.10, labelling was performed overnight at 16°C. Following overnight labelling 70 µl of nuclease free water was added to the ligation reaction and 100 µl of chloroform isoamyl alcohol to extract the RNA ligase. The RNA was then precipitated from the aqueous phase with 250 mM NaCl, 1 µl of glycogen and 300 µl of 100% ethanol and incubated for at least 1 hour at -20°C. The precipitated RNA was pelleted by centrifugation at 13,000 x g for 15 minutes, washed with 70% ethanol and then re-suspended in 20 µl of nuclease free water.

RNA oligo 1: GUGGUUUUUUGGCAAAAUCCCCAGAGCCAAGGCAGAAGAAAUGCUUAGCA

RNA oligo 2: AGCGGCACGAUGGGGGCCUUUCUUAUCCGAGAGAGUGAGAGCGCUCCUGGG

RNA oligo 3: GACUUCUCCCUCUCUGUCAAGUAAGUAUUUCCUGCUGCAGUUGCCUGGAA

Figure 2.3 Sequences of the RNA oligos derived from *GRB2* exon 4

Component	Volume (µl)	Final Concentration
Nuclease free water	3	n/a
10x RNA ligase reaction buffer	3	1x
RNase Inhibitor	1	40units
RNA oligo	5	50pmol
Biotinylated cytidine bisphosphate	1	1nmol
T4 RNA ligase	2	40units
Polyethylene glycol (30%)	15	15%

Table 2.10 Reaction components for 3' RNA labelling. Table summarises the volumes of each reaction component required for RNA labelling

2.2.5.2 RNA dot blot

Successful labelling of RNA was tested using a dot blot. A serial dilution of the labelled RNA was prepared and 2 µl was added to a nitrocellulose membrane. The RNA was crosslinked to the membrane by exposure to 254nm UV for 5 minutes. Labelled RNA was then detected using Thermo Scientific chemiluminescent nucleic acid detection module (89880). The membrane was blocked in blocking buffer for 15 minutes at room temperature. Following blocking the membrane was incubated with a streptavidin-horseradish peroxidase conjugate which binds to the desthiobiotinylated cytidine bisphosphate nucleotide on labelled RNA molecules. Excess streptavidin was removed by incubating in wash buffer for 5 minutes at room temperature, with shaking. The wash step was repeated four times before incubating the membrane in substrate equilibration buffer for 5 minutes. The membrane was then transferred to a plastic wallet and incubated with the peroxide substrate solution and imaged on the Syngene G.box.

2.2.5.3 RNA Protein pull downs

Labelled RNA was immobilised onto streptavidin coated magnetics beads and used to pull out proteins from cell lysates using Pierce magnetic RNA-protein pull down kit (20164). 50pmol of labelled RNA was used in each pull down reaction. The streptavidin coated magnetic beads were re-suspended by mixing and washed twice with 0.1 M NaOH, 50 mM NaCl solution. A second wash step was done using 100 mM NaCl. In each pull down 50 µl of re-suspended beads was used. Prior to RNA binding the beads were washed twice with 20 mM Tris pH 7.5. For RNA immoblization 50 µl of RNA capture buffer and 50pmol of labelled RNA was added and the beads were incubated at room temperature for 30 minutes with gentle agitation. Following incubation, any remaining unbound RNA was

removed and the beads were washed twice with 20 mM Tris pH 7.5. The RNA-protein binding reaction was set up according to Table 2.11 and incubated for 1 hour at 4°C. Unbound proteins were removed and the beads were washed twice with 100 µl of wash buffer. Bound proteins were eluted by incubation with 50 µl of elution buffer for 30 minutes at 37°C. Eluted proteins were identified using mass spectrometry or western blotting.

Reagent	**Volume per pull down**
10x Protein-RNA binding buffer	10 µl
50% glycerol	30 µl
Mammalian cell lysate (2 mg/ml)	30 µl
Nuclease-free water	30 µl
Total	100 µl

Table 2.11 Reaction components for the RNA-protein pull downs. Table summarises the volumes of each reaction component required for the pull down.

2.3 Cellular Biology

2.3.1 Mammalian cell culture

Mammalian cells were maintained as sub-confluent cultures at 37°C with 5% $C0_2$ in the appropriate medium. Colorectal Caco2 cells were cultured in Eagles modified essential media (EMEM) supplemented with 10% fetal bovine serum (FBS) and 50 µg/ml gentamycin. Colorectal colo320 cells were cultured in RPMI media supplemented with 10% FBS and 50ug/ml gentamycin. Modified Hek293t cells were cultured in Dulbecco's Modified Eagle medium (DMEM) media supplemented with 10% fetal bovine serum (FBS), 50 µg/ml gentamycin and 7 µg/ml puromycin to maintain knockdown of Grb2 or stable expression of FGFR2. Cells were passaged twice a week with trypsin and the culture medium was changed every three to five days. Cell stocks were frozen in 80% FBS, 20% DMSO and stored in liquid nitrogen. All cell culture work was carried out in a class II tissue culture hood.

2.3.2 Cell counting

Cells were counted using a heamocytometer. 10 µl of diluted cell suspension was added to the heamocytometer and the number of cells in each quartile was counted and an average taken. To calculate the number of cells per ml the average was multiplied by 10,000.

2.3.3 Mammalian cell transfection

Mammalian cells were transfected using liposome based transfection reagents optimized for the cell line. 24 hours prior to transfection mammalian cells were seeded onto treated 6 well culture plates at 50-60% confluency. Immediately before transfection fresh media was added to the cells. Caco2 cells were transfected using the TransfeX reagent, 5 µl of TransfeX reagent was used for 2.5 µg of plasmid DNA. Colo320 cells were transfected with lipofetamine 2000, 12.5

µl of lipofectamine was used for transfection of 2.5 µg of plasmid DNA. Hek293t cells were transfected with metafectene, 5 µl of metafectene was used for transfection of 2 µg of plasmid DNA; for co-transfection of two plasmids the amount of metafectene used was doubled. For lipofectamine and metafectene transfections plasmid DNA and the transfection reagent were each mixed with 250 µl of reduced serum media and incubated for 5 minutes at room temperature. The DNA:media solution was added to the transfection reagent media solution and incubated for a further 15 minutes before adding dropwise to the cells. For the TransfeX reagent DNA was added to 250 µl of reduced serum media, mixed by pipetting then the TransfeX reagent was added and the sample was incubated for 15 minutes at room temperature. Cells were incubated for a further 24 hours following transfection to allow protein expression.

2.3.4 Ligand stimulation of mammalian cells

Cells were cultured in 6 well plates and maintained at approximately 80% confluency for ligand stimulation. 24 hours prior to stimulation the cells were washed 3x in PBS, then placed in serum free media with the appropriate antibiotics. Cells were incubated with 20ng of ligand for the appropriate length of time (15 minutes for FGF9 or 5 minutes for all other ligands) at 37°C. Following incubation the media was removed and the cells washed with ice cold PBS then lysed according to section 2.3.5.

2.3.5 Mammalian cell lysis & quantification

Adherent cells were detached from tissue culture flasks by trypsinization or cell scraping. Cell suspensions were transferred to 15ml falcon tubes and pelleted by centrifuging at 3000 rpm for 5 minutes. Pellets were washed with ice cold PBS and then re-suspended in mammalian cell lysis buffer (RIPA lysis buffer Thermo

Fisher, supplemented with protease inhibitors). The volume of cell lysis buffer used was varied depending on the number of cells, for 6 well plates 200 µl of cell lysis buffer was used for 10 cm dishes 1 ml of buffer was used. Cells were incubated in lysis buffer for 15 minutes on ice, the insoluble material was removed by centrifugation at 15,000 rpm for 15 minutes.

Mammalian cell lysates were quantified using the BCA assay (Pierce BCA Protein assay kit 23225). This assay combines the reduction of Cu^{2+} ions to Cu^{1+} ions by protein with the reaction of bicinchoninic acid (BCA) with Cu^{1+} ions to produce a purple coloured reaction product that can be detected by measuring the absorbance at 562 nm. In this assay a standard curve is produced using known concentrations of bovine serum albumin (BSA) and used to determine the concentration of unknown protein samples. Standard curves were prepared according to table 2.12 and 100 µl of protein standards or 10 µl of cell lysates were added in duplicate to a 96 well plate. Working reagent was prepared by diluting BCA reagent B by 50x in reagent A. 200 µl of working reagent was added to each protein sample and the plate was incubated at 37°C for 30 minutes to allow the reduction reaction to occur. Absorbance readings were measured using a plate reader.

BSA standard	Volume of diluent	Volume of BSA stock solution (2 mg/ml)	Final BSA concentration (mg/ml)
A	0	300 µl of stock	2
B	125 µl	375 µl of stock	1.5
C	325 µl	325 µl of stock	1
D	175 µl	175 µl of standard B	0.75
E	325 µl	325 µl of standard C	0.5
F	325 µl	325 µl of standard E	0.25
G	325 µl	325 µl of standard F	0.125
H	400 µl	100 µl of standard G	0.025
I	400 µl	0	0

Table 2.12 BCA assay protein standards. The table shows how to prepare the standard curve for the BCA assay; standards were diluted in lysis buffer

2.3.6 Cell Fractionation

For cell fractionation studies mammalian cells were harvested by centrifugation and re-suspended in 1 ml of homogenisation buffer (250 mM sucrose, 3 mM imidazole pH7.4, 1 mM EDTA and protease inhibitors) and lysed using mechanical force produced with a tissue grinder and then passed through a 0.8 mm and 0.4 mm needle. Nuclear fractions were removed by centrifugation at 600 x g for 10 minutes. The supernatants were then transferred to ultracentrifugation tubes and made up to a total volume of 4 ml using PBS. Mitochondrial fractions were removed by ultracentrifugation at 20,000 x g for 30 minutes and microsomal membranes were pelleted by centrifugation at 100,000 x g for 90 minutes, the cytoplasmic fraction remains in the supernatant and was concentrated before use. Microsomal membranes were re-suspended in homogenisation buffer and applied to a sucrose gradient produced from 40-100% sucrose and centrifuged at 100,000 x g for 90 minutes to pellet the plasma membrane and remove the Golgi and ER. The plasma membrane was then re-suspended in mammalian cell lysis buffer. Cell fractions were quantified using the BCA assay and then analysed using immunoblotting (see Sections 2.3.5 and 2.4.10).

2.3.7 Cell Proliferation assay

Cell proliferation assays were performed using the BrdU cell proliferation kit (CST 6813). The assay measures the incorporation of 5-bromo-2'-deoxyuridine (BrdU) into newly synthesised DNA during cell proliferation. Cells were plated at a density of 10,000 cells per well in a 96 well plate. 24 hours after seeding cells were transfected with the appropriate plasmid according to section 2.3.3 and BrdU was added to each well at a final 1x concentration. 24 hours after transfection BrdU containing media was removed and replaced with 100 µl of fixing/denaturing solution and incubated at room temperature for 30 minutes.

Fixing/denaturing solution was removed and replaced with 100 µl of mouse anti-BrdU detection antibody diluted 1 in 100 in detection antibody diluent and incubated at room temperature for 1 hour. Following incubation the antibody was removed and the wells were washed 3x in wash buffer to remove any unbound antibody. Horse-radish peroxidase (HRP) conjugated anti mouse secondary antibody was diluted 1 in 100 with HRP linked antibody diluent and 100 µl was added to each well and incubated at room temperature for 30 minutes. Secondary antibody was removed and the wash procedure was repeated to remove unbound secondary antibody. 100 µl of TMB substrate was added to each well and incubated at room temperature for up to 30 minutes. 100 µl of stop solution was added to each well and the absorbance at 450nm was immediately measured using a plate reader.

2.3.8 Immunohistochemistry

Immunohistochemistry (IHC) was performed in order to look at the levels of Raf-1 serine 338 phosphorylation and the localisation in formalin fixed paraffin embedded tissue samples. Tissue sections were dewaxed and antigenic epitopes were retrieved using a pressure cooker. The pressure cooker was filled with 50 0ml of distilled water and glass slides with the tissue sections mounted on were added and incubated at 125°C for 2 minutes then gradually cooled to 90°C over 30 minutes. The slides were then placed briefly in a plastic container containing 20 ml of vector antigen solution and 200 ml of water then transferred to a plastic container containing wash buffer and rinsed under running water. Slides were then transferred to Sequenza racks and washed with Menarini wash buffer before blocking with bloxs-all reagent for 15 minutes at room temperature. Excess bloxs-all was removed by washing with an excess of Menarini wash buffer 3 times and then the slides were incubated in 2.5% horse serum for a further 15 minutes at

room temperature then the wash step was repeated. Primary antibody was diluted 1:100 in zymed and the slides were incubated in primary antibody for 1 hour at room temperature. Excess antibody was removed by washing 3 times; slides were then incubated in secondary antibody for 1 hour at room temperature before repeating the wash step. HRP solution was added to the tissue sections and incubated for 30 minutes at room temperature and then the wash step was repeated. Slides were then incubated in diaminobenzidine (DAB) for approximately 5 minutes. HRP catalyses the oxidation of DAB to produce a visible brown colour, excess DAB was removed using running water. Slides were then counterstained in haematoxylin to stain the nucleus for 30 seconds, rinsed under running water and then incubated in lithium carbonate for 1 minute, then rinsed again under running water. The stained tissue sections were then dehydrated by incubating in 100% ethanol for 3 minutes, this step was repeated 3 times in total then performed again but with xylene. Finally, coverslips were added on top of the tissue sections using DPX mounting medium. Stained slides were imaged on a light microscope.

2.4 Biochemical Techniques

2.4.1 Bacterial protein expression

Recombinant plasmids were transformed into *E.coli* BL21 (DE3) cells for protein expression and single colonies were used to inoculate 10 ml of LB media supplemented with the appropriate selection antibiotic and grown overnight at 37°C, 220 rpm. The overnight culture was used to inoculate larger culture volumes which were grown until an optical density (OD) of 0.6-0.8. Optical density was determined by measuring the absorbance of 1 ml of cell suspension in a cuvette at 600 nm relative to media only. At an OD of 0.6-0.8 protein expression was induced by the addition of IPTG to a final concentration of 0.1 mM

and the cells were grown overnight at 20°C. The following day cells were harvested by centrifugation at 4000 rpm for 20 minutes and cell pellets were frozen at -20°C until further use.

2.4.2 Protein purification

Affinity chromatography was used for purification of bacterially expressed recombinant proteins. Bacterial cells were thawed on ice and re-suspended in buffer A (20 mM Tris pH8.0, 150 mM NaCl and 1 mM 2-mecaptoethanol) and lysed by sonication (5 second bursts with 5 second intervals). The lysate was centrifuged at 13,000 x g for 1 hour and the supernatant (soluble fraction) was loaded onto an affinity column. Hexa-His tagged fusion proteins were purified using a nickel affinity column, glutathione-S-transferase (GST) fusion proteins were purified using a glutathione affinity column and maltose binding protein (MBP) fusion proteins were purified using amylose affinity columns. Affinity columns were washed with buffer A until no protein was detected using the absorbance at 280 nm and then specifically bound protein was eluted. His-tagged proteins were eluted using buffer A supplemented with 200 mM imidazole to compete out the histidine for binding to the nickel residues. GST tagged proteins were eluted in buffer containing 10 mM reduced glutathione, Tris-HCl pH 8.0 and 1 mM 2-mecaptoethanol. MBP fusion proteins were eluted using 20 mM maltose, 20 mM Tris-HCL pH 7.5, 150 mM Nacl and 1 mM 2-mecaptoethanol.

2.4.3 Size exclusion chromatography

Proteins were further purified using pre-packed Superdex 75 or Superdex 200 size exclusion chromatography columns and Akta FPLC system (Amersham Biosciences) at 4 °C. The column was equilibrated with 5 column volumes of gel filtration buffer (varied depending on the protein application). Protein samples

were concentrated to less than 5 ml and injected into the equilibrated column at a flow rate of 0.5 ml/minute and 1 ml fractions were collected. Fractions containing protein were identified using the UV absorbance at 280 nm and confirmed by SDS-PAGE analysis and coomasie blue staining (see method 2.4.7 and 2.4.8).

2.4.4 Protein quantification

Protein concentration was determined using a Nanodrop spectrophotometer; the absorbance of 2 µl of protein was measured at 280 nm in relation to the elution buffer where an absorbance of 1.0 corresponds to 1 mg/ml of BSA. The molar concentration was calculated by dividing the concentration (mg/ml) by the extinction coefficient. The extinction coefficient was determined using the amino acid sequence and the program Expasy ProtParam.

2.4.5 GST pull down

GST protein pull downs were performed to study protein-protein interactions under various conditions. GST-fusion proteins were expressed in bacteria according to method 2.4.1 and the cell pellets were lysed using sonication and centrifuged to separate the soluble fraction (see Section 2.4.2). The soluble fraction was added to glutathione cross-linked agarose beads and incubated for 1 hour at 4°C to capture the GST-fusion protein. Following incubation the beads were pelleted by centrifuging at 1000 rpm for 1 minute and unbound proteins were discarded, the beads were washed with 1 ml of PBS-T to remove any non-specifically bound proteins. The washing procedure was repeated 5x and the beads were re-suspended in 25% glycerol and stored at -20°C throughout the study.

For pull down reactions 20 µl of the GST-fusion protein bead suspension was added to 200 µl of cell lysate or purified protein and incubated overnight at 4°C on an orbital rotator. The beads were pelleted by centrifuging at 1000rpm for 1 minute and unbound proteins were removed. The beads were washed with 1 ml of PBS-T, pelleted by centrifuging (as above) and then the PBS-T was removed taking care not to remove the beads. This washing procedure was repeated 5x. Bound proteins were then eluted in 100 µl of 2x lamelli buffer and boiled at 95°C for 5 minutes. The eluted proteins were analysed by SDS-PAGE and western blotting according to Sections 2.4.7, 2.4.10 and 2.4.11.

2.4.6 RFP pull down

RFP pull downs were also performed to study protein-protein interactions. In contrast to GST pull downs in which binding partners are co-purified from a lysate onto glutathione beads, RFP pull downs precipitate the complex and therefore the binding event occurs within intact cells. Mammalian cells were transfected with RFP tagged fusion proteins, serum starved then treated with growth factors and cell lysates were prepared according to sections 2.3.3-2.3.5. The cell lysates were quantified using the BCA assay and diluted to the same concentration, 100 µl of lysate was added to 30 µl of agarose bead suspension pre-coupled with anti RFP antibodies and incubated at 4°C for 1 hour. The beads were pelleted by centrifugation at 1500rpm for 1 minute, unbound proteins were removed, and the beads washed with 1 ml of PBS-T. This washing procedure was repeated 5x, bound proteins were eluted in 100 µl of 2x lamelli buffer and boiled at 95°C for 5 minutes. The eluted proteins were analysed by SDS-PAGE and western blotting (see Sections 2.4.7, 2.4.9 and 2.4.10).

2.4.7 SDS polyacrylamide gel electrophoresis

Protein electrophoresis was used for size dependent separation of proteins from cell lysates prior to western blotting or to assess the purity of bacterially expressed recombinant proteins. Protein samples were mixed with 4x Laemmli loading dye (8% SDS, 250 mM TrisHCL pH 6.8, 35% glycerol, bromophenol blue, 50 mM DTT) to a 1x final concentration and denatured by boiling at 95°C for 5 minutes. Protein samples were loaded alongside molecular weight ladders onto 4-15% pre-cast gels and separated by electrophoresis at 180v in running buffer (25 mM Tris, 192 mM Glycine, 0.1% SDS) using the Biorad Mini-PROTEAN II system. Proteins on the gel were either transferred to nitrocellulose membranes for immunoblotting (see Section 2.4.10) or stained to be visualized (see Section 2.4.8).

2.4.8 Coomasie Blue staining

Coomassie Blue staining was used for non-specific staining of all proteins. Gels were transferred to a plastic container containing Coomassie Blue solution (0.2% w/v coomasie brilliant blue, 40% methanol and 10% acetic acid) and incubated for 60 minutes with gentle agitation. Unbound dye was removed using destaining solution (40% methanol, 10% acetic acid, 50% H_20) until the protein bands could be visualised clearly.

2.4.9 Western transfer

Western blot analysis was performed to specifically detect proteins of interest following electrophoresis. Polyacrylamide gels were transferred to nitrocellulose membranes using the iBlot 2 gel transfer device and iBlot2 transfer stacks. The iBlot 2 transfer stack consists of a bottom stack and a top stack. The bottom stack is composed of a plastic tray, copper electrode, transfer gel layer and transfer

membrane. The polyacrylamide gels were soaked in western transfer buffer (48 mM Tris, 39 mM glycine, 0.037% SDS and 20% methanol) and then placed on top of the transfer membrane in the bottom stack. A piece of white filter paper soaked in water was then placed on top of the gel and the top stack was added. The top stack consists of a plastic separator, transfer gel layer and copper electrode. An absorbent pad was then placed on top of the top stack. Transfers were performed at 25 volts for 7 minutes. Nitrocellulose membranes were then removed, placed in a plastic container and processed according to Section 2.4.10.

2.4.10 Immunoblotting

Membranes were washed for 30 minutes in ultrapure H_20 to remove transfer buffer before incubating at room temperature for 1 hour in 3% BSA in TBS-T (3 g of BSA per 100 ml of TBS-T) to block the remaining protein binding sites on the membrane. Primary antibodies were added at a dilution of 1:1000 prepared in blocking buffer (3% BSA in TBS-T) and incubated overnight at 4°C. Table 2.13 shows the antibodies used in this study. Non-specifically bound antibodies were removed by washing 5x for 10 minutes in TBS-T buffer. The membrane was then incubated in Horseradish Peroxidase conjugated immunoglobulin G antibody diluted in 3% BSA for 1 hour at room temperature. Non-specifically bound secondary antibody was removed by washing 5x for 10 minutes in TBS-T buffer. The membrane was incubated with Pierce ECL western blotting substrate for 2 minutes. Detection was carried out using a Syngene G:BOX.

Name	Species	Epitope	Source
Phospho-C Raf (IHC)	Rabbit	Synthetic phosphopeptide corresponding to residues surrounding serine 338	Sigma
RFP antibody	Rabbit	15-residue peptide VNGHEFEIEGEGEGR, derived from N terminus of RFP	Genscript
Sos (C23) antibody	Rabbit	C-terminus of human Sos1	SantaCruz
E-Cadherin antibody	Rabbit	Epitope surrounding Pro780 of human E-cadherin	CST
Beta Actin antibody	Rabbit	Amino terminus	CST
Grb2 CSH3 antibody	Rabbit	C terminal SH3 domain	SantaCruz
Phospho cRaf antibody (western blots)	Rabbit	Synthetic phosphopeptide corresponding to residues surrounding serine 338	CST
Phospho Erk antibody	Rabbit	Synthetic phosphopeptide corresponding to residues surrounding Thr202/Tyr204	CST
Total ERK	Rabbit	Synthetic peptide from the c-terminus of rat p44 MAPK	CST
Total EGFR	Rabbit	Synthetic peptide syrrounding tyrosine 1068 in the human EGF receptor	CST

Phospho EGFR Y1068	Rabbit	Synthetic phosphpeptide syrrounding tyrosine 1068 in the human EGF receptor	CST
hnRNPC ANTIBODY	Rabbit	Peptide corresponsing to amino acids 100-200 from human hNRNPC1 and C2	Abcam
Strep antibody	Rabbit	Produced by immunising rabbits with a Strep tagged fusion protein	Generon
Tubulin Antibody	Rabbit	Produced by immunising rabbits with human alpha tubulin	CST
Ras Antibody	Rabbit	Produced by immunising rabbits with the amino terminus of K-Ras	CST
GST Antibody	Rabbit	Produced by immunising rabbits with a GST fusion protein	CST
His antibody	Mouse	His tagged fusion proteins	Clontech
Anti-mouse -HRP conjugated	Horse	Mouse primary antibodies	CST
Anti-rabbit -HRP conjugated	Goat	Rabbit primary antibodies	CST

Table 2.13 Antibodies used in this study. Table provides information on the supplier of the antibody, the host species used to generate the antibody and the region of the target protein that the antibody recognizes.

2.4.11 Image J quantification of immunoblots

Images of immunoblots were quantified using Image J software. A square box was drawn around each lane containing a band. The intensity of each band was measured by calculating the area under the curve. The intensity of variable protein bands were normalised by dividing by the intensity of housekeeping proteins whose expression remains constant under different conditions.

2.5 Biophysical techniques

2.5.1 Microscale thermophoresis (MST)

Microscale thermophoresis was used to measure the affinity of protein-protein and protein-peptide interactions. MST measurements were recorded using a NanoTemper Monolith NT.115 instrument. MST measures the temperature induced change in fluorescence of a fluorescently labelled molecule in response to titration of an unlabelled molecule and uses the measurements to produce a binding curve. The change in fluorescence of the labelled molecule is influenced by two factors; 1) the chemical environment of the fluorophore which changes upon complex formation and 2) the movement of the labelled molecule in a temperature gradient (thermophoresis) which is also affected by binding reactions as larger complexes move at a slower rate.

Proteins and peptides used in MST experiments were prepared in 20 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT. Fluorescent labelling of proteins was carried out in a total volume of 100 µl containing 100 µM of purified protein, 80 µl of labelling buffer (0.1mM sodium bicarbonate pH 8.3) and 10 µl of Atto488 NHS ester and incubated at room temperature for 1 hour in the dark. The labelled protein was separated from unbound dye using gel filtration.

MST experiments were set up using 16 samples, all of which contained the fluorescently labelled protein at a final concentration of 100 nM. The unlabelled protein was added to all but one of the samples and was serially diluted by 2 fold. The samples were then transferred into glass capillaries and inserted into the MST machine. On the MST machine the LED colour was set to green and the MST power was selected so that the maximum fluorescence measured was between 200-12000 counts. Analysis of the MST data was performed using NanoTemper softwear.

2.5.2 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) was used to measure the thermodynamic parameters of binding interactions. ITC experiments were performed using a VP Microcal instrument (Northampton, MA). The instrument has two calorimetry cells, molecules to be studied were placed in the sample cell and buffer was placed in the reference cell. The sample cell was filled with 200 µl of protein at 10 µM concentration prepared in 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT and 0.05% Tween-20. The syringe was filled with 20 µl of protein at 100 µM concentration prepared in the same buffer. Experiments were initiated by injecting protein from the syringe into the sample cell. Binding induces a heat and enthalpy change in the sample cell which is adjusted to bring the conditions to the same as the reference cell. This heat and enthalpy change is recorded and can be used to determine the affinity of the interactions and the stoichiometry of binding events. Experiments were performed at a fixed temperature in the range of 15 to 25°C and repeated 3 times. Data analysis was performed using Origin software version 9.1

2.5.3 Fluorescence resonance energy transfer (FRET)

Fluorescence resonance energy transfer (FRET) was used to study protein interactions in the cellular environment. FRET describes the transfer of energy between two chromophores. To study protein-protein interactions one protein is tagged with a donor fluorophore and the other protein is tagged with an acceptor fluorophore. The donor fluorophore is excited with light energy which can be transferred to the acceptor fluorophore resulting in fluorescent emission at its emission wavelength. Energy transfer can only occur between molecules that are within 1-10 nm proximity of one another which is physiologically relevant for studying protein-protein interactions.

Mammalian cells were co-transfected with two fluorescently tagged proteins and then grown on glass coverslips for 24 h before experimentation. As controls, cells expressing vector alone with the fusion protein were also analysed. Experiments were carried out under serum starved and stimulated conditions. Cells were fixed by addition of 4% (w/vol) paraformaldehyde, pH 8.0. Cells were washed three times with PBS, pH 8.0, and then mounted onto a slide with mounting medium containing DAPI. Cells were imaged using a LSM880 confocal microscope. DAPI was excited at 405 nm and used to locate the cells, GFP was excited at 488 nm, RFP was excited at 561 nm. For FRET analysis, GFP was excited at 488nm and RFP emission was detected.

2.6 *In silico* analysis

2.6.1 Ensembl genome browser

Ensembl genome browser is an online database containing genome annotations and information about DNA, cDNA and protein sequence and different splice variants of the gene. The browser was used to study *GRB2* splicing and for primer design.

2.6.2 SpliceAid 2

SpliceAid 2 is an online database containing information about splicing factor expression and their RNA binding sequences. The database contains 2220 target sites for 62 unique splicing factor proteins and there expression in 320 tissues [22-24]. This database was used to identify splicing factor binding sites within *GRB2* alternate exon 4. The RNA sequence corresponding to *GRB2* exon 4 was input into SpliceAid 2. The database searches for known RNA binding sequences within the input sequence and provides a table with the sequence element, its position within the input sequence and the splicing factor that recognises it.

Chapter 3: Analysis of *GRB2* splicing in colorectal cancer

3.1 Introduction

Alternate splicing (AS) is a critical step during gene expression and a major source of transcriptomic and proteomic diversity. In cancer AS becomes disrupted, leading to the production of protein isoforms that facilitate tumour development and progression. Many of the hallmarks of cancer [160] are affected by alternate splicing [161] including cell proliferation, angiogenesis and evasion of apoptosis [162-164]. Genome wide comparisons of AS in normal and cancer tissues have identified thousands of cancer associated AS events [45]. However, the cellular function of most alternate splicing events and their contribution to cancer biology remains unknown. Understanding the regulation and molecular consequences of cancer-associated splicing events is important and may lead to the development of a new class of anti-cancer therapeutics that target AS [64, 69, 72].

This thesis aims to characterize the molecular role of *GRB2* splicing in colorectal cancer. Grb2 is a ubiquitously expressed adaptor protein responsible for linking activated RTKs to the Ras-Erk signalling cascade. Its splice variant, Grb3-3 produced by skipping of exon 4 has been reported to inhibit EGF induced expression of Ras-responsive genes, suggesting that the two isoforms may have antagonistic functions in Ras signalling [89]. Ras signalling is up regulated in a large proportion of CRCs as well as many other cancers. Despite this, there has been no previous analysis of the role of this splicing event in cancer development, which may represent a novel route for hyperactivation of the Ras signalling pathway.

In this Chapter, a splice sensitive PCR assay was developed to enable the expression of the splice variants Grb2 and Grb3-3 to be quantified in colorectal

cancer and adjacent normal tissue samples. Immunohistochemical analysis of Raf-1 serine 338 phosphorylation was also used to assess Ras signalling in these samples.

3.2 Developing a splice sensitive PCR assay

There are several experimental methods that can be used to analyse alternate splicing [165]. Northern blotting and ribonuclease protection assays were amongst the first methods developed. In northern blotting, total RNA is isolated and electrophoresis is used for size-dependent RNA separation. Subsequently transcripts of interest are detected using hybridization probes complementary to the target gene. The probes anneal at exon-exon or exon-intron junctions to allow different splicing events to be analysed. In ribonuclease protection assays total RNA is first incubated with probes complementary to genes of interest. Following hybridisation of the probes the RNA is treated with nucleases that only cleave single stranded RNA (ssRNA). ssRNA is separated from the RNA-probe duplexes by electrophoresis allowing the levels of the undigested transcripts to be compared under different conditions or across different samples. These techniques are low throughput and suffer from a number of limitations. For instance, both techniques require large quantities of RNA and are not therefore suitable for low abundance transcripts, and in addition they can only be used to study splice variants that have a relatively large difference in size that can be resolved by electrophoresis [165]. Splice-sensitive DNA microarrays overcome some of these limitations. Microarrays are higher throughput, enabling spicing of a large number of genes to be measured simultaneously, and are more sensitive, requiring lower amounts of sample [166, 167]. In DNA microarrays, multiple probes against several different genes are immobilised onto a solid surface. cDNA is added to the surface and the probes bind to their complementary sequences.

Binding is detected using chemiluminescent detection methods, producing an array of black spots, where the intensity of the spots provides a measure of the abundance of that particular transcript. DNA microarrays can be used to analyse total gene expression and alternate splicing. For gene expression analysis the probes anneal at the 3' end of the gene, to analyse alternate splicing probes are designed that anneal at unique exon-exon or exon-intron boundaries [167]. RNA sequencing is another high-throughput method that can be used to compare alternate splicing events, and to identify novel splicing events. However, for studying a single splicing event, these methods are not cost effective.

To assess *GRB2* splicing, a splice sensitive, quantitative PCR assay was developed, using isoform specific primers [165, 168]. This method was chosen because qPCR is the gold standard for mRNA analysis. qPCR is extremely sensitive, allowing subtle changes in transcript abundance to be quantified, and making it suitable for low abundance transcripts. Furthermore, it has a large dynamic range; important for assessing *GRB2* splicing, since Grb3-3 is reportedly expressed at up to 50x lower levels than Grb2 [89]. In qPCR RNA is extracted and reverse transcribed into cDNA using oligo dT primers. Gene-specific primers are then used to exponentially amplify the specific cDNAs of interest. The concentration of the amplicon in the reaction is monitored using a DNA intercalating fluorescent dye (sybr green). Template quantitation is based on the number of PCR cycles required for the fluorescence to reach an arbitrary threshold. This is recorded as the Ct value shown schematically in Figure 3.1. Highly expressed transcripts reach this threshold after a lower number of PCR cycles than low abundance transcripts.

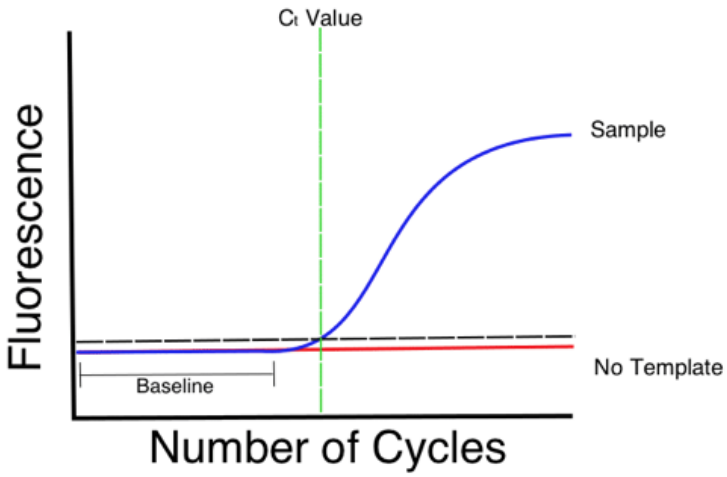


Figure 3.1 Schematic representation of qRT-PCR assay. Fluorescence is measured over a fixed number of PCR cycles. An example trace for a sample is shown in blue, and for the no-template control in red, the threshold is shown by the black dashed line, the Ct value shown by the green line is the cycle number when the fluorescence measurement crosses the threshold.

3.2.1 Primer Design

In order to measure the abundance of Grb2 and Grb3-3 splice variants in CRC by RT-qPCR transcript specific primers were designed. The two transcripts differ only by the presence (Grb2) or absence (Grb3-3) of exon 4. Two primer pairs were designed each specific to one single variant, shown schematically in Figure 3.2. For Grb3-3 this was achieved by designing a forwards primer complementary to the unique boundary between exon 3 and exon 5 created from skipping of exon 4. For Grb2 the forwards primer annealed to the unique sequence in the alternate exon 4. An additional primer pair against the reference gene GAPDH was designed as an internal control for variations in the amount of cDNA in each sample. All primers were between 20 and 25 nucleotides in length with melting temperatures between 50 °C and 60 °C. All primers showed a GC content of 40-60% with a GC clamp. The maximum amplicon produced with each primer pair was less than 250bp.

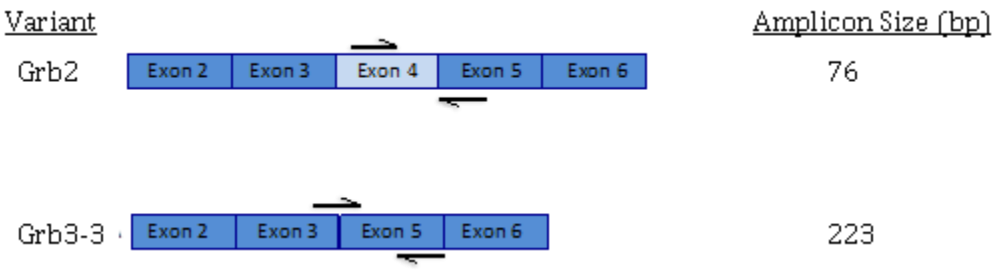


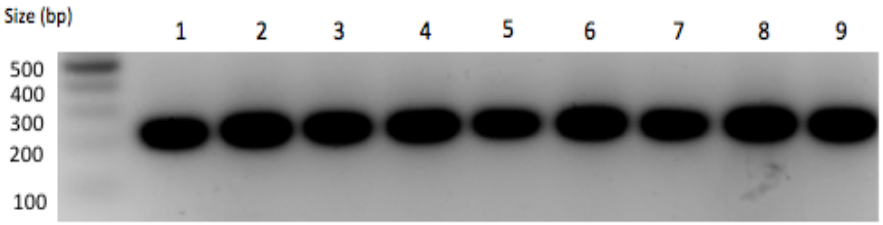
Figure 3.2 Splice sensitive PCR primer design. Schematic representation of *GBR2* splice variants. Constitutive exons shown in dark blue and alternate exons in light blue, primers are shown as black arrows above the exon they anneal too, and on the right hand side is the expected size of the amplicons.

3.2.2 Assessing primer specificity

Agarose gel electrophoresis was used to determine the specificity of the primers and to ensure there was no cross-reactivity between the two transcripts, or miss-priming of non-specific genes. Little is known about the expression of Grb3-3 in cell lines, although it is reported to be expressed in a range of tissues [89]. Therefore, to assess the specificity of the Grb3-3 primer pair Hela, cells were transfected with exogenous Grb3-3 cDNA prior to RNA extraction. The amount of Grb3-3 cDNA transfected was gradually reduced from 5 µg to 0 µg in order to determine the sensitivity of the primers. Although Grb2 is known to be expressed in Hela cells, some samples were also transfected with Grb2 prior to RNA extraction. This ensured no cross-reactivity under conditions where the abundance of Grb2 far exceeds Grb3-3. Figure 3.3a shows the PCR products produced with the Grb3-3 primers. In all samples a single band was produced corresponding to an amplicon between 200 and 300 base pairs in size, similar to the expected amplicon size of 223 base pairs (Figure 3.2). This band was seen in all samples, including those not transfected with exogenous Grb3-3 cDNA, indicating that the primers are sensitive enough to detect endogenous levels of the Grb3-3 transcript. The PCR fragment from sample 9 corresponding to endogenous Grb3-3 was gel extracted, cloned into a TA vector and sequenced confirming that the PCR product produced was Grb3-3. The specificity of the Grb2 specific primers was also assessed in Hela cells. Grb2 is known to be expressed at high levels in Hela cells, therefore no exogenous Grb2 cDNA was transfected prior to RNA extraction. However in order to ensure there was no cross reactivity with Grb3-3, one sample was transfected with 5 µg of Grb3-3 cDNA. As shown in Figure 3.3b, a single band was produced in both samples indicating no cross reactivity. The band observed was < 100 base pairs, which is similar to the

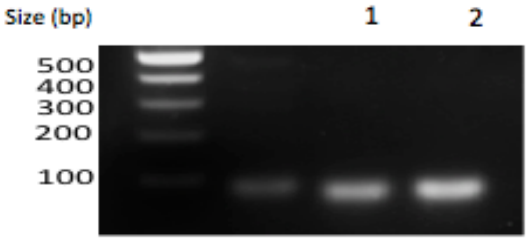
expected size amplicon of 76 base pairs. Sequencing of the PCR product confirmed it was Grb2 mRNA. Analysis of the PCR products, via agarose gel electrophoresis, showed the primer pairs were specific. They do not cross-react with the other splice variant or with non-specific genes because a single band corresponding to the desired product was produced. This was confirmed via melting curve analysis. This measures the temperature-dependent dissociation of two DNA strands during heating; if a single PCR product is produced a single peak is observed. For all primer pairs a single, well-defined peak was observed.

(a)



Lane	1	2	3	4	5	6	7	8	9
Grb2 (µg)	2.0	2.0	2.0	2.0	2.0	2.0	1.0	0.5	0.0
Grb3-3 (µg)	5.0	2.5	1.25	0.6	0.3	0.0	5.0	5.0	0.0

(b)



Lane	1	2
Grb2 (µg)	0.0	0..0
Grb3-3 (µg)	0.0	5.0

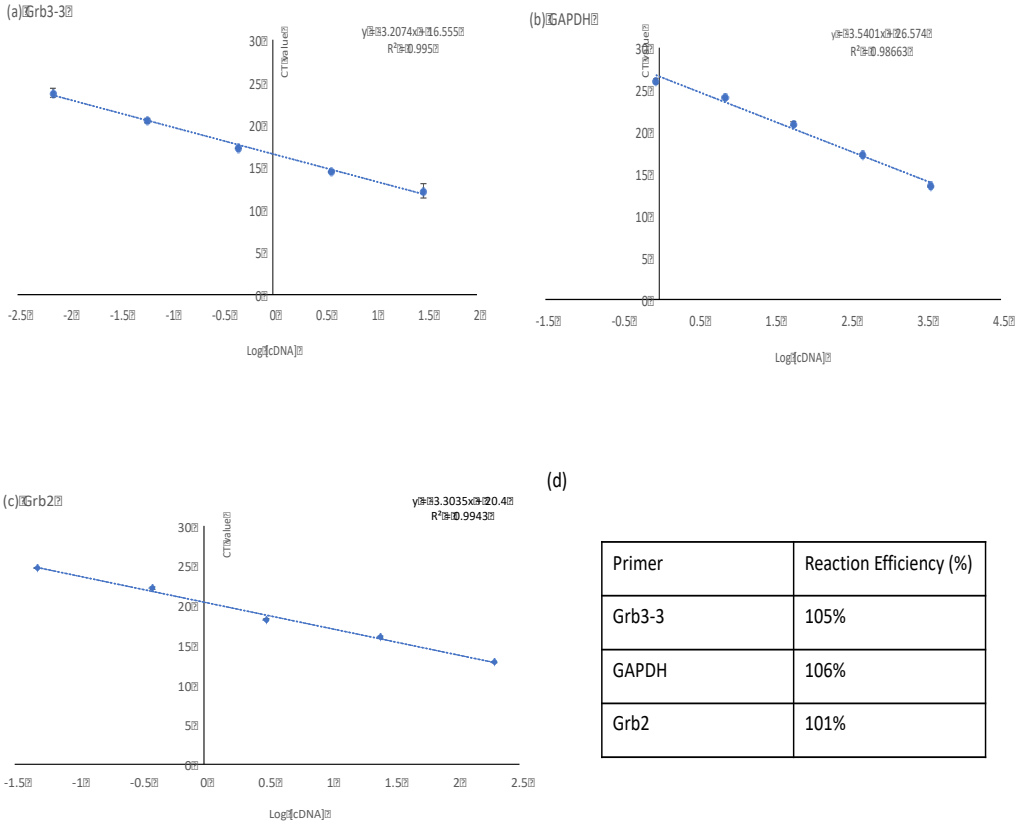
Figure 3.3 Determining the specificity of the isoform specific sensitive primers. 2% agarose gel showing the PCR products produced with the (a) Grb3-3 specific primer pairs and (b) Grb2 specific primers. The table underneath the gel images shows the amount of each plasmid transfected into HELA cells prior to RNA extraction and cDNA synthesis

3.2.3 Absolute Quantification

The level of Grb2 and Grb3-3 splice variants was determined using absolute quantification. Absolute quantification utilises a standard curve prepared from samples of known template concentration to interpolate the concentration of unknown samples. This method of quantification was chosen as it allowed the ratio of the two alternate transcripts to be measured and compared across different samples. Plasmid DNA containing a single copy of the Grb2, Grb3-3 or GAPDH (control) cDNA was used to produce the standard curves. Plasmid DNA was treated with RNases to remove contaminating RNA, and the concentration was determined using the absorbance at 260nm. Standard curves were produced using 8-fold serial dilutions of the plasmid resulting in a large dynamic range of four orders of magnitude. This gave rise to standard curves with a range of CT values between 15 and 30. Example standard curves are shown in Figure 3.4, and all raw PCR data can be found in Appendix 1.

The standard curves were also used to determine the efficiency of each PCR run. Efficiencies were calculated using the slope of the amplification curve (gradient) and the following equation: Reaction efficiency= 100 x 1 – ($10^{-1/\text{gradient of line}}$). Gradient values between -3.3 and -3.6 produced reaction efficiencies between 90-110%, deemed as acceptable by the MIQE (minimum information for publication of qRT-PCR data) guidelines [169]. PCR efficiencies are influenced by multiple factors including; the master mix, the specificity of the primers, the annealing temperature and the quality of the cDNA sample. A PCR efficiency of 100% indicates that the target sequence doubles during each cycle. Figure 3.4 shows examples of the standard curves and reaction efficiencies for each set of primers. The R^2 value measures how well the data points fit the line of best fit., For all three standard curves the R^2 value was greater than 0.99, indicating a strong

linear correlation between $\log_{10}$[plasmid DNA] and the CT value measured. The reaction efficiencies shown in Figure 3.4d were all between 90 and 110%, indicating that the amount of template doubles each PCR cycle.



(d)

Primer	Reaction Efficiency (%)
Grb3-3	105%
GAPDH	106%
Grb2	101%

Figure 3.4 Efficiency of the Primers. Scatter plot showing linear relationship between CT value and log[cDNA] for, (a) Grb3-3, (b) GAPDH, and (c) Grb2 primer pairs. (d) Summary Table of PCR efficiencies calculated using the equations of the line displayed on graphs a-c.

3.3 Analysis of GRB2 splicing in colorectal cancer

The splice sensitive RT-qPCR assay developed in Section 3.1 was used to quantify the expression of the splice variants Grb2 and Grb3-3 in 26 colorectal cancer patient samples. Only patients that lacked mutations in K-Ras, B-Raf and C-Raf, as determined by next generation sequencing were used in this study. RNA was extracted from the tumour tissue and the adjacent normal tissue from each patient and used to synthesise cDNA.

Grb2 mRNA abundance ranged from 0.18 - to 428 pg in the normal colonic tissue and 4.2 - 3977 pg in the tumour tissue, suggesting that Grb2 transcript levels are increased in the tumour tissue. Grb3-3 mRNA levels ranged from 1.12 - 1542 pg in the normal colonic tissue and 1.21 - 820 pg in the tumour tissue. Suggesting that in contrast to Grb2, Grb3-3 is more abundant in the normal tissue than the tumour tissue.

Figure 3.5 shows the fold change in Grb2 mRNA levels in the tumour tissue relative to the adjacent normal tissue. In 42% of patients (11 out of 26) Grb2 expression increased. In 7 out of the 11 patients the increase was greater than 2-fold and the maximum increase was 62-fold. In the remaining 58% (15 out of 26) of patients Grb2 expression decreased in the tumour tissue relative to the surrounding normal, however in only 6 of these the fold decrease was greater than, or equal to 2-fold. The maximum fold decrease in expression was 3.6.

Figure 3.6 shows the fold change in Grb3-3 mRNA levels in the tumour tissue relative to the adjacent normal tissue. In 38% of patients (10 out of 26) Grb3-3 expression increased, however in only half of these (5 patients, 19%) the increase was greater than 2-fold. In the remaining 62% (16 out of 26) of patients Grb3-3 expression decreased in the tumour tissue relative to the surrounding normal. In

9 out of the 16 patients the decrease was greater than 2-fold, and by a maximum of 41-fold.

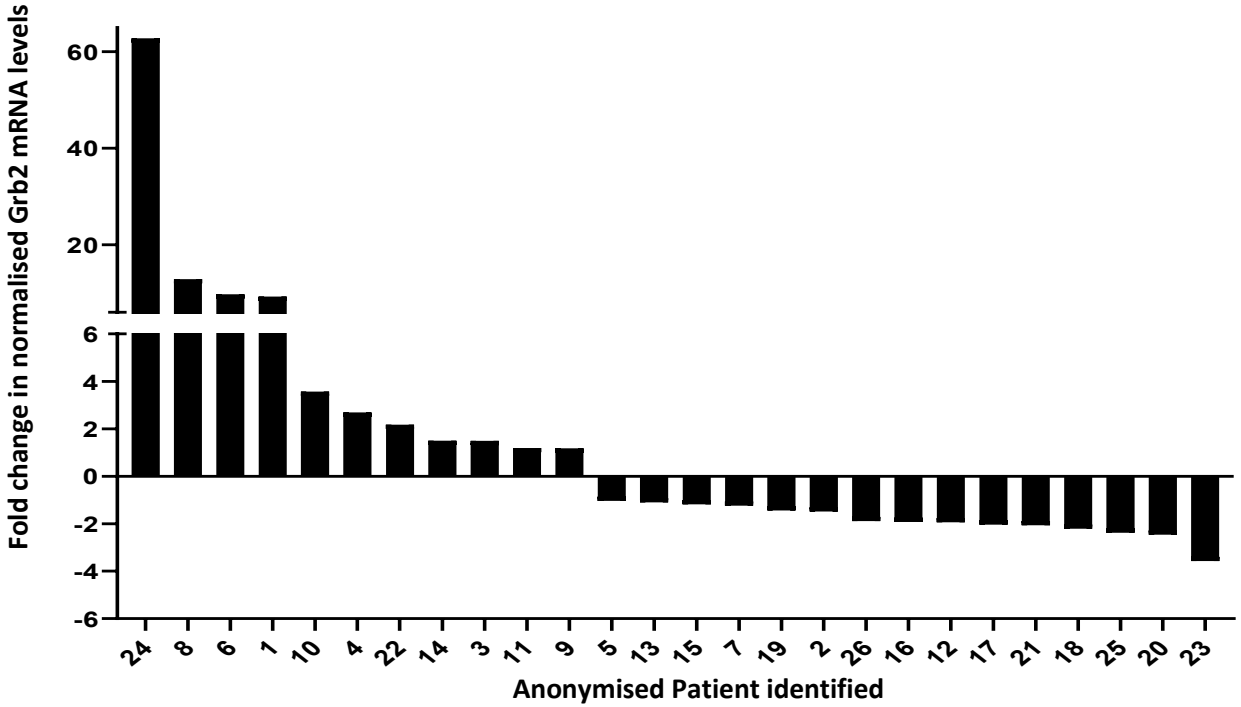


Figure 3.5 Fold change in normalised Grb2 mRNA levels in the tumour tissue relative to the normal. Waterfall plot showing the fold change in Grb2 mRNA abundance in tumour tissue relative to the associated normal for 26 colorectal cancer patients, anonymised patient identifiers are shown on the x-axis, negative values represent down regulation in mRNA level in the tumour tissue, positive values represent up regulation.

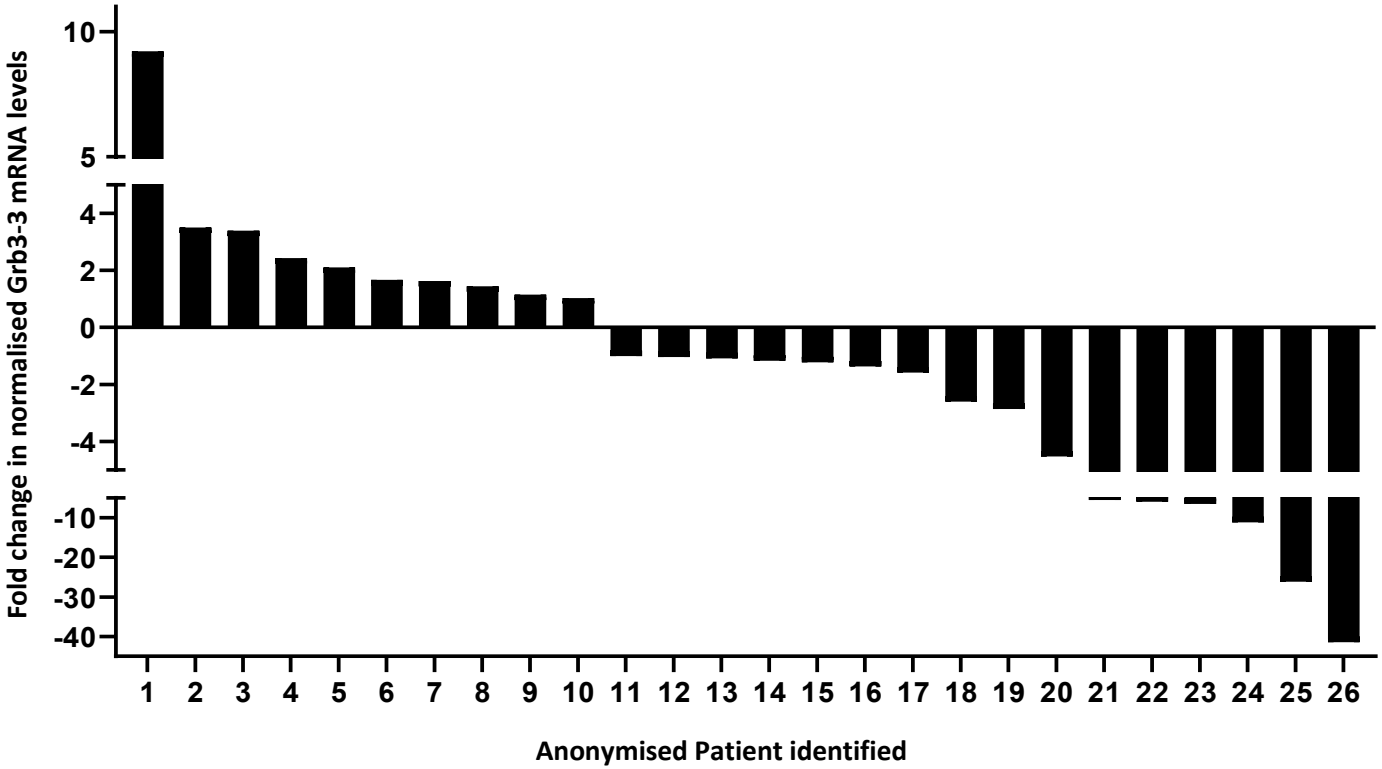


Figure 3.6 Fold change in normalised Grb3-3 mRNA levels in the tumour tissue relative to the normal. Waterfall plot showing the fold change in Grb3-3 mRNA abundance in tumour tissue relative to the associated normal for 26 colorectal cancer patients, anonymised patient identifiers are shown on the x-axis, negative values represent down regulation in mRNA level in the tumour tissue, positive values represent up regulation.

Figure 3.5 and 3.6 show that the levels of the 2 splice variants changes in colorectal cancer, and therefore the ratio of the alternatively spliced transcripts will also change. The ratio of the two transcripts was calculated by dividing the amount of each splice variant by the total expression (total expression describes the sum of Grb2 and Grb3-3 mRNA amount). The ratio of Grb2 transcript abundance as a proportion of total expression is plotted in Figure 3.7. A value of one 1 indicates that only the full-length transcript containing exon 4 is being expressed (Grb2), a value of 0.5 indicates that the two transcripts Grb2 and Grb3-3 are expressed equally and a value of 0 indicates that only the shorter splice variant Grb3-3 that lacks exon 4 is being expressed. In the normal tissue the Grb2 ratio ranged from 0 (not expressed) to 0.93, and in the tumour tissue the ratio ranged from 0.35-0.96. The median ratio for the population of patients analysed in this study was 0.66 in the normal colonic tissue and in the tumour tissue 0.78, therefore inclusion of exon 4 increased by 16% in colorectal cancer and skipping of exon 4 is reduced by 35%. Grb2 was the dominant transcript in both the normal colonic tissue and the tumour tissue as the average ratios were greater than 0.5 as for all but 5 patients. This is in agreement with previous reports that Grb3-3 is present at lower levels than Grb2 [89].

In the tumour tissue the spread of ratios observed is much less varied than in the normal tissue. This is likely to be due to heterogeneity of the samples, “normal” non-cancerous tissue is composed of epithelial cells, blood cells, immune cells and fibroblasts. All of these different cells were used for RNA extraction. While these cells are also present in the tumour tissue, these samples were macro-dissected prior to RNA extraction. This involves the removal of non-cancerous cells resulting in a purer sample. In addition, in the tumour tissue all the cells will adopt a proliferative phenotype whereas in the normal tissue some cells may be

senescent or apoptotic and therefore will have different splicing profiles. Overall, the analysis of *GRB2* splicing in colorectal cancer, using the splice sensitive PCR assay developed during this study, showed a decrease in exon 4 skipping in CRC. This resulted in reduced levels of Grb3-3 in the tumour tissue relative to the adjacent normal tissue and increased Grb2 levels. Grb3-3 has been reported to inhibit the expression of Ras responsive genes [89] and therefore may function as a negative regulator of Ras signalling. The affect the change in Grb2 splicing had on Ras signalling was investigated in a subset of patients using immunohistochemistry.

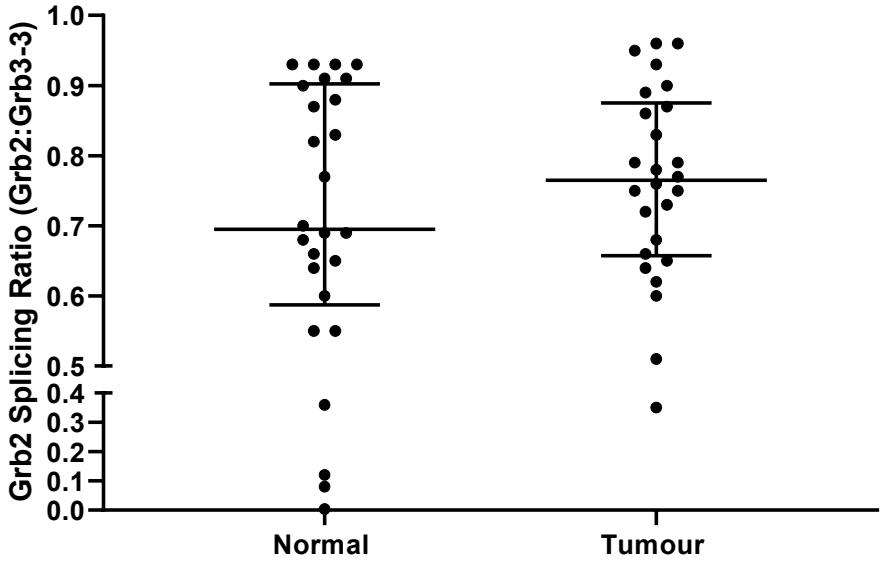


Figure 3.7 Scatter plot showing Grb2 splicing in colorectal cancer. Each black dot corresponds to an individual patient. The values correspond to the proportion of Grb2 transcripts as a ratio of total gene expression (Grb2 + Grb3-3 expression), black lines represent the median and interquartile range.

3.4 Analysis of Ras activation in colorectal cancer

Phosphorylation of serine 338 in Raf-1 (pRaf-1^{s338}) was used as a surrogate for Ras activation in colorectal tissue samples. Raf-1 interacts with GTP bound active Ras resulting in phosphorylation of multiple activatory residues including serine 338. The level of pRaf-1^{s338} was analysed by immunohistochemistry in 11 patients: 3 patients where Grb3-3 expression was reduced in the tumour tissue relative to the surrounding normal, 5 where there was less than 2-fold change and 3 where the levels of Grb3-3 increased in the tumour tissue (Figure 3.8). An inverse correlation between the levels of pRaf-1^{s338} and Grb3-3 expression was observed. Patients 25, 24 and 20 that had high levels of Grb3-3 in the normal colonic tissue (Figure 3.6) had low levels of pRaf-1^{s338}. Down-regulation of Grb3-3 expression in the tumour tissue by 26- ,11- and 5-fold respectively correlated with an increase in pRaf-1^{s338} levels, indicating that loss of Grb3-3 expression is sufficient to induce Ras-Erk signalling. In contrast patients 11, 10, 15, 16 and 12 that showed less than a 2-fold change in Grb3-3 expression (Figure 3.6) showed no change in pRaf-1^{s338} staining. Interestingly, these patients had very low levels of Grb3-3 expression in both the normal and tumour tissue and had high levels of pRaf-1^{s338} providing further evidence that Grb3-3 negatively regulates Ras signalling. Finally, patients where Grb3-3 expression increased in the tumour tissue (including patient 4, 3 and 5) the levels of pRaf-1^{s338} decreased, providing further evidence that Grb3-3 negatively regulates Ras signalling. A qualitative assessment of the level of pRaf-1^{s338} is shown in Table 3.1.

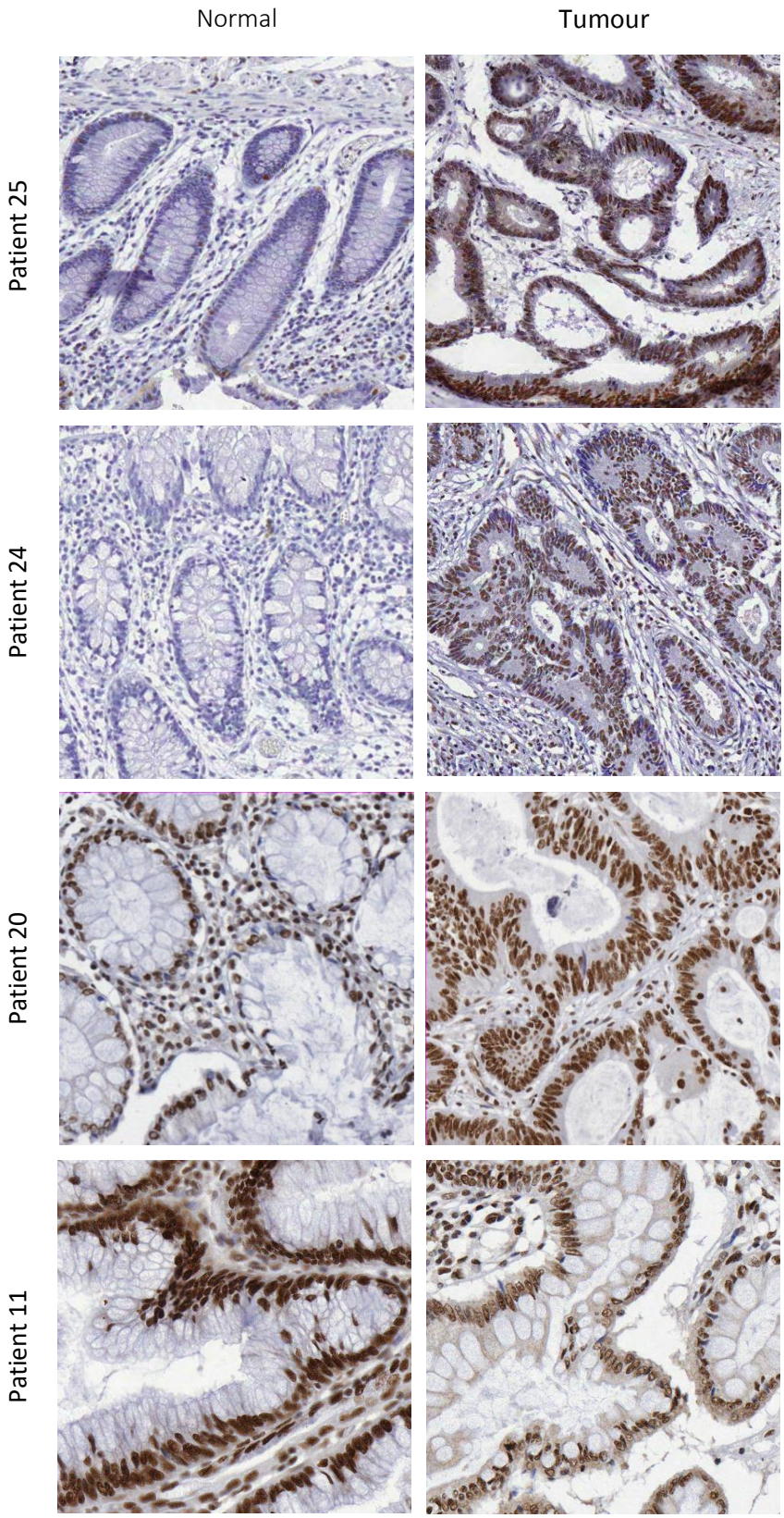


Figure 3.8 Representative immunohistochemistry images for colorectal cancer and associated normal tissues stained with anti pRaf-1^{s338} antibody (continued)

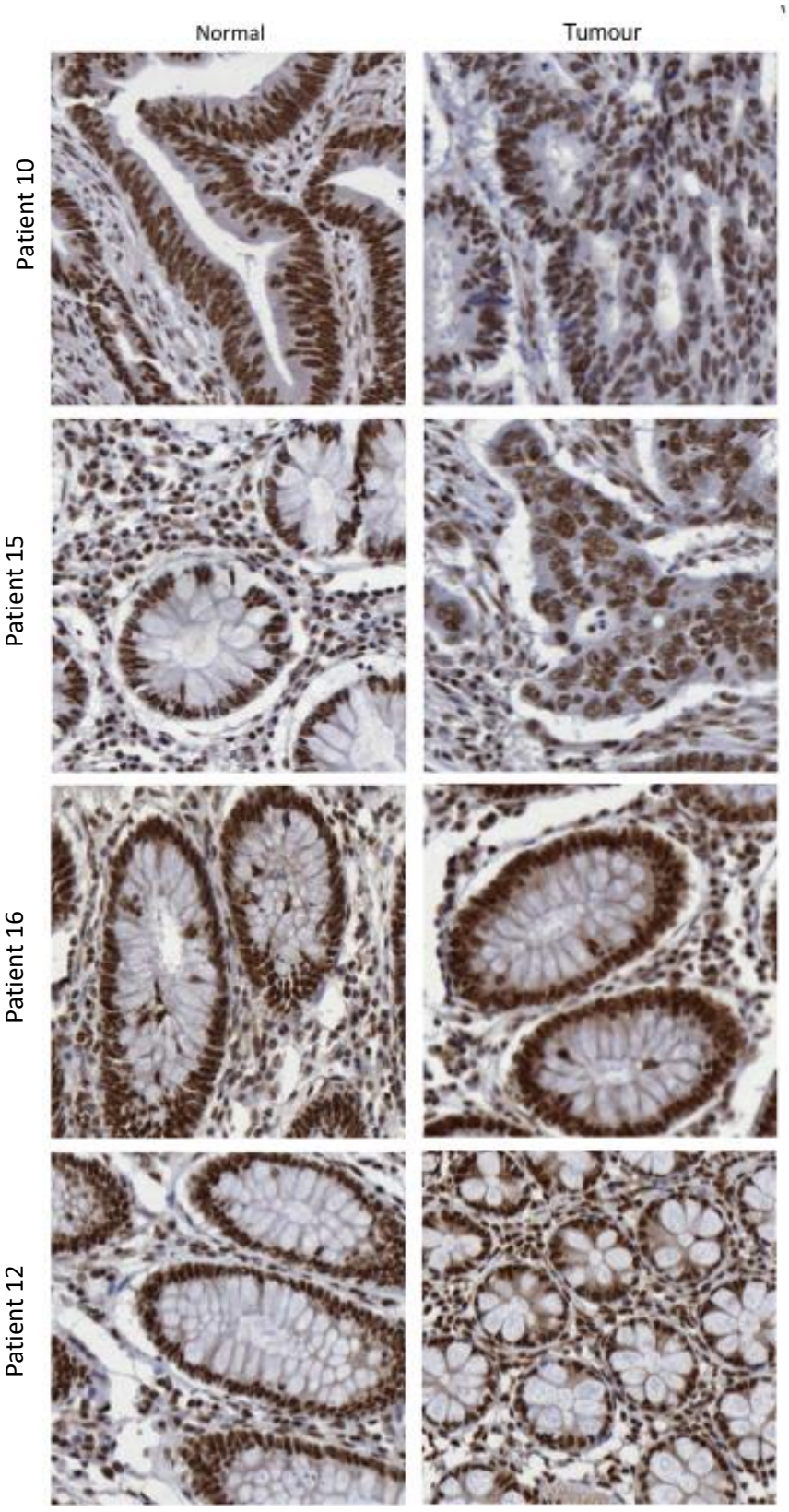


Figure 3.8 Representative immunohistochemistry images for colorectal cancer and adjacent normal tissues stained with anti pRaf-1^{s338} antibody (continued)

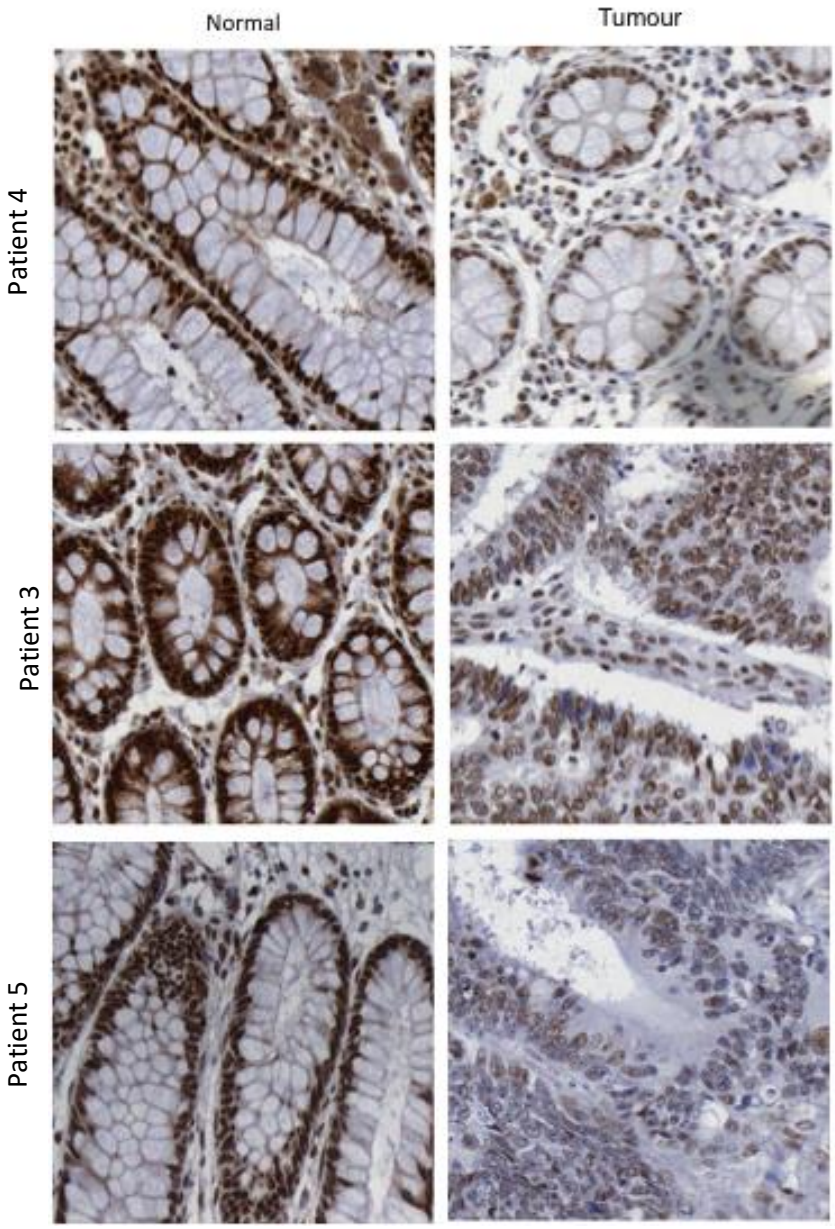


Figure 3.8. Representative immunohistochemistry images for colorectal cancer and adjacent normal tissues stained with anti pRaf-1^{s338} antibody. Images of tissue sections at 20x magnification, from normal colonic tissue (left) and tumour tissue (right). Brown staining represents pRaf-1^{s338} staining, purple staining is haematoxylin which shows the nucleus of cells.

Anonymised Patient Identifier	Normal	Tumour
25	Negative	++
24	Weak	+
20	+	+ (++ in some areas of mucosa)
11	++	++
10	+	+
15	++	+
16	+	+
12	++	++
4	++	+
3	++	++
5	++	+

Table 3.1 Qualitative analysis of pRaf-1^{s338} staining in colorectal cancer and associated normal tissues. Table shows the patient IDs and a qualitative analysis of the levels of pRaf-1^{s338} staining, weak or negative staining is stated, + denotes medium staining ++ represents strong staining**.**

Interestingly, the colonic crypts, which are tube-like structures that line the colonic mucosa (described in more detail in Section 1.5.1) showed a gradient of increased staining towards the base of the crypt (Figure 3.9). The crypts are highly organized monolayers of epithelial cells. At or near the base of the crypt is a stem cell population and their progeny are highly proliferating cells that occupy the lower third of the crypt. As the cells move up the crypt they differentiate forming mature epithelial cells. Colorectal cancers originate in the colonic crypt and then invade through subsequent tissue layers of the bowel wall. Higher levels of Raf-1 s338 phosphorylation were observed in the lower third of the colonic crypts, where the cells are known to be proliferating (Figure 3.9). The Ras-Erk signalling cascade is well established as controlling cell proliferation suggesting that up-regulated Ras signalling in these cells contributes to hyper-proliferation and that changes in *GRB2* splicing may be a persuasive event driving hyper-proliferation and tumour development.

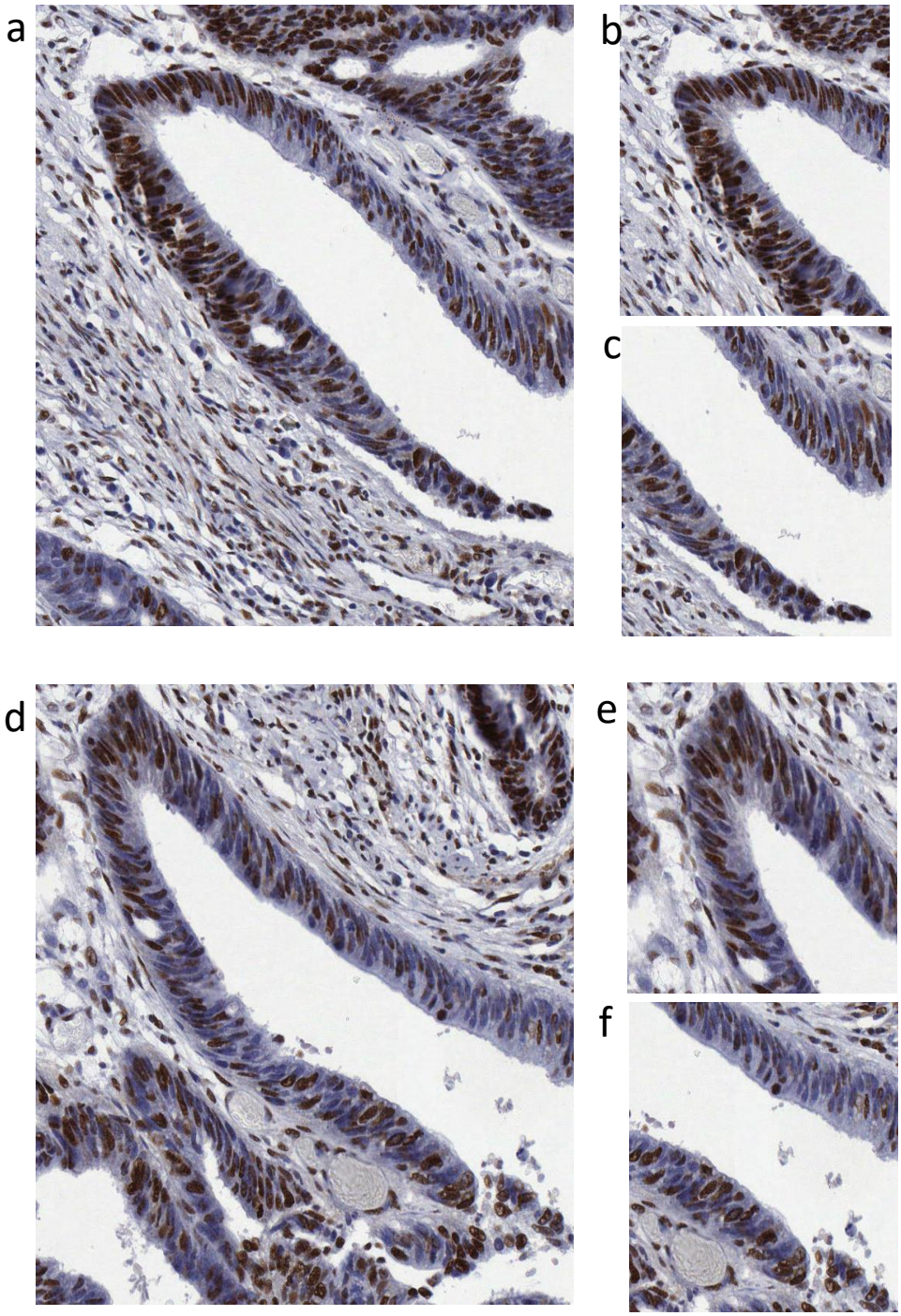


Figure 3.9 Representative images of colonic crypts stained for pRaf-1^{s338}. Images taken from tumour tissue of patient 26493 at 20x magnification, (a & d) entire colonic crypt (b & e) lower third of the colonic crypt, (c & f) upper third of the colonic crypt

3.5 Analysis of *GRB2* splicing in a panel of tumours and cell lines

Data presented in Section 3.3 and 3.4 highlight a novel link between altered *GRB2* splicing and Ras activation in colorectal cancer. Upregulated Ras signalling is a common feature of many different tumours. To determine if altered *GRB2* splicing occurs in other tumours Grb2 and Grb3-3 expression were investigated in a panel of tumour tissues and cell lines (Figure 3.10). In breast, kidney liver and ovarian cancer Grb2 was the dominant transcript and Grb3-3 was expressed at low levels. However, the expression in the adjacent normal tissue is unknown. In lung cancer Grb3-3 was expressed at equivalent levels to Grb2. This was previously reported to induce apoptosis, suggesting Grb3-3 may also have a regulatory role in lung cancer [89]. Interestingly in hepatocellular carcinoma (HCC) matched tumour and associated normal tissue samples were obtained from the same patient. In the normal liver tissue Grb3-3 was the dominant transcript, whereas in the tumour tissue Grb2 was the dominant transcript and expression of Grb3-3 was down regulated in HCC by 109 fold. This is consistent with the trends observed in CRC suggesting that changes in *GRB2* splicing may also be important in the development of HCC. The Ras signalling pathway is up-regulated in 50-100% of HCCs and is associated with worse prognosis [170-172]. Importantly up-regulation is reported to occur through loss of endogenous inhibitors of Ras-Erk signalling and not through mutations in Ras/Raf genes which are rare in HCC [173, 174]. This is consistent with the hypothesis that Grb3-3 functions as an endogenous inhibitor of Ras signalling, and loss of Grb3-3 represents a novel route for up regulated Ras signalling in colorectal cancer. However, further work is required to determine if there is a correlation between Grb3-3 levels and Ras activation in HCC.

Cell line analysis showed Grb3-3 was not expressed in Hela or Pnt1a cells, but was present in all other cell lines at 2-45 fold lower levels than Grb2. This data suggests that while Grb2 expression is ubiquitous, the splice variant Grb3-3 is expressed in a cell or tissue specific manner.

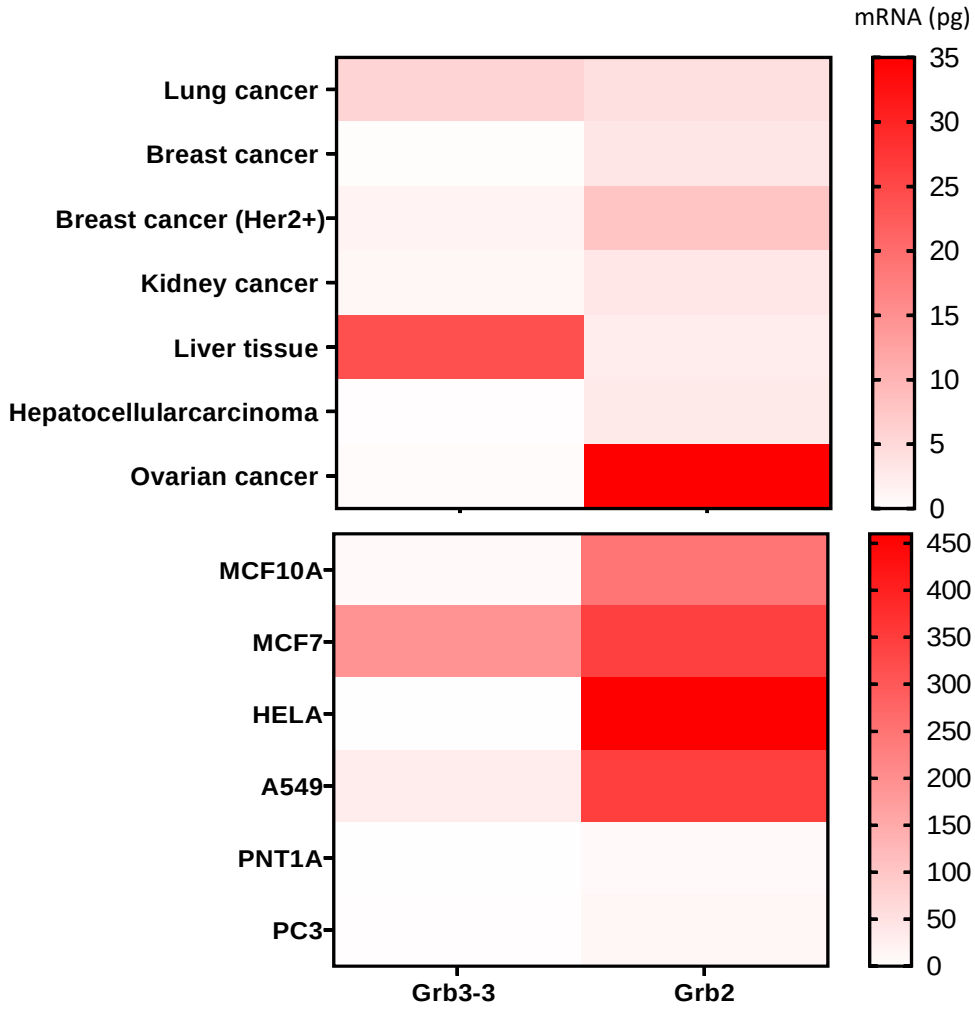


Figure 3.10 Grb2 and Grb3-3 isoform expression in a panel of tumour tissues and cell lines. Heat map showing the amount in picograms of Grb3-3 and Grb2 mRNA present in a panel of tissues (top) and cell lines (bottom), the higher the expression the darker red the bands are, sample IDs are shown on the vertical axis

3.6 Summary

This Chapter reveals that splicing of the adaptor protein Grb2 is altered in a subset of colorectal cancer cases and may also be disrupted in hepatocellular carcinoma. In CRC reduced skipping of the alternate exon 4 in the tumour tissue resulted in reduced levels of Grb3-3 and increased levels of Grb2. This correlated with increased activation of Ras signalling, suggesting that Grb3-3 has antagonistic functions to Grb2 in the Ras-Erk signalling cascade. Therefore the ratio of the two alternate transcripts may function to maintain homeostatic Ras signalling, thus a change in this ratio, due to changes in *GRB2* splicing may represent a novel route for oncogenic Ras signalling.

Aberrant alternate splicing is known to be a pervasive event in the development of cancer and an increasing number of AS events have been linked to colorectal cancer development and progression [40, 159]. This is the first study showing that *GRB2* splicing is altered in colorectal cancer. Analysis of a larger number of patient samples is required to determine the frequency that this splicing event is disrupted. Preliminary data suggests that *GRB2* splicing may also be altered in hepatocellular carcinoma, another Ras driven tumour. Further work is required to determine the molecular role of the splice variant Grb3-3 in Ras signalling. This may reveal novel strategies to therapeutically target Ras signalling in CRC directed towards correcting *GRB2* splicing.

Chapter 4: The role of Grb3-3 in Erk signalling

4.1 Introduction

Data presented in Chapter 3 revealed that alternate splicing of the adaptor protein Grb2 is altered in colorectal cancer. Expression of the truncated splice variant Grb3-3 that lacks exon 4 was reduced in CRC in 62% of the patients analysed. Reduced expression of Grb3-3 in the tumour tissue correlated with increased activation of the Ras effector Raf-1, while increased levels correlated with reduced activation. This suggests that the Grb2 splice variant functions as a negative regulator of Ras signalling. These observations are consistent with previous reports from Fath et al [89] that Grb3-3 inhibits EGF induced transactivation of a Ras responsive element in Chinese hamster lung fibroblasts. However, direct evidence that Grb3-3 negatively regulates Ras activation or a molecular mechanism for how it does so is lacking. In this Chapter a combination of biophysical, molecular and cellular biological approaches were adopted to understand the role of the truncated splice variant Grb3-3 in the Ras-Erk signalling cascade.

4.2 Grb3-3 inhibits growth factor induced Ras-Erk signalling

To investigate whether Grb3-3 has any effect on growth factor-induced Ras-Erk signalling, Hek293t cells were transfected with exogenous RFP-tagged Grb3-3 or RFP alone (control). Following transfection, cells were serum-starved for 24 hours prior to stimulation with various growth factors to induce signalling. Grb3-3 inhibited growth factor-induced Erk activation shown by reduced levels of pErk in Grb3-3 expressing cells (Figure 4.1). This inhibition was not specific to the growth factor receptor stimulated upstream of Erk, suggesting that Grb3-3 acts on the signalling pathway at a point downstream of receptor activation where all three

growth factor receptors converge. The proportion of total Erk that is in the active phosphorylated form is quantified in Figure 4.1b, allowing the level of inhibition by Grb3-3 to be determined. Grb3-3 inhibited Erk activation downstream of FGFR2 by 68%, downstream of the Met receptor by 80% and downstream of the IFN-λ receptor by 71%.

To determine if the inhibitory effect of Grb3-3 on Erk activation also occurs in colorectal cancer cell lines, (suggested by the tissue analysis in Chapter 3), RFP-Grb3-3 was transfected into human Caco2 colorectal adenocarcinoma cells (isolated from a 72 year old male). Erk signalling was induced by stimulation with the epidermal growth factor (EGF) as these cells express high levels of the EGF-receptor. Consistent with the effect observed in Hek293t cells, overexpression of Grb3-3 in Caco2 cells inhibited EGF-induced Erk activation (Figure 4.2). The level of inhibition was approximately 30%, much lower than observed in Hek293t cells (Figure 4.2b). Differences in the level of inhibition between the two cell lines are likely to be due to differences in the transfection efficiencies (Figure 4.3). Approximately 80% of Hek293t cells expressed RFP-Grb3-3 in comparison to around 20% of Caco2 cells.

Immunoblotting for phosphorylation of tyrosine at position 1068 (pY1068) in the EGF-receptor was used to determine the effect of Grb3-3 on EGFR activation (Figure 4.2a). pY1068 is found in the c-terminus of EGFR and is trans-phosphorylated by the receptor in response to EGF stimulation. Additionally pY1068 and a pY1086 form the binding site for Grb2 SH2 domain [175]. Immunoblotting for pY1068 on the EGF receptor showed that Grb3-3 did not inhibit EGFR activation confirming that the splice variant inhibits Erk signalling at point downstream of the receptors.

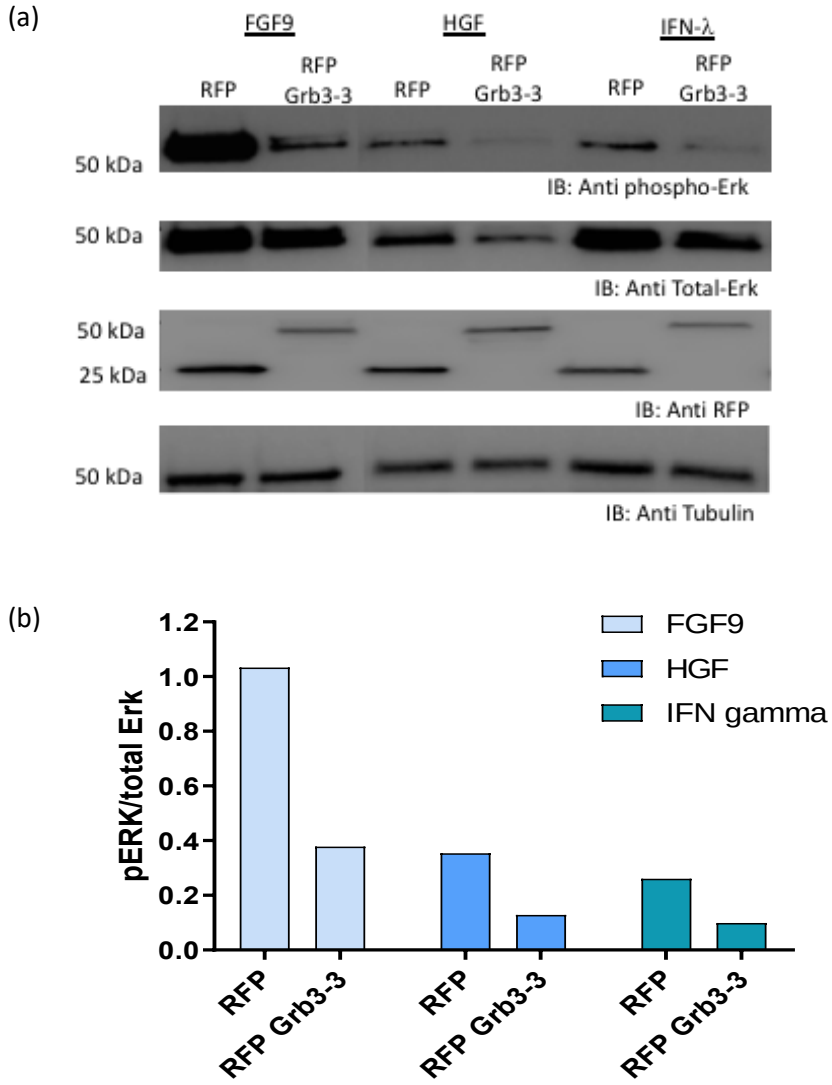
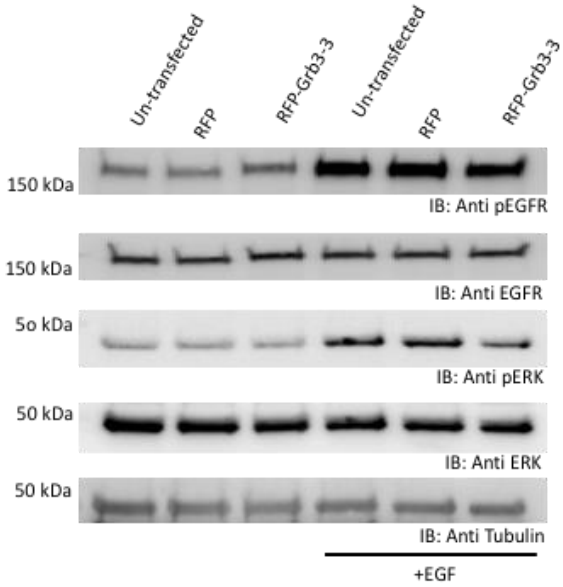


Figure 4.1 Effect of Grb3-3 on growth factor induced Erk activation in Hek293t cells (a) Immunoblots of Erk activation in Hek293 cells overexpressing RFP-Grb3-3 or RFP (control) stimulated with FGF9, HGF and IFN-λ (each condition represents a single experiment). Immunoblots were probed with antibodies against RFP to assess transfection, tubulin to assess total protein, pErk and total Erk to measure activation of the Erk signalling pathway (b) Quantification of the relative amount of total Erk that is phosphorylated, calculated using Image J, Erk and pErk levels are normalised separately to tubulin.

(a)



(b)

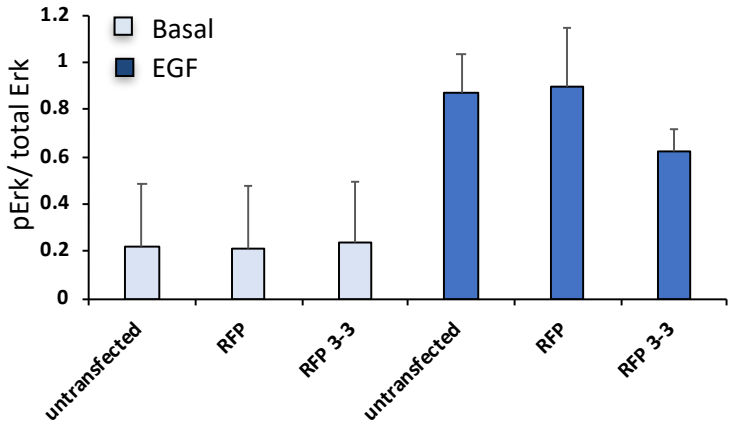


Figure 4.2 Effect of Grb3-3 on EGF induced Erk activation in caoc2 cells (a) Immunoblots of Erk activation in Hek293 cells overexpressing RFP-Grb3-3 or RFP (control) stimulated with FGF9, HGF and IFN-λ. Immunoblots were probed with antibodies against RFP to assess transfection, tubulin to assess total protein, pErk and total Erk to measure activation of the Erk signalling pathway and pEGFR/total EGFR to measure receptor activation. (b) Quantification of the relative level of phospho-Erk calculated using Image J, data represents the mean plus standard deviation for biological duplicates.

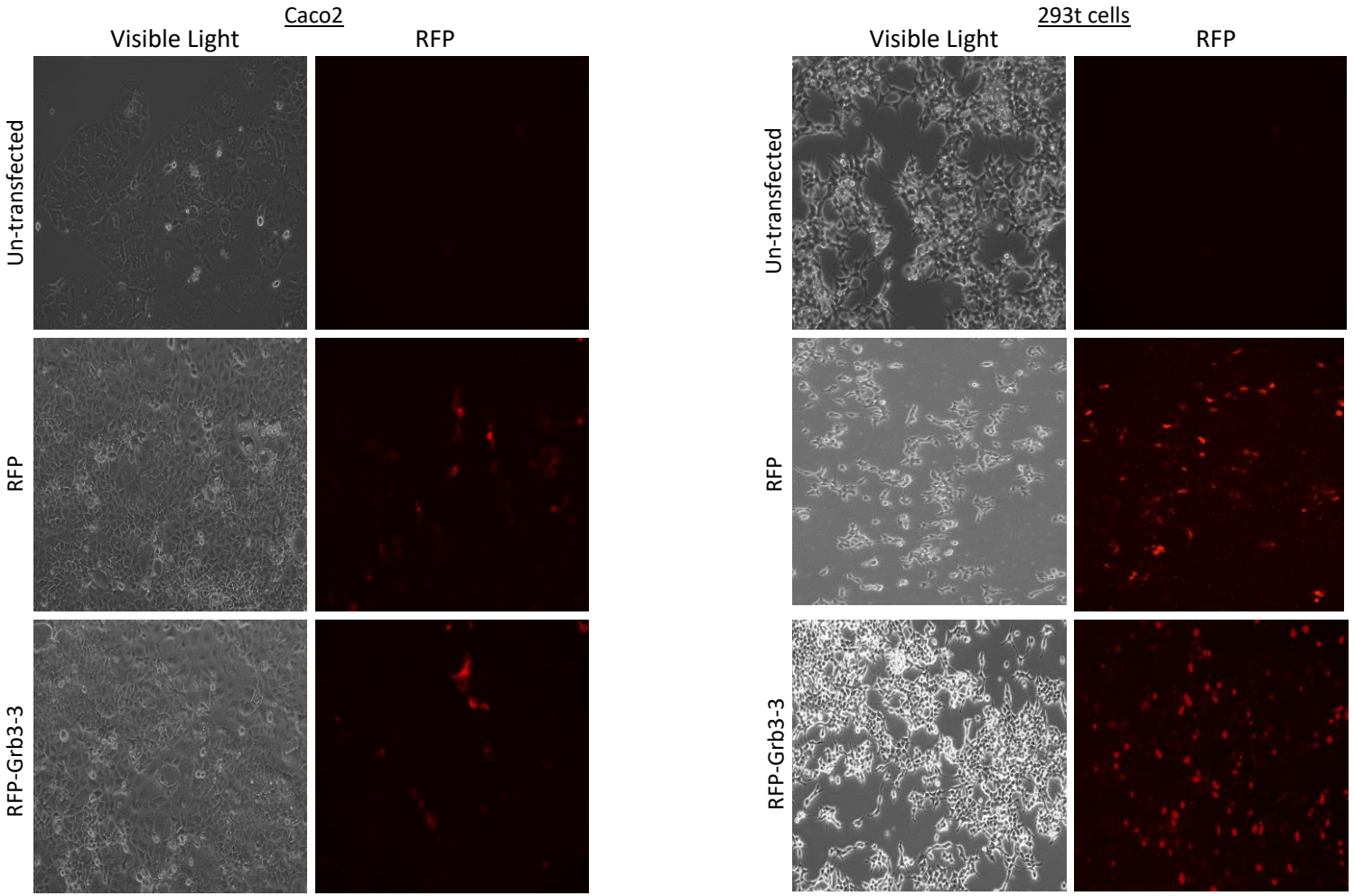


Figure 4.3 Transfection efficiency of Caco2 and Hek293t cells. Panel on the left shows the transfection of RFP/RFP Grb3-3 in the Caco2 cells, and on the right shows the Hek293t cells.

The Ras proteins are activated downstream of activated RTKs and upstream of the Raf-Mek-Erk signalling cascade. Therefore the effect of Grb3-3 on Ras activation was investigated. In cells the Ras proteins act as molecular switches cycling between inactive GDP-bound states and active GTP-bound states controlled by the opposing action of guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs). The levels of active GTP-Ras in cells was measured using a pull-down assay with GST-tagged Raf1 Ras binding domain (RBD). The Raf proteins only bind to Ras in its GTP-bound active form and the RBD corresponding to residues 1-149 is the minimal region required for this interaction. Grb3-3-inhibited Ras activation resulting in reduced levels of GTP-Ras in Grb3-3 expressing cells (Figure 4.4). Grb3-3 resulted in complete inhibition of EGFR- and IFN-λ-induced Ras activation (Figure 4.4b) but did not completely inhibit Erk activation (Figure 4.1 and 4.2). The low level of Erk activation observed in Grb3-3 expressing cells may be due to Ras-independent activation of Erk signalling, for example protein kinase A (PKA) and C (PKC) can directly activate the Raf-Mek-Erk signalling cascade bypassing Ras activation.

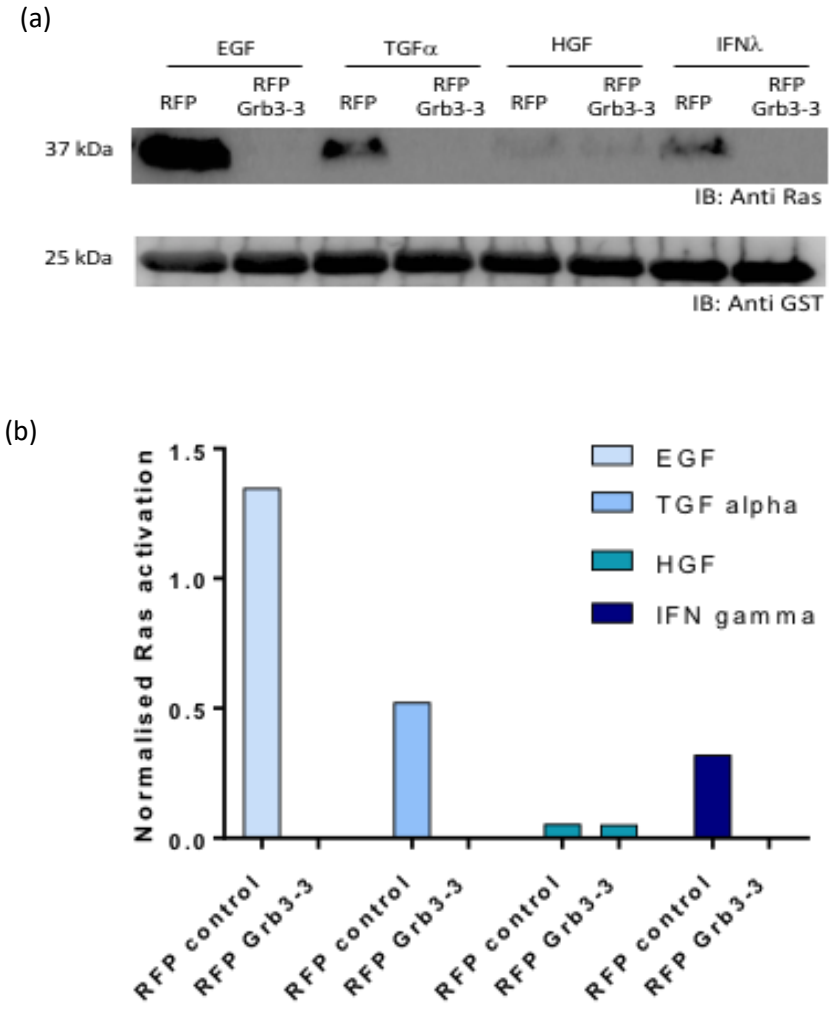


Figure 4.4 Effect of Grb3-3 on growth factor induced Ras activation (a) Western blots showing the amount of active Ras immunoprecipitated with GST-RBD. Anti-GST was used as control to show that equal amounts of GST-RBD were used in the pull-down. (b) Quantification of the amount of Ras bound to GST-RBD in the presence and absence of Grb3-3 following stimulation with various growth factors. Data represents a single experiment.

To confirm that Grb3-3 inhibits Ras activation, and that loss of Grb3-3 expression is responsible for the increased Raf1 phoshorylation observed in the colorectal cancer tissue samples (Chapter 3), the effect of Grb3-3 on Raf-1 activation was analysed. Upon formation of the GTPRas.Raf complex, shown to be inhibited by Grb3-3 in Figure 4.4, the Raf proteins become phosphorylated on multiple residues including serine 338 by an as of yet unknown kinase. Overexpression of strep-tagged Grb3-3 in Hek293t cells reduced pRaf-1^{s338} levels in response to FGF9 stimulation (Figure 4.5), providing further evidence that Grb3-3 inhibits Ras activation.

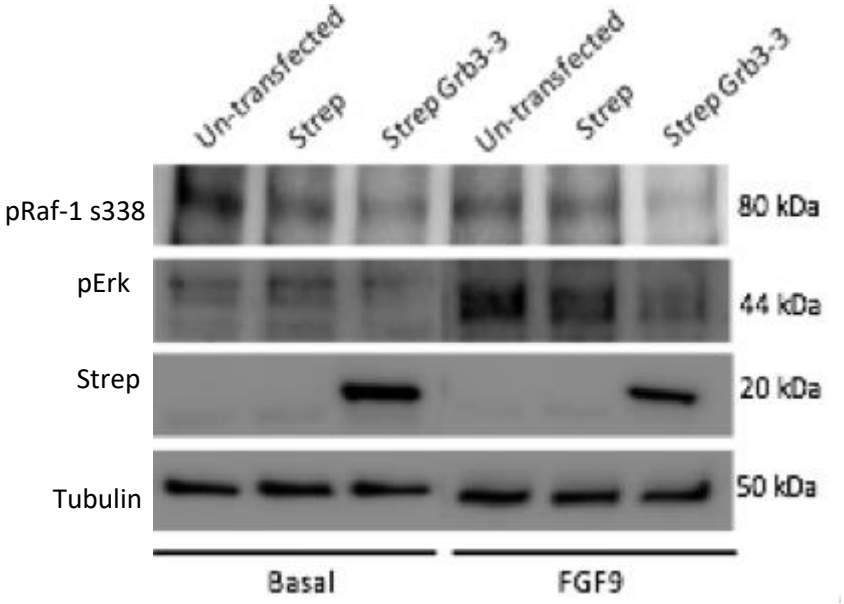


Figure 4.5 Effect of Grb3-3 on growth factor induced Raf activation. Immunoblots of Raf activation in Hek293 cells overexpressing Strep-Grb3-3 or Strep (control) stimulated with FGF9. Immunoblots were probed with antibodies against Strep to assess transfection, tubulin to assess total protein, pRaf and pERK to measure Raf activation and downstream Erk signalling.

4.3 Characterising the Grb3-3 Sos interaction

The Sos proteins are the major guanine nucleotide exchange factors responsible for catalysing Ras activation. They bind to inactive GDP Ras and induce a conformational change in the nucleotide-binding region that promotes release of GDP. GTP is present in the cell at much higher levels and therefore associates with nucleotide free Ras activating it [99, 105]. In order to gain a deeper understanding of the molecular mechanism for inhibition of Ras activation by Grb3-3 its ability to bind Sos was investigated.

Binding of Grb3-3 to endogenous Sos1 was assessed in Hek293t cells. Cells were transfected with RFP-tagged Grb3-3 or the truncated protein binding domains: NSH3-SH2$\Delta_{40,}$ SH2Δ_{40} and SH2Δ_{40}-CSH3 (Δ_{40} represents the 40 residue deletion in the SH2 domain). RFP-tagged proteins were precipitated under serum-starved and stimulated conditions using RFP trap A beads, and co-immunoprecipitation of endogenous Sos was assessed via immunoblotting (Figure 4.6). Co-immunoprecipitation of Sos by Grb3-3 under both non-stimulated and stimulated conditions shows that Grb3-3 is constitutively bound to Sos. This differs to the previously reported Grb2-Sos interaction which is inhibited under non-stimulated conditions by Grb2 dimerization [103]. Under non-stimulated conditions binding of Grb3-3 appeared to be mediated by the N-terminal SH3 domain whereas under stimulated conditions binding occurred through the C-terminal SH3 domain, suggesting there are two distinct binding sites for Grb3-3 on Sos.

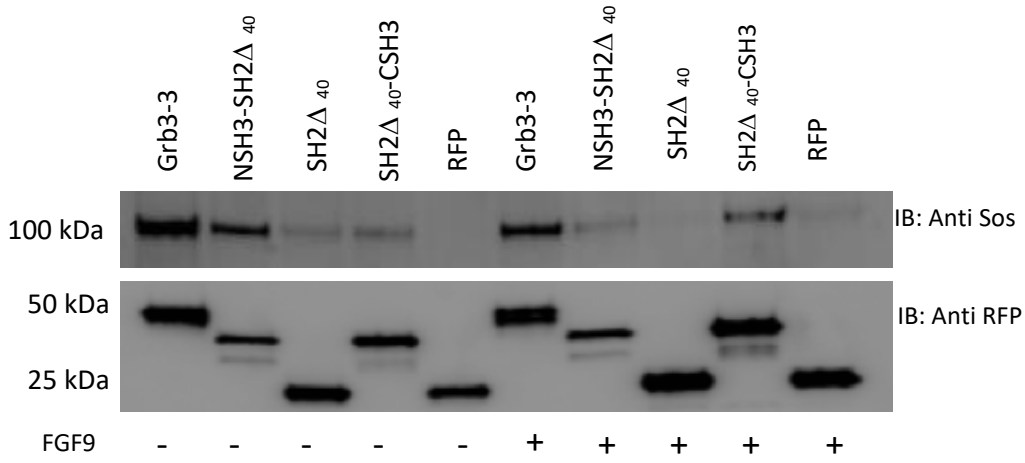


Figure 4.6 Characterisation of the Grb3-3-Sos interaction in cells. Hek293t cells stably expressing FGFR2 were transfected with RFP-tagged Grb3-3 constructs. The RFP proteins were immunoprecipitated using RFP Trap A beads from cells under serum starvation (lanes 1-5) and FGF9 stimulation (lanes 6-10). Anti-RFP was used as a control to show equal expression of the constructs. Bound endogenous Sos was measured using the anti-Sos antibody.

Fluorescence resonance energy transfer (FRET) was used to confirm the constitutive interaction between Grb3-3 and Sos. RFP-tagged Grb3-3 and GFP-tagged Sos were co-transfected into Hek293t cells and FRET was measured under serum-starved and EGF stimulated conditions. FRET occurs upon binding of the two fusion proteins because excitation of GFP produces an emission spectrum that overlaps the excitation wavelength of RFP, and therefore when in close proximity energy can be transferred between the two chromophores. FRET analysis confirmed Grb3-3 is constitutively associated with Sos (top panel, Figure 4.7). FRET was also observed between GFP-Sos and RFP-tagged NSH3-SH2Δ_{40} or SH2Δ_{40}-CSH3 domains confirming that binding to Sos is mediated by both the N and C-terminal SH3 domains of Grb3-3 (middle and bottom panels, Figure 4.7) although no difference in binding of the two SH3 domains was observed between serum-starved and stimulation conditions.

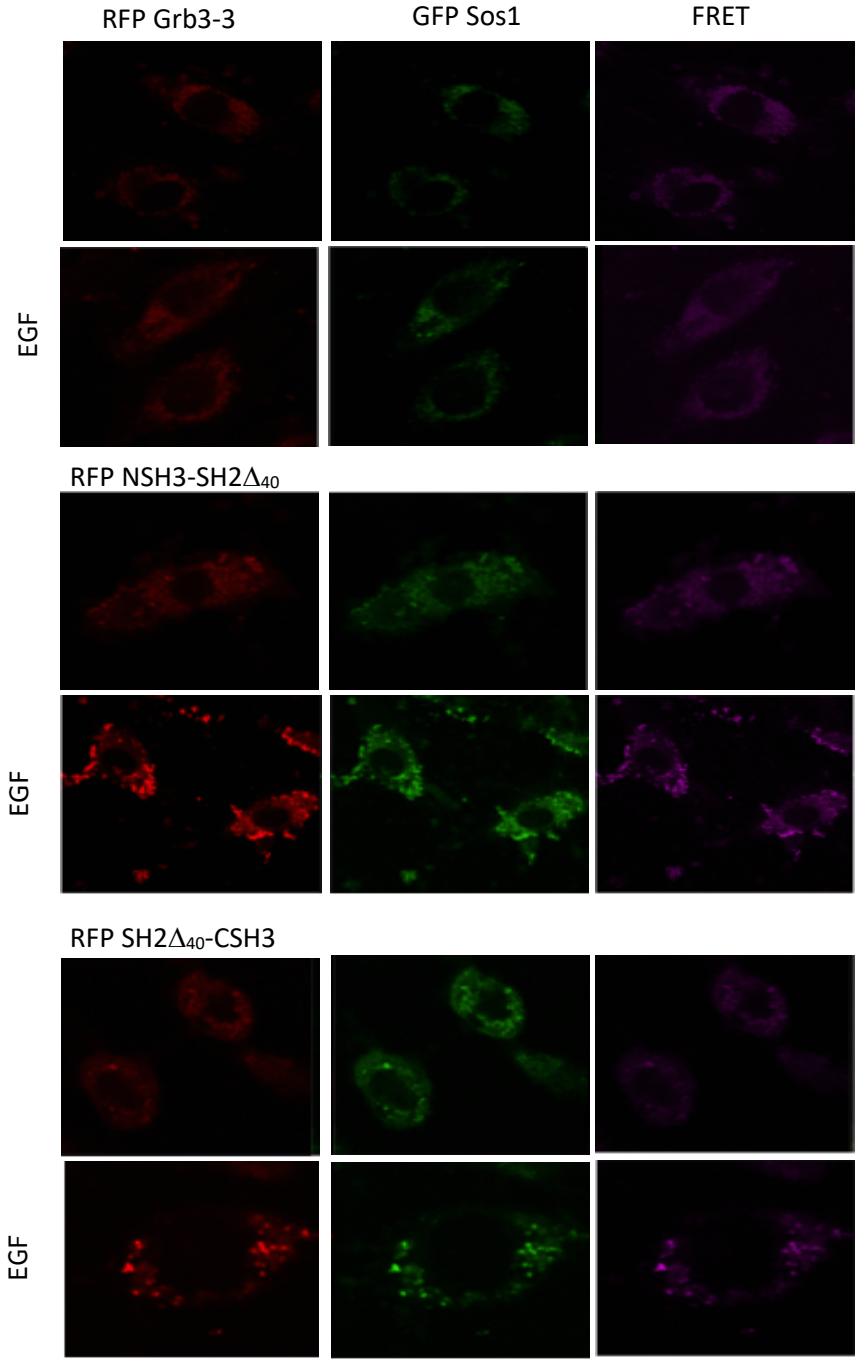


Figure 4.7 Representative fluorescent images of the interaction between Grb3-3 and the individual domains with Sos in Hek293t cells. Green fluorescence shows GFP tagged Sos expression, red fluorescence shows RFP tagged Grb3-3 (top panel) NSH3 SH2Δ_{40} domain (middle panel) and SH2Δ_{40}-CSH3 (bottom panel) expression. FRET is shown in purple. Cells that were stimulated with EGF are labelled on the left.

SH3 domain-mediated protein interactions typically occur through the recognition of proline-rich motifs. Analysis of the amino acid sequence of Sos1 identified 14 proline-rich motifs predominantly in the C-terminal tail, but also in the PH and CDC25 domains. To identify the site of interaction on Sos more precisely, pull down experiments were performed using GST-tagged proline-rich peptides derived from Sos. Sequences of the proline-rich peptides are shown in Figure 4.8a. The ability of the GST-tagged peptides to bind to purified His-tagged- Grb3-3 was assessed and compared to binding of the full-length isoform Grb2 (Figure 4.8b). No binding was observed in the GST control sample, meaning any binding to Grb3-3 and Grb2 occurs via a specific interaction with the peptide and not the GST tag. Only Grb3-3 bound to Sos peptides 2 and 3, derived from the catalytic CDC25 domain of Sos. Both Grb3-3 and Grb2 bound to Sos peptides 5, 6 and 8 derived from the C-terminus (Figure 4.8b).

(a)

Sos1	MICCKSNHGQPRLPGASNAEYRLK
Sos2	DHYKKYLAKLRSINPPCVPFFGIYLTNILKTE
Sos3	TDYLFNKSLEIEPRNPKPLPRFPKKYSYPLKSPGVRPSNPRPGTM RHPTPLQQEP
Sos4	ESETESTASAPNSPRTPLTPPPASGASSTTDVCSVFDSDHSSPFH SSNDTVFIQVTLPHGP RSASVSSISL
Sos5	TKGTDEVPVPPPVPPRRRPESAPAESSPSKIM
Sos6	SKHLDSPPAIPPRQPTSKAYSPRYSISDRTSISDPPESPPLLPPREP VRTPDVFSSSP
Sos7	LHLQPPPLGKKSDHGNAFFPNSPSPFTPPPPQTPSPHGTRRHLP SPPLTQEVDLHSI
Sos8	EVDLHSIAGPPVPPRQSTSQHIPKLPPKTYKREHTHPSMH

(b)

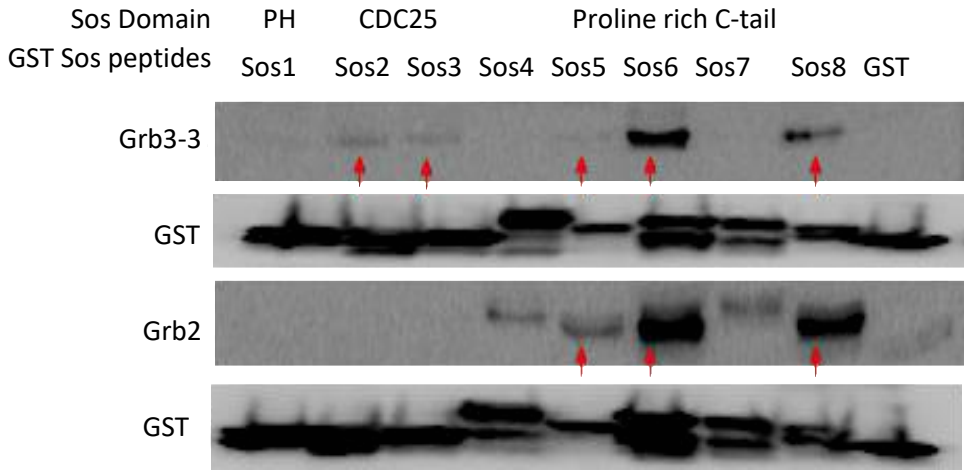


Figure 4.8. Binding of Grb3-3 and Grb2 to proline-rich peptides from Sos C-terminal tail. (a) Amino acid sequence of Sos proline-rich peptides, proline-rich motifs are underlined (b) Immunoblot showing binding of GST-Sos peptides to purified His-tagged Grb2 and Grb3-3

4.3.1 The Grb3-3 CDC25 interaction

Microscale thermophoresis (MST) was used to further characterise the unique interaction between Grb3-3 and proline-rich motifs in the catalytic CDC25 domain of Sos. MST is a powerful, biophysical technique that can be used to determine the affinity of biological interactions by measuring the change in fluorescence (intrinsic or extrinsic) of a sample in response to a change in temperature of a highly localised position in the solution, the fluorescent properties of proteins and their movement in a temperature gradient are altered upon the formation of complexes. MST was used to determine if Grb3-3 was able to bind to the proline-rich motifs identified in the GST pull-down in the context of the full CDC25 domain corresponding to residues 564-1049 of Sos. This domain was recombinantly expressed in E. coli and fluorescently labelled with Atto488. The binding to increasing concentrations of unlabelled Grb3-3 was measured using MST (Figure 4.9a). MST experiments confirmed a direct interaction between Grb3-3 and Sos CDC25 domain. Two separate binding events were observed with affinities of 7.9 μM and 238 μM suggesting that Grb3-3 can bind to the CDC25 domain at two sites (Figure 4.9a). This is consistent with the pull down that showed binding to two separate proline-rich peptides. To identify the precise binding site for Grb3-3 within the CDC25 domain three proline-rich peptides (shown in Table 4.9b) were designed encompassing the individual proline-rich motifs in this domain, and MST was used to assess the binding of Grb3-3 to these peptides (Figure 4.9c). Grb3-3 bound to peptide 1 (SosCat1) corresponding to residues 917-941 of Sos with a high affinity of 2.95 μM and to peptide 3 (SosCat3) with an affinity of 152 μM resembling the two binding events observed with the entire CDC25 domain. The high affinity interaction between Grb3-3 and Sos-Cat1 was confirmed by isothermal titration calorimetry (ITC). The peptide was titrated into Grb3-3 and

an equilibrium dissociation constant of 1.94 μM was measured which closely resembles the affinity determined using MST (Figure 4.9d).

Interestingly the binding site for Grb3-3 on Sos CDC25 domain has been previously reported to form the allosteric Ras binding site, specifically leucine 919, isoleucine 922 and asparagine 923 are essential for the interaction with allosteric Ras. Furthermore, the proline residues at position 924 and 925, suspected to be binding to Grb3-3, have been previously reported as being essential for orientating these residues in a conformation permissive for Ras binding [109]. Binding of GTP-Ras at the allosteric site in Sos increases the affinity of the interaction between Ras and the catalytic site. It also increases the rate of nucleotide exchange, creating a positive feedback loop [109]. The interaction between Grb3-3 and the Sos CDC25 domain may function to block allosteric Ras binding and inhibit the positive feedback activation of Sos, however this was not investigated further in this thesis.

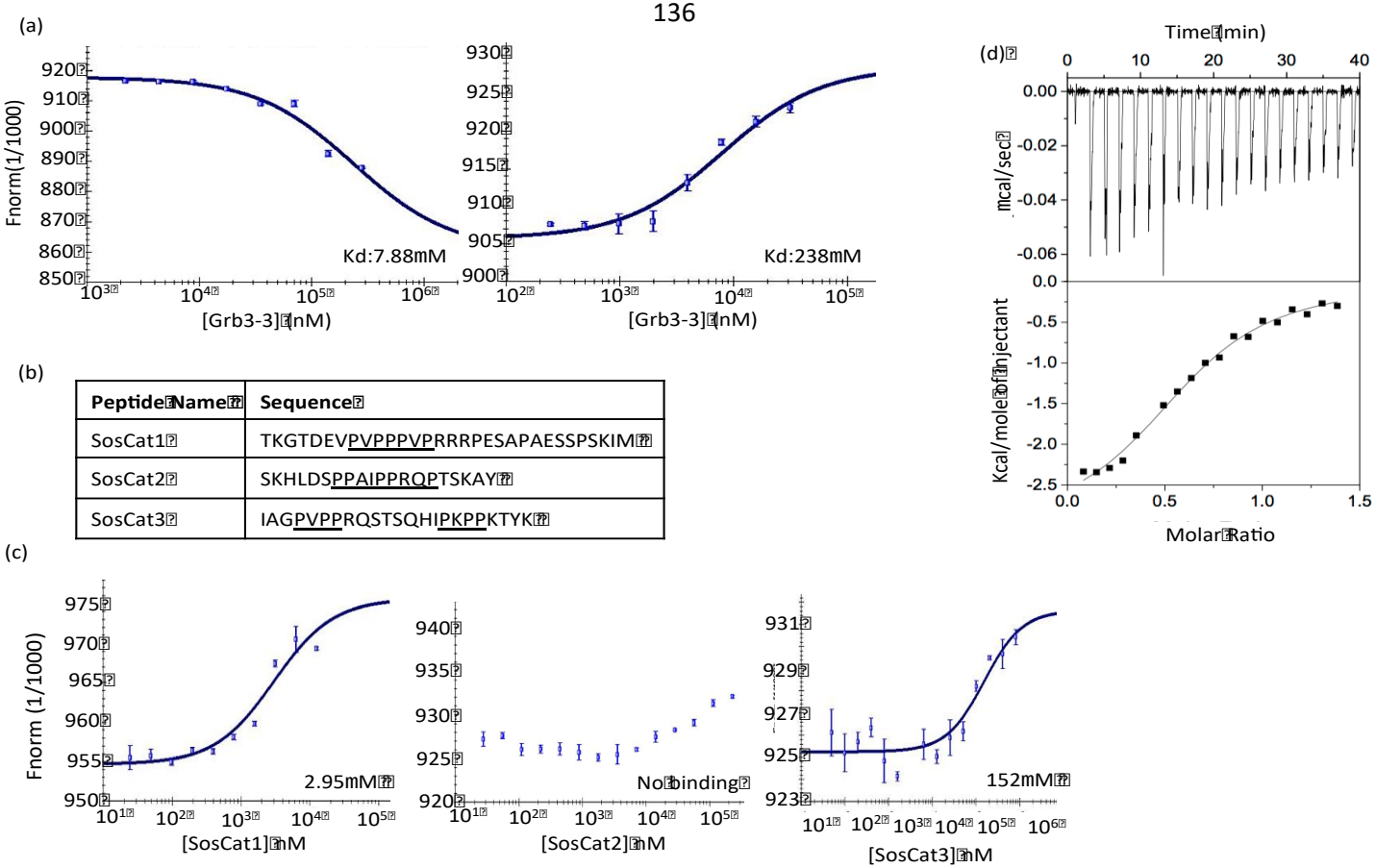


Figure 4.9. Biophysical characterisation of Grb3-3 Sos CDC25 interaction (a) MST graph showing two binding events for the interaction between purified His-tagged Grb3-3 and the fluorescently labelled CDC25 domain of Sos. (b) Sequences of proline-rich peptides derived from Sos CDC25 domain. (c) MST graph showing binding of fluorescently labelled His-tagged Grb3-3 to peptides derived from Sos CDC25 domain. All MST data plotted represents the mean plus standard deviation for 3 technical repeats (d) ITC binding isotherm for the interaction between Sos Cat peptide 1 and Grb3-3

4.3.2 Characterising the interaction with Sos proline-rich tail

Pull down analysis revealed that both Grb2 and Grb3-3 were able to bind to three proline-rich peptides derived from the c-terminus of Sos (Figure 4.8). Therefore in cells where both protein isoforms are expressed competition for binding to c-tail of Sos may occur. The affinities of the interactions between the two isoforms and the c-terminus of Sos were measured using MST. The sequences of the c-terminal proline-rich peptides used in the MST experiments are shown in Table 4.1

Peptide Name	Sequence
Sos-Ctail-1	TKGTDEVPVPPPVPRRRPESAPAESSPSKIM
Sos-Ctail-2	SKHLDSPPAIPPRQPTSKAY
Sos-Ctail-3	IAGPVPPRQSTSQHIPKPPKTYK

Table 4.1 Sequences of proline-rich peptides derived from Sos C-terminus used in MST studies. Amino acid sequences of each peptide are shown with the proline-rich motifs underlined.

MST experiments confirmed a direct interaction between both isoforms and the three peptides derived from the C-terminus of Sos (Figure 4.10). The affinities of the interactions ranged from 5-84 μM and both protein isoforms bound with similar affinities to each peptide. This suggests that in the cellular environment where both protein isoforms are present, the ability to bind to the C-terminus of Sos will be largely influenced by the concentrations and cellular localisation of the two isoforms. It is unknown if in the context of full-length Sos all of the proline-rich motifs would be exposed for binding, or if Grb3-3 and Grb2 preferentially bind to a single site. Both Grb2 and Grb3-3 bound with highest affinity to Sos-Ctail1 peptide, this peptide contains the canonical PVPP binding motif. With the exception of the interaction between Grb2 and the Sos-Ctail3 peptide all binding curves were characteristic of a single binding event. Grb2 showed two site binding to the Sos-Ctail3 peptide, this peptide had 2 separate proline-rich motifs, the two binding events were fitted separately to a high affinity interaction (1.51 μM) and a low affinity interaction (41.7 μM).

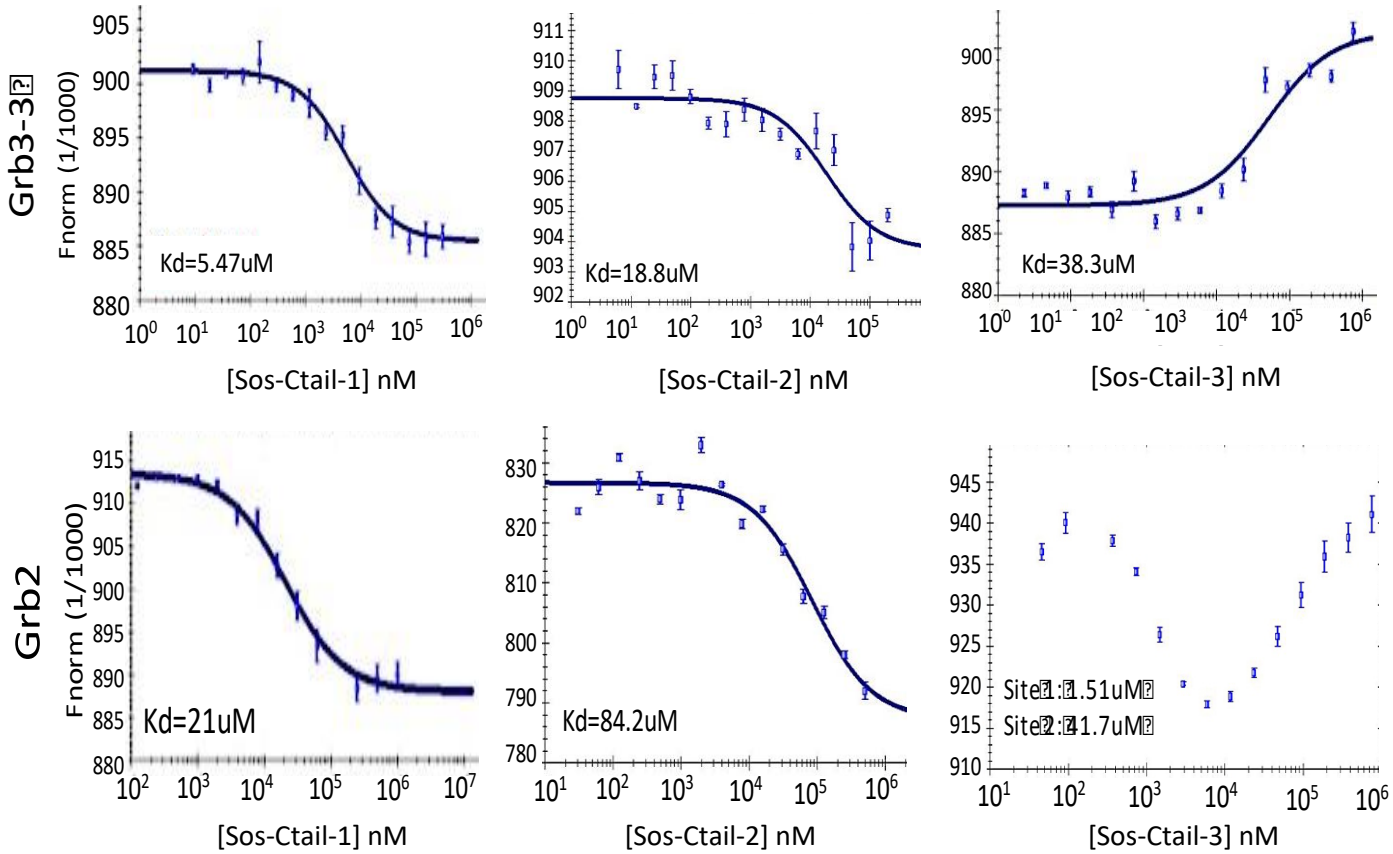


Figure 4.10. Biophysical characterisation of the interaction between Grb2 and Grb3-3 with proline-rich peptides derived from c-terminus of Sos C-terminus. MST binding graph for the interactions between Grb3-3 (row 1) and Grb2 (row 2) with three proline-rich peptides derived from the C-terminus of Sos, binding affinities for the interactions are shown on each plot. All MST data plotted represents the mean plus standard deviation for 3 technical repeats.

Binding of Grb2 to the proline-rich C-terminus of Sos has been previously reported and is known to be mediated by the N-terminal SH3 domain [101]. To identify the domain responsible for Grb3-3 binding MST experiments were performed using the truncated protein binding domains (NSH3-SH2Δ_{40} and SH2Δ_{40}-CSH3). Both the N and C-terminal SH3 domains were able to bind to the C-terminal proline-rich peptides however, the affinity of the NSH3-mediated interaction was up to 2 orders of magnitude higher (Figure 4.11). Consistent with the full-length protein, the highest affinity interaction occurred between the NSH3-SH2Δ_{40} domain and the canonical PVPP-containing Sos-Ctail1 peptide. The interaction between the CSH3 domain and Sos-Ctail peptides 2 and 3 was extremely weak with affinities of 145 μM and 297 μM respectively (Figure 4.11) these interactions are unlikely to occur under physiological conditions since the concentration of Grb2 in the cell has been reported to be up to 10μM and Grb3-3 levels are reported to be 10-50x lower than Grb2 [89]. Therefore, binding of Grb3-3 to the C-terminus of Sos is likely to occur primarily through the NSH3 domain reminiscent of Grb2 however, under conditions where the local concentration of Grb3-3 is high, binding may also occur via the CSH3 domain.

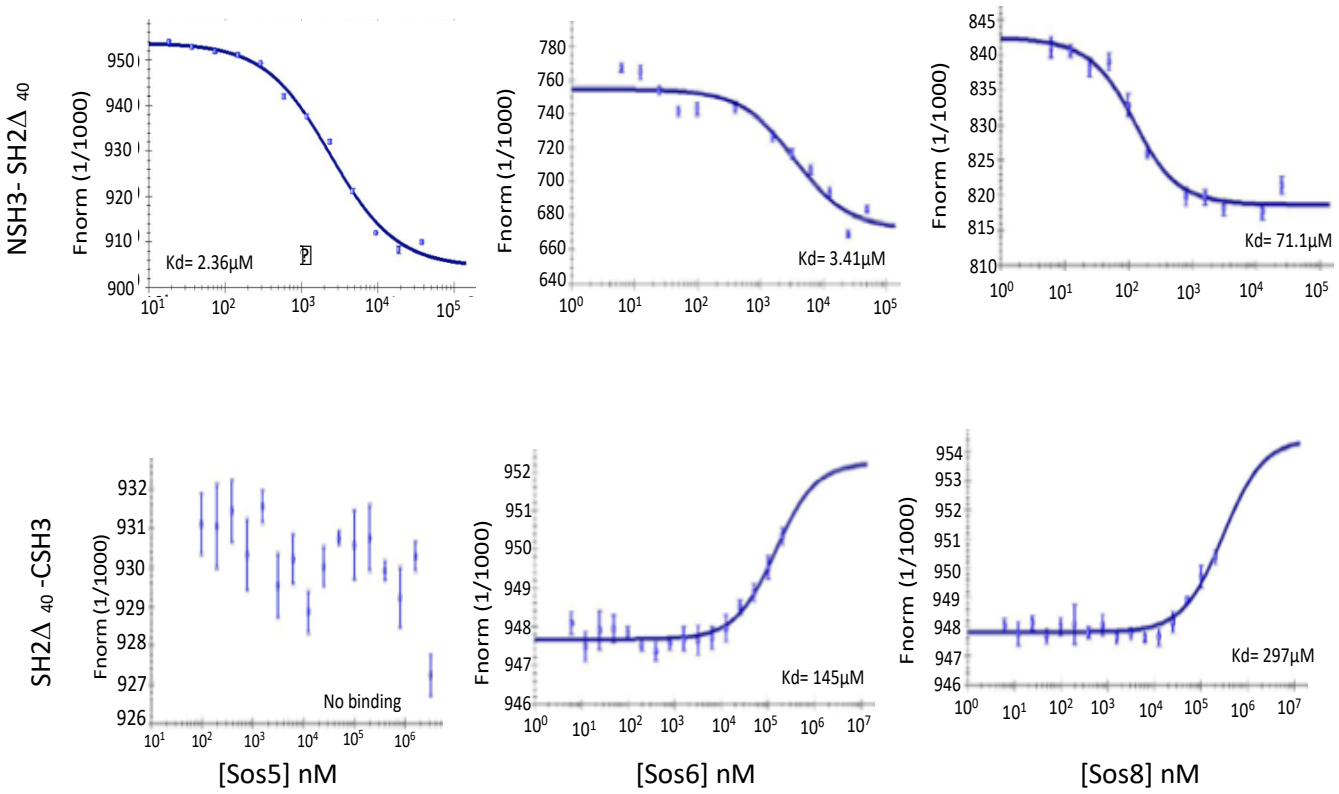


Figure 4.11 Biophysical characterisation of the interaction between Grb3-3 domain and proline-rich peptides derived from Sos C-terminus. MST binding trace for the interactions between Grb3-3 NSH3SH2Δ_{40} domain (row 1) and SH2Δ_{40}CSH3 domain (row 2) with three proline-rich peptides derived from the C-terminus of Sos, binding affinities for the interactions are shown on each plot. All MST data plotted represents the mean plus standard deviation for 3 technical repeats.

4.4. Competition between Grb3-3 and Grb2 for binding to Sos c-tail

Having established that both Grb3-3 and Grb2 bind with similar affinities to the C-terminal proline-rich tail of Sos *in vitro*, the physiological relevance of this interaction in cells was investigated. In a cellular context where both isoforms are present competition for binding to this site on Sos may occur. To investigate this the effect of Grb3-3 on the ability of Grb2 to immunoprecipitate endogenous Sos was assessed in Hek293t cells under both serum-starved and FGF9-stimulated conditions (Figure 4.12a). Under non-stimulated conditions in the absence of Grb3-3, low levels of the Grb2-Sos complex were present (lane 2), consistent with previous reports that in the absence of stimulation Grb2 dimerisation reduces its ability to bind to Sos [103]. Increased Grb2-Sos complex formation was observed under stimulated conditions (lane 5). Overexpression of Strep tagged Grb3-3 reduced Grb2-Sos complex formation under both serum-starved (lane 1) and stimulated conditions (lane 4), suggesting that the physiological function of the interaction between Grb3-3 and the C-terminus of Sos is to inhibit Grb2 binding. The relative amounts of the Grb2-Sos complex formed under different conditions is quantified in Figure 4.12b allowing the level of inhibition of complex formation by Grb3-3 to be determined. Exogenous expression of Grb3-3 inhibited complex formation by roughly 65% under non-stimulated conditions and 80% under stimulated conditions. Direct competition of the two isoforms for binding to the C-terminus of Sos was confirmed using MST. Grb2 bound to the Sos-Ctail 1 peptide with an affinity of 21 μM in the absence of Grb3-3, but pre-saturation of the peptide, by incubation with an excess of Grb3-3 completely abrogated Grb2 binding (Figure 4.12c and d).

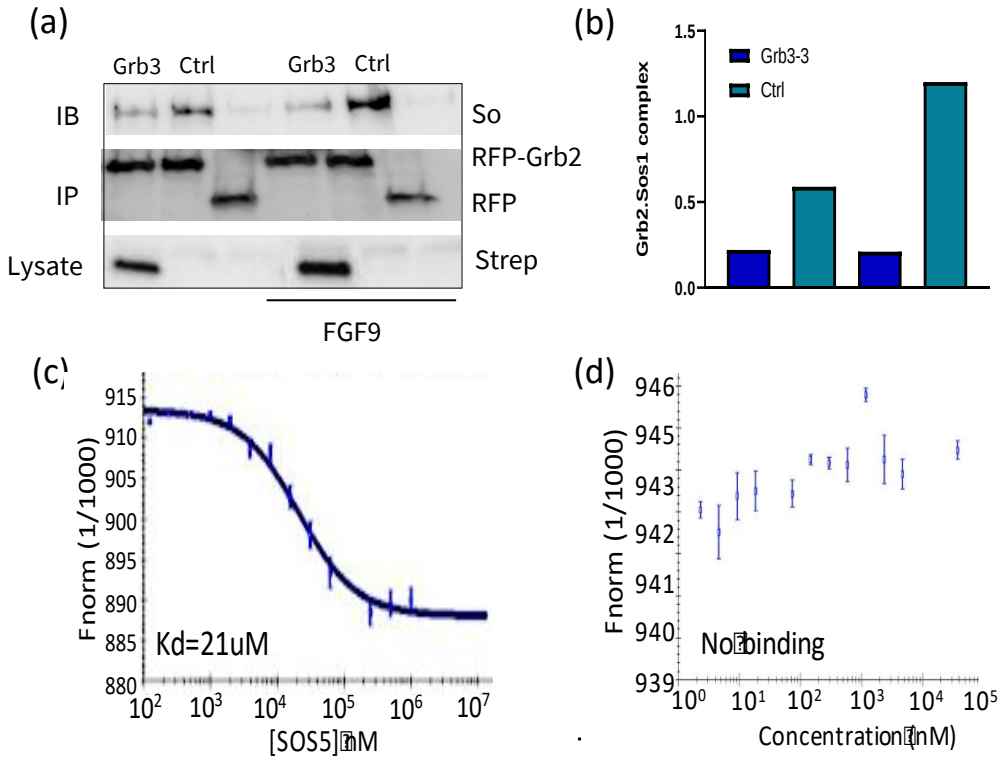


Figure 4.12 Competition between Grb2 and Grb3-3 for binding to Sos. (a) affinity purification of RFP tagged proteins and immunoblot for co-purified Sos shows the presence of Grb3-3 reduces the Grb2 Sos interaction. This is quantified in b) the levels of Sos are normalised to the amount of RFP protein present. c) MST graph for Grb2 binding to Sos peptide 5 from the C-terminal tail. d) Pre-saturation of the peptide with Grb3-3 prevents Grb2 binding

4.5 Inhibition of Grb2 mediated Sos membrane localisation

The interaction between Grb2 and Sos is responsible for the recruitment of Sos to the plasma membrane through the SH2-mediated interaction with phosphotyrosine residues on activated RTKs as well as the NSH3 mediated interaction with Sos. This interaction brings Sos into close proximity with its membrane-localised substrate Ras, where it catalyses the release of GDP enabling GTP binding and Ras activation [90, 95-98]. Figure 4.12 shows that Grb3-3 inhibits the Grb2-Sos interaction, and therefore may function to inhibit Grb2 mediated membrane localisation of Sos. Furthermore, since Grb3-3 does not bind phosphotyrosine residues due to the truncation of the SH2 domain [89], it is unable to be recruited to RTKs hence the Grb3-3-Sos complex will inhibit recruitment of Sos to the plasma membrane. The effect of Grb3-3 on Sos localisation was investigated by cellular fractionation. Hek293t cells, where the endogenous levels of Grb2 were knocked down to approximately 50% using stable expression of Grb2 shRNA, were used in the cell fractionation experiments enabling either Grb2 or the truncated isoform Grb3-3 to be reconstituted in these cells. Following transfection the cells were serum starved prior to stimulation with FGF9. The membrane and cytoplasmic fractions were separated via ultracentrifugation and immunoblotting was used to assess the levels of endogenous Sos in the two cellular compartments (Figure 4.13). Probing for the plasma membrane marker E-cadherin and the cytoplasmic marker beta-actin confirmed successful cellular fractionation. The levels of these proteins were also used as a control to normalise the amount of total protein in each cellular fraction. Overexpression of Grb2 in these cells only resulted in a small increase in the levels of Sos at the membrane relative to the RFP control. This may be because Grb2 is present in the cells at a higher concentration than Sos (despite

the expression being knocked down) and therefore Grb2-Sos complex formation will already be saturated. It may also be that the RTK population is saturated with Grb2-Sos in WT cells. In contrast overexpression of Grb3-3 caused a clear increase in the levels of Sos in the cytoplasm and a decrease in membrane localised Sos, providing evidence that Grb3-3 prevents Grb2-mediated Sos membrane localisation through competitive inhibition of Grb2 binding.

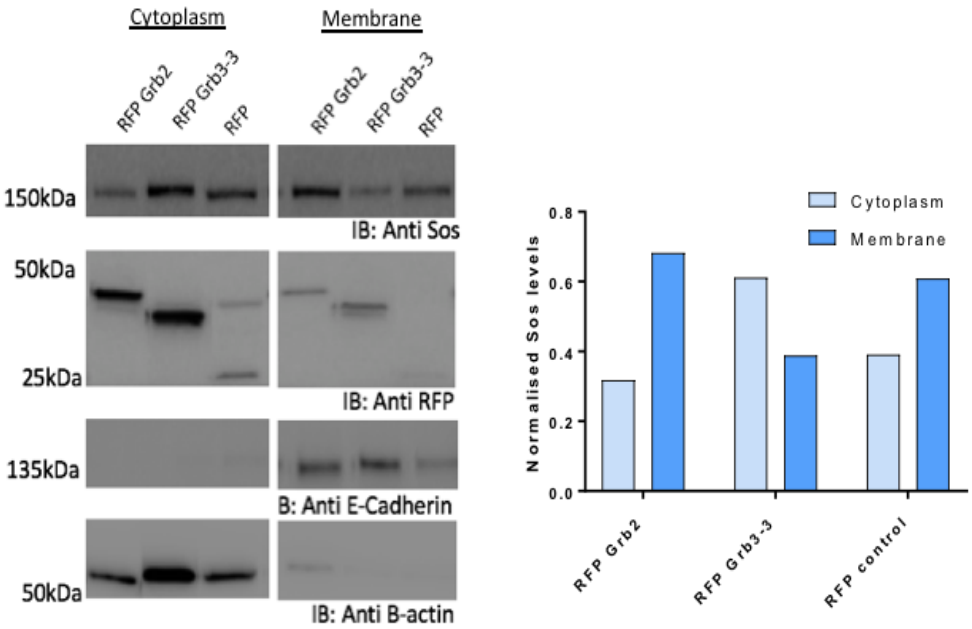


Figure 4.13. Effect of Grb2 and Grb3-3 on Sos cellular localisation. (a) Immunoblot for Sos in the membrane and cytoplasmic fractions of Hek293t FGFR2 GFP G2i cells transfected with RFP Grb2, RFP Grb3-3 or RFP under serum-starved or stimulated conditions. E-Cadherin is used as a membrane marker and beta actin is used as a cytoplasmic marker. (b) Quantification of Sos protein levels normalised to the organelle markers. Data represents a single experiment.

4.6 Inhibition of cellular proliferation

The Ras-Erk signalling cascade regulates cell proliferation by inducing the expression of genes involved in cell cycle progression. Therefore Grb3-3-induced inhibition of Ras activation, through cytoplasmic sequestration of Sos, should prevent cell proliferation. The effect of Grb3-3 on cell proliferation was investigated in human Colo320 colorectal adenocarcinoma cells (Dukes type C, isolated from a 55 year old female). These cells were used as they have a faster doubling time than the Caco2 cells thereby allowing a change in proliferation to be observed within the time scale of the assay. Colo320 cells were transfected with RFP-Grb3-3 or RFP alone and cell proliferation was measured 24 hours following transfection using the brdU cell proliferation assay. This assay detects the incorporation of 5-bromo-2-deoxyuridine (brdU) into newly synthesized DNA during cell proliferation. BrdU incorporation is then detected using an antibody-based method resulting in the development of a yellow colour that can be quantified using the absorbance at 450 nm; higher absorbance readings indicate more BrdU incorporation thus greater cell proliferation. Grb3-3 significantly inhibited Colo320 cell proliferation by 49% (Figure 4.14a). The effect of Grb2 on Colo320 cell proliferation was also investigated; however overexpression of Grb2 had no significant effect on the rate of cell proliferation (p value 0.26, data not shown), likely to be due to already high expression of the adaptor protein in the Colo320 cells. To determine if the inhibition of cell proliferation by Grb3-3 is due to the non-functional SH2 domain a Grb2 R82L SH2 mutant was produced. This mutation inactivates the SH2 domain. Expression of mutant Grb2 also significantly inhibited Colo320 cell proliferation p value 0.004, although the level of inhibition was lower (Figure 4.14b), confirming that the inhibition of cell

proliferation caused by Grb3-3 is, at least in part, due to the inability of the SH2 domain to bind phosphotyrosine residues.

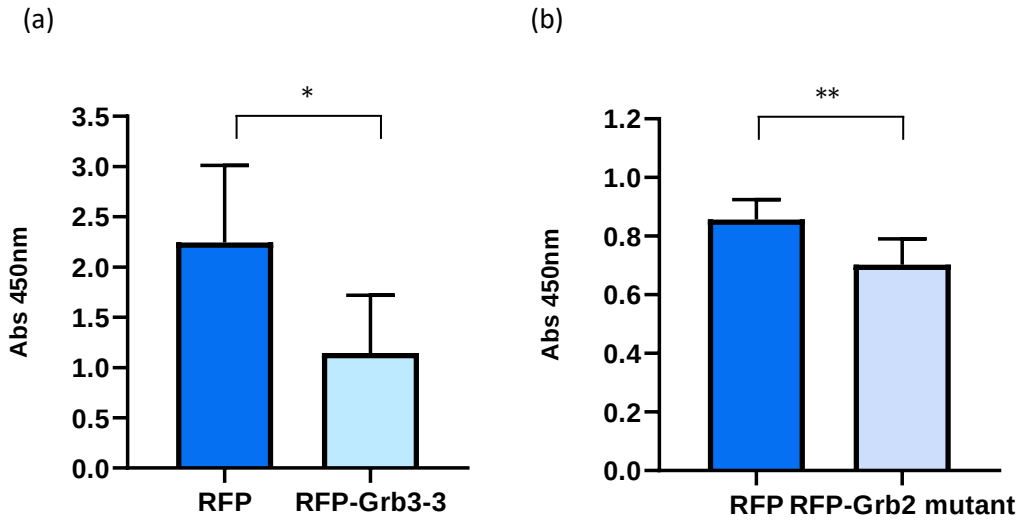


Figure 4.14. Effect of Grb3-3 on colo320 cell proliferation. Bar chart showing average absorbance at 450 nm of Colo320 cells cultured in Brdu for 24 hours post transfection of (a) RFP or RFP-Grb3-3 plasmids, and (b) RFP or RFP Grb2 R82L mutant. Data shown as mean plus standard deviation, n=6, * represents p value <0.05 and ** p value <0.005.

4.7 Summary

This Chapter reports on a novel molecular mechanism for inhibition of growth factor-induced Ras activation by the Grb2 splice variant Grb3-3. Grb3-3 binds via its N-terminal SH3 domain to the proline-rich C-terminus of the Ras guanine nucleotide exchange factor Sos. Binding of Grb3-3 to this site inhibits Grb2 binding and Grb2 mediated recruitment of Sos to the plasma membrane, thus sequestering Sos in the cytoplasm away from its substrate Ras. Cytoplasmic sequestration of Sos prevents GTP loading of Ras, resulting in inhibition of downstream Erk signalling and cellular proliferation. Grb3-3 is also able to bind to the catalytic CDC25 domain of Sos, however the molecular consequence of this interaction is unknown.

The opposing function of the two splice variants in the Ras-Erk signalling cascade therefore, highlights a novel mechanism for tight control of Ras activation which is dictated by *GRB2* splicing. In cells with high levels of Grb2 the Grb2-Sos complex will prevail facilitating Sos membrane recruitment and Ras activation. In contrast, in cells with high levels of Grb3-3, the Grb3-3-Sos complex will prevail resulting in cytoplasmic sequestration of Sos and inhibition of Ras signalling. Therefore, in the absence of Ras mutations, loss of Grb3-3 regulation of Ras activation represents a novel route for oncogenic Ras signalling, shown in Chapter 3 to occur in a subset of colorectal cancers. Further work is required to understand the regulation of *GRB2* splicing, which may reveal novel strategies to therapeutically inhibit oncogenic Ras signalling by manipulating the levels of the two splice variants Grb2 and Grb3-3.

Chapter 5: Regulation of *GRB2* splicing

5.1 Introduction

Alternate splicing is regulated by trans-acting splicing factors that either promote (enhancers) or repress (silencers) splicing. The splicing factors recognise short sequences found in the introns and exons of pre-mRNA, typically in close proximity of exon intron boundaries. Two major families of splicing factors are the serine/arginine rich proteins (SR proteins), and the heterogenous nuclear ribonucleoproteins (hnRNPs). The SR proteins typically function as splicing enhancers, while the hnRNPs function as splicing silencers, although their regulatory consequences are context specific and can vary dependent on where they bind within the pre-mRNA. Altered expression or mutation of splicing factors is frequently observed in cancer and can contribute to aberrant splicing.

GRB2 splicing is altered in colorectal cancer, which appears to contribute to aberrant activation of Ras signalling in the absence of mutation through the loss of Grb3-3 regulation of Sos membrane recruitment. In this Chapter the regulation of *GRB2* splicing was investigated. *In silico* analysis and RNA:protein pull-downs were used to identify the splicing factors that bind to the alternate exon 4. Subsequently knockdown experiments and transient overexpression of the splicing factors were used to study their functional role in the regulation of exon 4 splicing. Understanding the regulation of cancer-associated splicing events is critical and may pave the way to novel and innovate RNA-based therapies to correct aberrant splicing.

5.2 In silico analysis of splicing factor binding to *GRB2* exon 4

Online databases exist that contain information on putative RNA binding proteins and their preferential binding sequence based on published experimental data. SpliceAid2 is an example of an online database that can be used to predict which RNA binding proteins bind to particular RNA sequences of interest [23]. SpliceAid2 was used to identify known splicing factor binding sites within the *GRB2* alternate exon 4. The 123 nucleotide mRNA sequence corresponding to exon 4 was used in the analysis.

GRB2 exon 4 sequence:

GUGGUUUUUUGGCAAAAUCCCCUGUGCCUUGGCAGAAGAAAUGCUUAGCAAACUGC GGCUCGAUGGGGCCUUUCUUAUCCGAGAGAGUGAGAGCGCUCCUGGGGACUUCUCC CUCUCUGUCAA

SpliceAid2 identified 20 splicing factor binding sites within exon 4, shown in Table 5.1 along with their position and sequence. Almost half (9/20) of the splicing factor proteins identified belonged to the hnRNP family. Four SR proteins were predicted to bind in this exon; SRSF9, 2 10 and 5. Several other splicing factors were identified which do not belong to either family of splicing regulatory proteins, including two K homology domain containing, RNA binding signal transduction-associated proteins (KHDRBS); KHDRBS1 and 3. The K homology domain is an RNA binding domain that was first identified in hnRNPK. KHDRBS1 and 3 have been reported to play a role in signal transduction as well as regulating splice site selection and exon inclusion during alternate splicing. They are recruited and tyrosine phosphorylated by receptor proteins and can bind SH2- and SH3-containing proteins [176, 177].

Splicing Factor	Recognised Sequence	Sequence Position
ELAVL1	UUUUU	5-9 or 6-10
HNRNP C	UUUUU(U) UUUUUG	5- 9 or 10 6-10
KHDRBS1	UUUUUU (CA)AAAU	5-10 13-18 or 15-18
HNRNP G	AAAAU AUCCCC	14-18 17-22
KHDRBS3	AAAU	15-18 or 39-42
HNRNP E2	UCCCCA	18-23
SRSF9	AGGCA	30-34
SRSF2	AGAAG	34-38
SRSF10	(G)AAGAA	35-40 or 36-40
Tra2alpha	AAGAA	36-40
HNRNP A1	CUUAG	44-48
ELAVL2	CUUUC	70-74
TIA-1	CUUUC	70-74
MBNL1	AUGCUU GCGCUC	41-46 93-98
HNRNP F	UGGGG GGGGC UGGGG GGGGA	64-68 65-69 100-104 101-105
HNRNP H1, H2, H3	AAGAA UGGGG GGGGC AGAGA UGGGG GGGGA	36-40 64-68 65-69 82-86 100-104 101-105
HNRNP I	CUUUCUU CUCUCU UCUCU	70-76 113-118 114-118
SRSF5	ACAGC	52-56

Table 5.1 Splicing factors predicted to bind to sequences located within *GRB2* exon 4 using SpliceAid2

Comparison of the sequences of the splicing factor binding sites suggests that they predominantly bind to polynucleotide regions. Figure 5.3 shows schematically where, within the exon, the binding sites are positioned. There is significant overlap between different splicing factor binding sites suggesting there may be some redundancy amongst them. Furthermore the majority (13/20) of splicing factor binding sites were identified within 50 nucleotides of the intron-exon boundary.

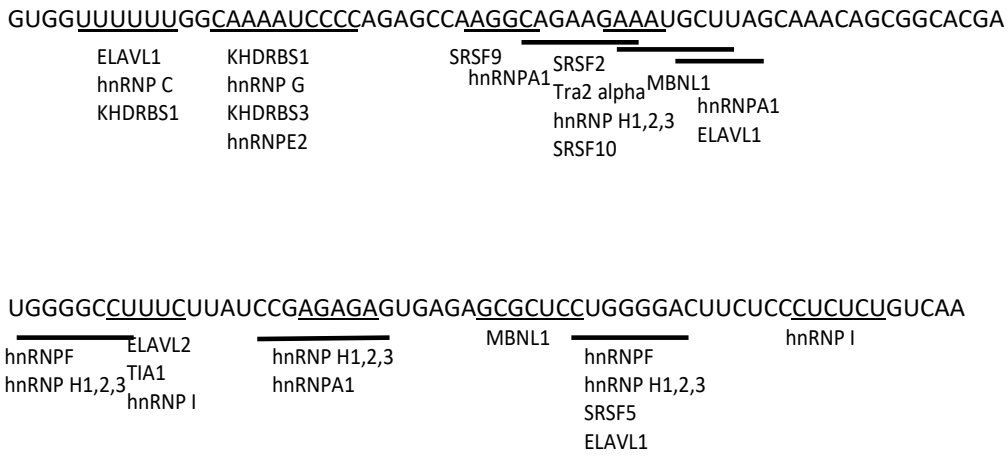
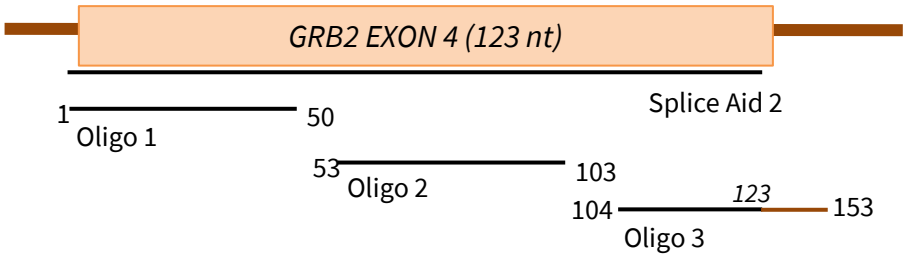


Figure 5.1 Location of splicing factor binding sites in *GRB2* exon 4 identified using SpliceAid2

5.3 Identification of splicing factors that bind *GRB2* exon 4

Whilst bioinformatic tools are useful to predict which splicing factors bind to a sequence of interest they need to be validated experimentally. There are several techniques that can be used to study interactions between RNA and proteins [178-180]. Electrophoretic mobility shift assays (EMSA) can be performed to study interactions between purified proteins and specific RNA molecules. EMSA exploits the fact that RNA:protein complexes migrate more slowly in native agarose gels than the free RNA, so a shift in mobility is indicative of binding to a recombinant protein. RNA immunoprecipitation can be performed to identify RNA molecules that bind to a specific protein of interest. Antibodies are used to immunoprecipitate the protein of interest along with any bound RNAs which can then be eluted and sequenced. The technique used in this study was an RNA pull down. This technique uses synthetic RNA oligonucleotides immobilised onto magnetic beads to pull out binding partners from cell lysates. This technique was chosen because, coupled with mass spectrometry, it allows the identification of unknown proteins that bind to a specific RNA sequence. Three synthetic RNA oligonucleotides that spanned *GRB2* alternate exon 4 and part of the 3' intron each of 50 nucleotides in length were designed, as shown schematically in Figure 5.2.



RNA_Oligo 1: GUGGUUUUUUGGCAAAAUCCCCAGAGCCAAGGCAGAAGAAAUGCUUAGCA

RNA_Oligo 2: AGCGGCACGAUGGGGCCUUUCUUAUCCGAGAGAGUGAGAGCGCUCCUGGG

RNA_Oligo 3: GACUUCUCCCUCUCUGUCAAGUAAGUAUUUCCUGCUGCAGUUGCCUGGAA

Figure 5.2 Schematic representation of RNA:protein pull down experiment. The Figure shows the alternate exon 4 of *GRB2* along with the sequence of three 50 base oligonucleotides used to pull down RNA binding proteins

The RNA oligos were labelled at the 3’ end with a single biotinylated cytidine (bis)phosphate and then immobilised onto streptavidin beads and used to pull out bound proteins which were identified using mass spectrometry. Successful labelling of the RNA oligos was tested using a dot blot with streptavidin-coupled horse radish peroxidase substrate (Figure 5.1). As a negative control the labelling reaction was set up without RNA.

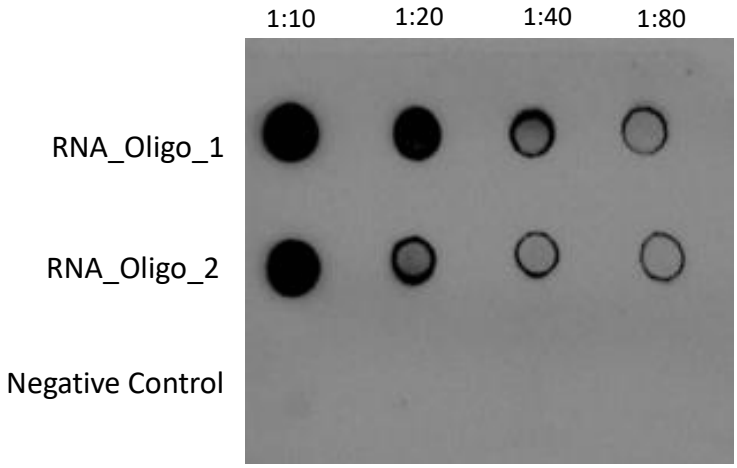


Figure 5.3. Dot blot showing labelled synthetic RNA oligonucleotides. Serial dilution of desthiobiotin labelled RNA oligos 1 and 2, and the negative control sample (set up in the absence of RNA) on a nitrocellulose membrane. The samples were probed with streptavidin horse-radish peroxidase conjugate

Mass spectrometry identified a number of RNA binding proteins that bound to *GRB2* exon 4. Gene ontology analysis was used to sort them based on the biological processes they are involved in. RNA binding proteins involved in transcription, RNA stabilisation, 3' polyadenylation as well as RNA splicing were identified. The RNA binding proteins with an established in RNA splicing that bound to each of the RNA oligos are shown in Figures 5.4-5.6 and are sorted into splicing factor families. Only proteins that were identified with more than 2 unique peptides, and that did not bind to the negative control (magnetic beads only) were deemed as true binding events. The mass spectrometry analysis identified a number of splicing factors that were not predicted to bind using the *in silico* analysis. These splicing factors may be binding indirectly to the RNA through the formation of protein-protein complexes, and therefore will not be identified using the *in silico* analysis, which only identifies direct binding partners. Both the hnRNP and SR protein families contain protein-binding domains and have been reported to form protein complexes [25, 30, 181]. 48 splicing factors bound to RNA oligo 1, 26 to RNA oligo 2 and 52 to RNA oligo 3. Therefore, approximately twice as many splicing factors were identified binding within 50 nucleotides of the intron-exon junctions. The majority of splicing factors that bound to the RNA oligos belonged to the hnRNP family.

Comparison of the mass spectrometry data with the *in silico* analysis showed 11 out of the 20 splicing factors predicted to bind with the *in silico* analysis were experimentally validated as binding to the RNA oligo containing the splicing factor binding site. These splicing factors are highlighted in red in Figures 5.4-5.6. Binding sites for some of the splicing factors identified using the *in silico* analysis, that were not shown to bind experimentally, overlapped with the binding sites for splicing factors identified experimentally as binding. This suggests that there may

be competition for binding between splicing factors. For example KHDRBS3 was predicted to bind AAAU at position 14-18 of oligo 1, this sequence was also predicted to be the binding site for KHDRBS1 and hnRNPG. Both of these proteins were shown to bind experimentally, suggesting that they may have blocked KHDRBS3 binding. The same was true for ELAVL2. This protein was predicted to bind CUUUC within oligo 2, which was also the binding site for hnRNPI that was shown to bind experimentally.

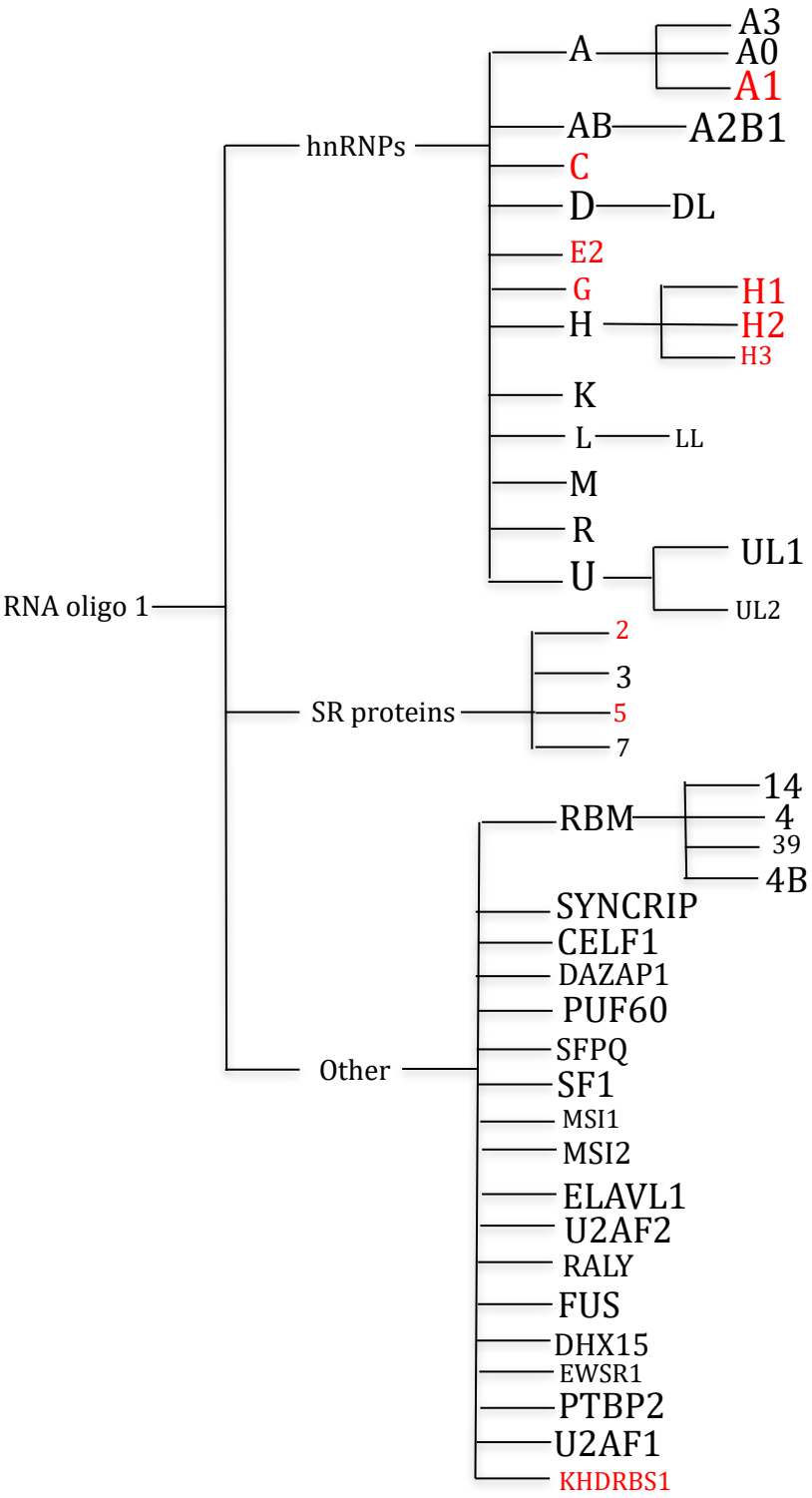


Figure 5.4 Splicing factors that bind to RNA oligo 1. The size of the font corresponds to the number of peptides identified for that protein, (larger font greater number of peptides). Splicing factors in red were predicted to bind using the *in silico* analysis

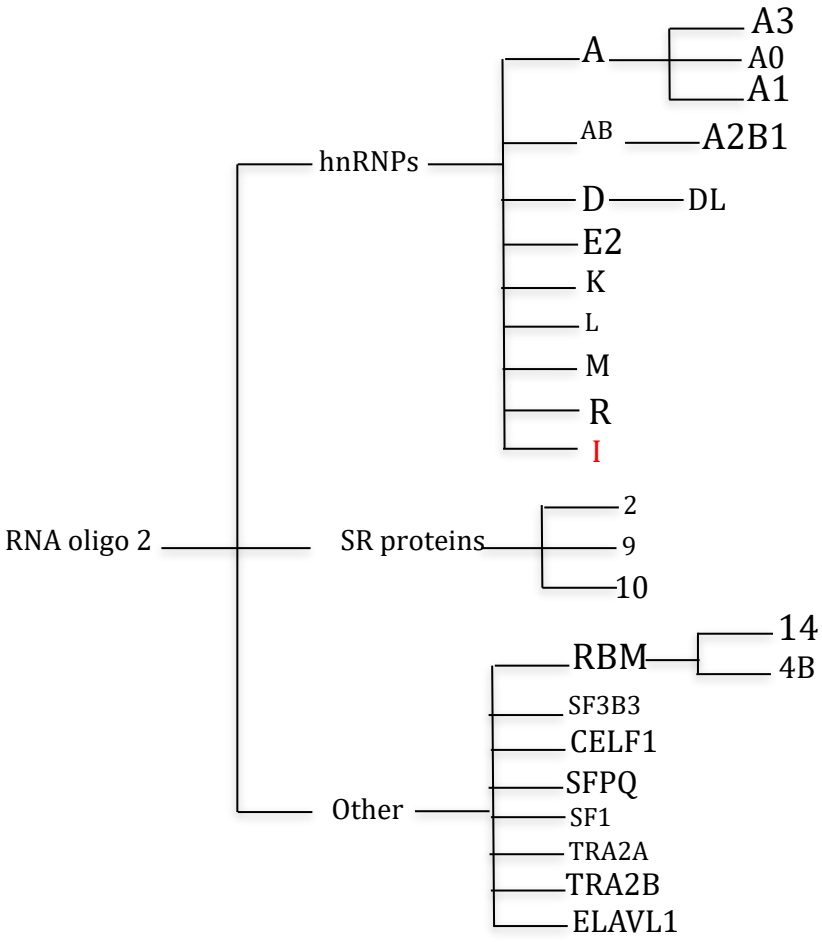


Figure 5.5 Splicing factors that bind to RNA oligo 2. The size of the font represents the number of peptides identified for that protein (larger font greater number of peptides). Splicing factors in red were predicted to bind using the *in silico* analysis.

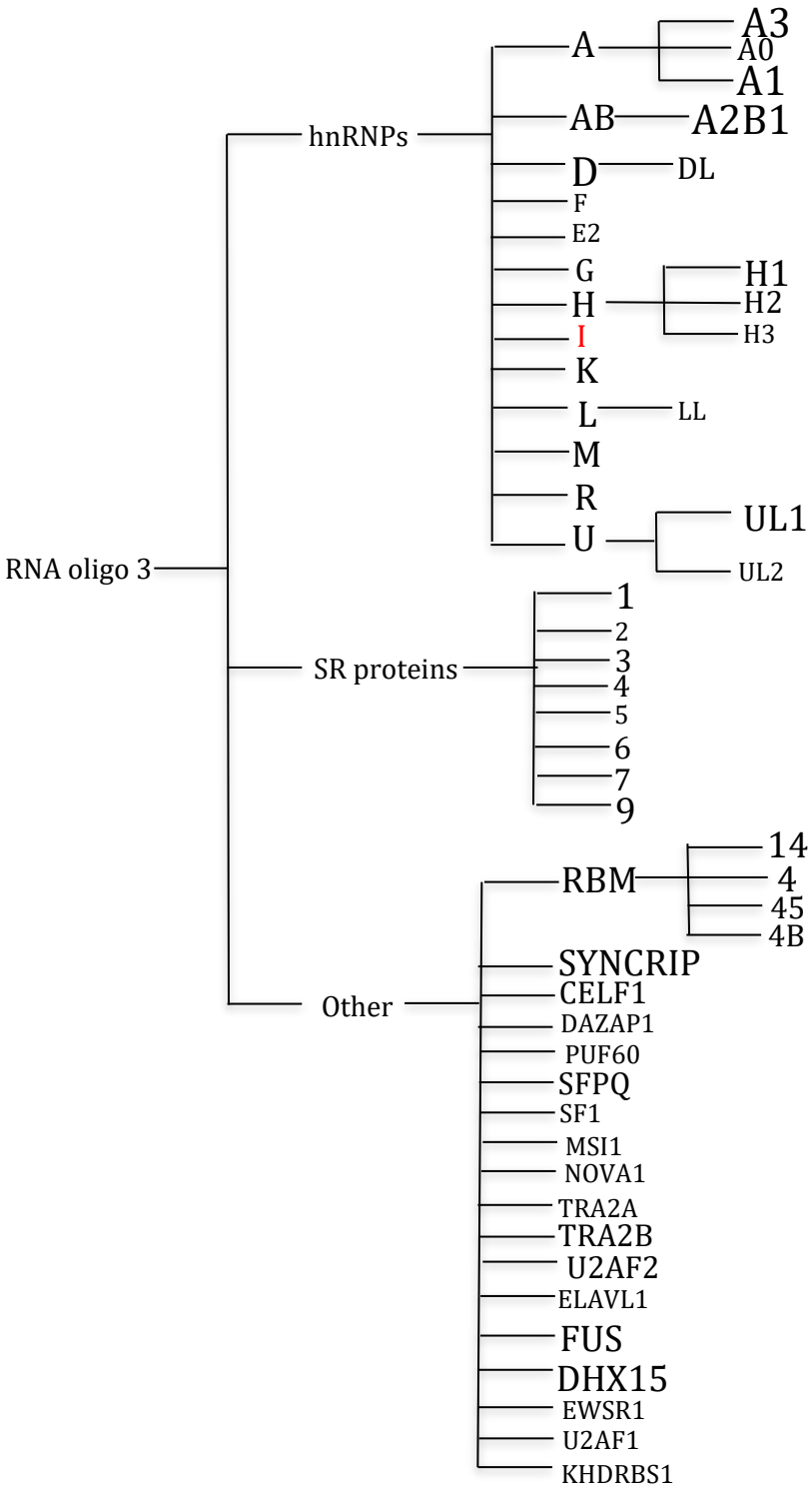


Figure 5.6 Splicing factors that bind to RNA oligo 3. The size of the font represents the number of peptides identified for that protein (larger font greater number of peptides). Splicing factors in red were predicted to bind using the *in silico* analysis

5.4 Functional characterisation of splicing factors that bind *GRB2* exon 4

The *in silico* analysis and experimental method identified 11 splicing regulatory proteins that bound to *GRB2* exon 4. All but one of the splicing factors bound to RNA oligo 1 encompassing the first 50 nucleotides of exon 4. hnRNPI bound to both RNA oligo 2 and 3. In order to investigate the functional role of these splicing factors in the regulation of *GRB2* exon 4 splicing, publically available RNA sequencing data, where each of the splicing factors were individually knocked down using shRNA or CRISPR/Cas9 gene editing prior to sequencing the entire transcriptome, was analyzed from K562 and HepG2 cells. This work was performed by Nostrand et al. and the raw RNA sequencing data was made publically available in the ENCODE database [182, 183]. Unfortunately, there was no data for knockdown of SRSF2 or hnRNPH splicing factors, and therefore these were excluded from our analysis. Further work would be required to determine if these splicing factors play a role in regulating splicing of *GRB2* exon 4. A statistical analysis, called replicate multivariant analysis of transcript splicing (rMATS), was applied to the RNA sequencing data available for the other splicing factors. rMATs is a robust statistical model which is able to assign and identify significant differences in splicing events between the two RNA sequencing datasets obtained from the knockdown cells and control cells treated with a scrambled shRNA [184]. rMATs can identify differences in five splicing events; exon skipping (SE) alternative 5’ splice site (A5SS), alternative 3’ splice site (A3SS), mutually exclusive exons (MXE) and retained introns (RI). The rMATs analysis revealed that knockdown of hnRNPC significantly altered skipping of *GRB2* exon 4, with a p value of 0.008984. The results of the rMATs analysis in the hnRNPC knockdown cells are shown in Table 5.2. Knockdown of the other RNA binding proteins did

not significantly alter *GRB2* exon 4 splicing, suggesting that hnRNPC is the major RNA binding protein responsible for controlling *GRB2* splicing.

FileID	HNRNPC-BGKLV21-K562__ENCFF213JWB/MATS_output/SE.MATS.JunctionCountOnly.txt
ID	27123
GeneID	ENSG00000177885.9
geneSymbol	GRB2
chr	chr17
strand	-
exonStart_0base	73321978
exonEnd	73322101
upstreamES	73317875
upstreamEE	73317908
downstreamES	73328780
downstreamEE	73328878
ID	27123
IJC_hnRNPC shRNA	198,190
SJC_hnRNPC shRNA	1,5
IJC_control	291,364
SJC_control	0,0
IncFormLen	197
SkipFormLen	99
PValue	0.008984227
FDR	0.108942659
IncLevel1	0.99,0.95
IncLevel2	1.0,1.0
IncLevelDifference	-0.03

Table 5.2. rMATs analysis of significantly different splicing events for hnRNPC knockdown and control experiments (analysis performed by Dr Elton Rosas de Vasconcelos).

5.5 Investigating the regulation of *GRB2* exon 4 splicing by hnRNPC

To confirm the role of hnRNP C in regulating *GRB2* exon 4 splicing Myc-tagged hnRNPC was overexpressed in Hek293t cells. RNA was isolated from the cells, 24 48 and 72 hours post transfection and used to synthesise cDNA. The two splice variants were amplified using the isoform-selective primers described and validated in Chapter 3. GAPDH was also amplified as a control for variants in the amount of cDNA used in each PCR assay. The PCR products were separated via agarose gel electrophoresis, as shown in Figure 5.7a. The normalised levels of the 2 splice variants is quantified in Figure 5.7b. Overexpression of hnRNPC resulted in a 2.2-fold increase in the levels of the shorter transcript Grb3-3 relative to the myc-tag-expressing control cells 24 hours post-transfection and 3.2-fold increase 48 hours post-transfection. No further increase in Grb3-3 levels were seen 72 hours post-transfection, and the levels of Grb2 remained constant at each time point. In the control cells Grb3-3 levels were 5-fold lower than Grb2. However, 48 hours following overexpression of hnRNPC, Grb3-3 was expressed at almost equal levels to Grb2. This data infers that hnRNPC suppresses inclusion of exon 4 resulting in increased levels of Grb3-3; minigene analysis could be used to confirm the effect of hnRNPC on *GRB2* exon 4 skipping.

Mass spectrometry analysis identified that hnRNPC bound to RNA oligo 1 derived from the first 50 nucleotides of exon 4. hnRNPC was predicted to bind to a pentameric polyuracil sequence at position 5-10 of this oligonucleotide. To confirm that hnRNPC does recognise the polyU sequence in exon 4 an RNA:protein pull down was performed using an RNA oligo that flanked the predicted binding site and a mutant oligonucleotide where the uracil bases at position 5-10 were mutated to cytosine. Binding was assessed via western blotting using a specific antibody to detect hnRNPC (Figure 5.8). Immunoblotting

confirmed that hnRNPC does bind to the polyuracil sequence at position 5-10, as reduced binding was observed in the U-Cmutant RNA oligo. However the U-C mutation did not completely abolish binding, suggesting that hnRNPC is also able to bind to polyC sequences but with reduced affinity, or that there is an additional hnRNPC binding site within the 20 nucleotide oligo1.

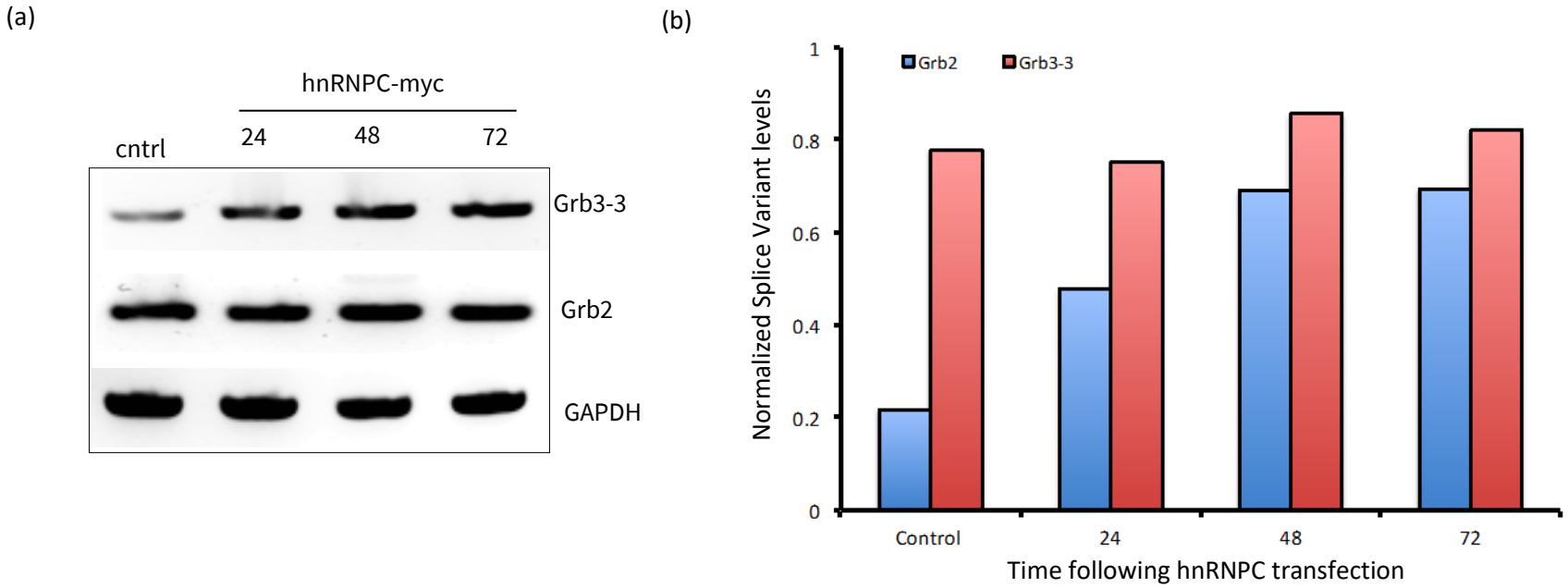


Figure 5.7. Effect of hnRNPC on *GRB2* splicing in Hek293t cells (a) Agarose gel image of Grb3-3, Grb2 and GAPDH PCR products produced from RNA isolated from Myc (cntrl) or hnRNPC-Myc transfected Hek293t cells. (b) Quantification of the mRNA levels of the two splice variants normalised to GAPDH expression. Data represents a single experiment.

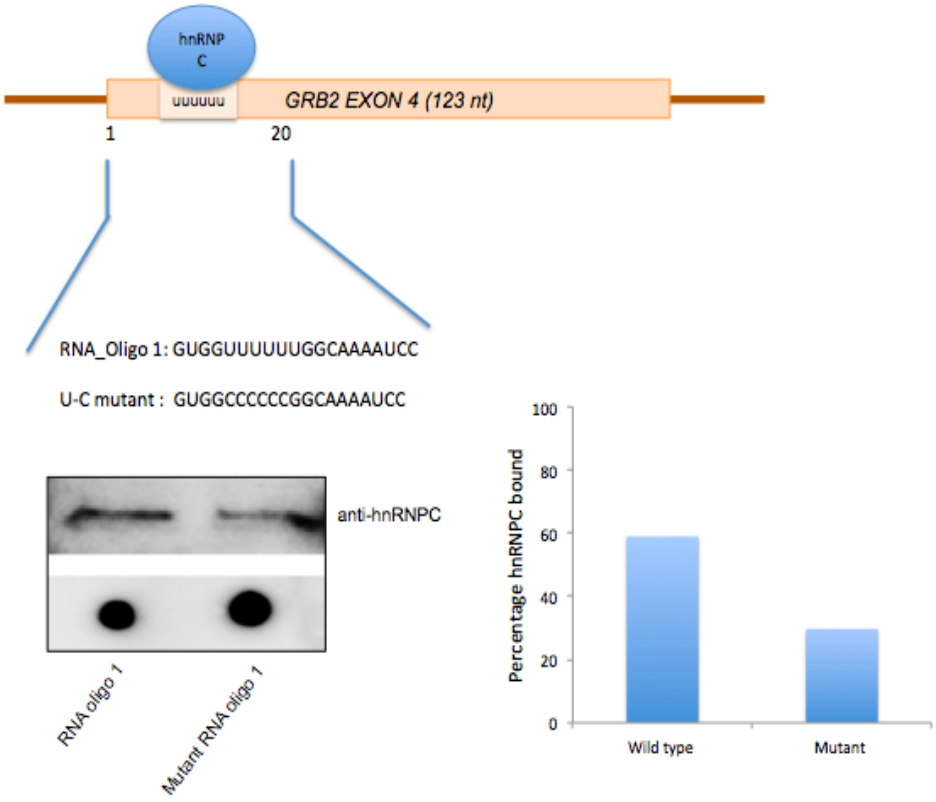


Figure 5.8 Analysis of hnRNPC binding to *GRB2* oligo 1, or C-U mutant oligo 1 by western blot. (a) Schematic representation showing oligonucleotides used in the RNA:protein pull-down spanning the predicted hnRNPC binding site in *GRB2*. (b) Western blot of pull-down samples: dot blot shows the labelled RNA oligos used in the pull down. Top bands correspond to bound hnRNPC. (c) Quantification of the amount of hnRNPC bound. This was calculated using the intensity of the band measured in Image J, normalised against the amount of labelled RNA. Date represents a single experiment.

5.6 Analysis of hnRNPC expression in colorectal cancer

Data presented in this Chapter shows that hnRNPC binds and suppresses the inclusion of *GRB2* exon 4 resulting in increased Grb3-3 levels in cells overexpressing hnRNPC. Grb3-3 levels were shown in Chapter 3 to be reduced in colorectal tumour tissue relative to the surrounding normal colonic tissue. Differential expression of genes encoding splicing factors is frequently observed in cancer and can contribute to changes in alternate splicing [35]. For this reason RT-qPCR was used to analyse hnRNPC expression in the matched colorectal cancer and surrounding normal tissue samples. Figure 5.9 shows the relative expression of hnRNPC in each patient. hnRNPC expression was significantly down-regulated in the tumour tissue relative to the surrounding normal colonic tissue. Reduced expression of hnRNPC in the tumour tissue may contribute to the reduced Grb3-3 expression observed in these samples.

To determine if altered hnRNPC expression influenced patient survival, data was extracted from the human protein atlas (Figure 5.10). hnRNPC expression was measured by RNA sequencing in 597 colorectal cancer patients and correlated to the survival time following diagnosis. Patients with low expression of hnRNPC had reduced survival probability relative to patients with high expression. The mean 5 year survival for patients with high hnRNPC expression was 73% relative to 57% for patients with low hnRNPC expression. Therefore, reduced hnRNPC expression is associated with a worse overall prognosis. Taken together these results suggest that hnRNPC may play a protective role against colorectal cancer development.

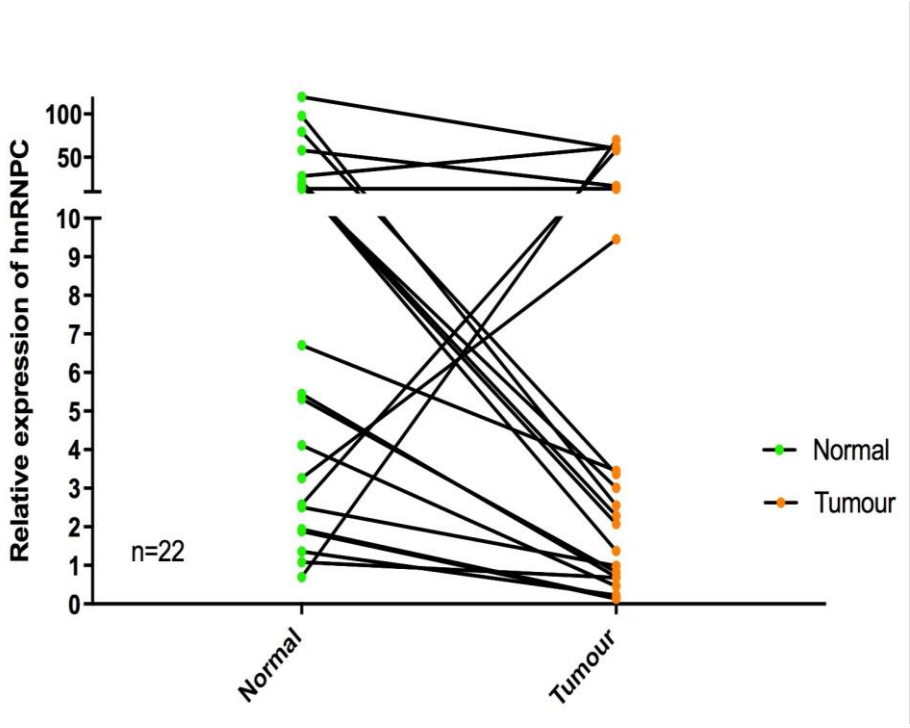


Figure 5.9. hnRNPC expression in colorectal cancer patient samples. Relative expression of hnRNPC in the normal colonic tissue and matched tumour tissue. Solid circles represent patients; black lines connect the normal and tumour tissue samples.

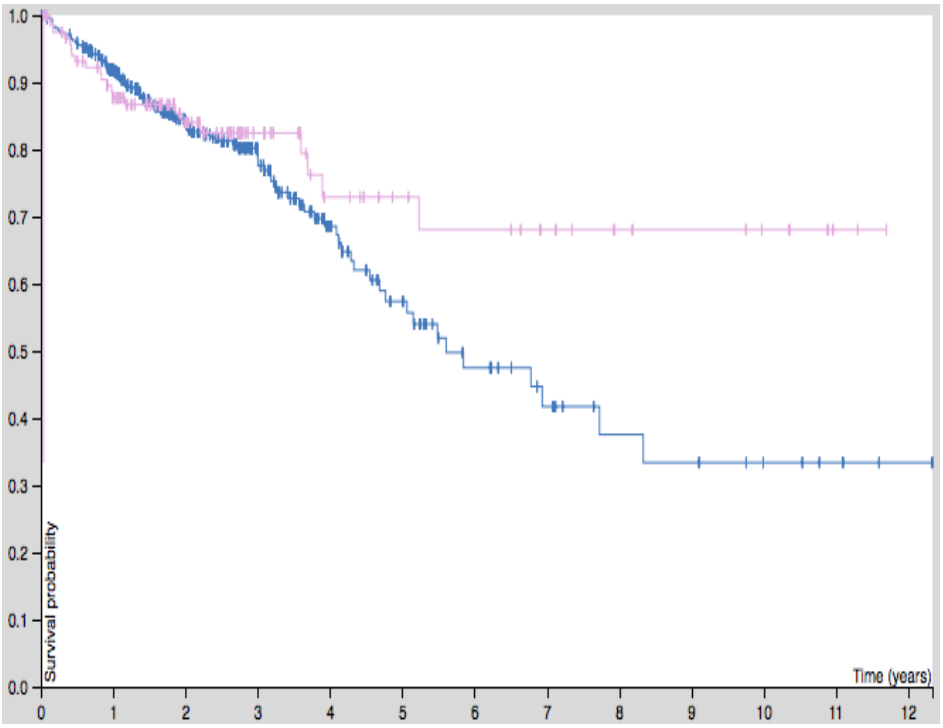


Figure 5.10. Kaplan Meier plot of colorectal cancer survival in patients with high (pink) or low (blue) hnRNPC expression. hnRNPC expression measured using RNA sequencing of 597 colorectal cancer patients, data taken from the human protein atlas [185].

5.7 Summary

In this Chapter a novel mechanism for regulation of *GRB2* exon 4 splicing by hnRNPC is highlighted. Several splicing factors were identified that bind to this exon, predominantly belonging to the hnRNP family of splicing factors. However, only knockdown of hnRNPC was shown to significantly alter *GRB2* exon 4 splicing. The hnRNP family commonly function as splicing suppressors and thus promote exon skipping and hnRNPC is known to promote exon skipping of a number of target genes including, but not limited to, the muscle-specific RTK (MUSK) and spinal muscular atrophy gene SMN1 [186, 187]. However, this is the first report showing that skipping of *GRB2* exon 4 is regulated by hnRNPC. hnRNPC was shown to bind to a petameric uracil sequence at position 5-10 of exon 4, and overexpression of hnRNPC increased Grb3-3 levels by up to 3.2-fold. hnRNPC expression was reduced in colorectal cancer tissue samples which correlated with reduced Grb3-3 levels, supporting the idea that hnRNPC acts to supress exon 4 inclusion. Furthermore, low expression of hnRNPC was associated with worse prognosis in colorectal cancer. Taken together these results suggest that hnRNPC may play a protective role against colorectal cancer development by enhancing *GRB2* exon 4 skipping, resulting in the production of the truncated isoform Grb3-3 that negatively regulates Ras signalling. Reduced hnRNPC expression results in reduced Grb3-3 production facilitating Ras activation. The regulatory mechanism highlighted here suggests than hnRNPC may serve as a prognostic indicator for colorectal cancer and may also function as a marker for Ras activation that could be used to predict response to anti-Ras therapies.

Chapter 6: Discussion and future work

6.1 Discussion

Alternate splicing represents a major source of transcript variation and proteomic diversity. On average a single gene encodes 7 different transcripts explaining how the human genome, with around 24,000 protein-coding genes, can encode more than 100,000 proteins. Current estimates suggest that 95% of human genes undergo alternate splicing. Mapping of alternate exons onto proteins revealed that they often encode loop regions, or surface regions involved in protein-protein interactions. Therefore, alternate splicing can have a profound effect on protein function and cellular signalling.

More recently, aberrant alternate splicing has emerged as an important contributor to various diseases and in particular cancer. Advances in RNA sequencing technologies have enabled entire transcriptomes to be mapped in health and disease which has highlighted numerous cancer-associated alternate splicing events. Splicing events altered in cancer often facilitate the acquisition of the hallmarks of cancer including evasion of apoptosis, sustained proliferation and angiogenesis; suggesting that alternate splicing contributes directly to tumour development and progression.

In this study alternate splicing of the adaptor protein Grb2 was investigated in colorectal cancer. Grb2 is known to link activated RTKs at the cell surface to the Ras signalling pathway, a pathway frequently dysregulated in colorectal cancer. It does so through a bipartite interaction between its central SH2 domain with phosphotyrosine residues on activated receptors, and its NSH3 domain with proline-rich motifs in the Ras guanine exchange factor (GEF) Sos. This interaction results in the recruitment of Sos to the plasma membrane, where it catalyses the

exchange of GDP for GTP on Ras. Once activated Ras initiates the Raf-Mek-Erk signalling cascade and induces cell proliferation.

Alternate splicing of Grb2 produces a truncated isoform denoted Grb3-3 that contains a 40-residue deletion in the central SH2 domain containing the phosphotyrosine binding site. This isoform is produced by the exclusion of alternate exon 4 during splicing. Interestingly the deleted region in Grb3-3 is also encoded by a single exon in *Caenorhabditis elegans sem5* and Drosophila *drk* genes, suggesting that the splice variant may be conserved. In this study a splice-sensitive PCR assay was developed to enable the abundance of the two transcripts to be quantified in normal and matched tumour samples. This revealed that *GRB2* splicing was dysregulated in colorectal cancer clinical samples. Grb3-3 expression was reduced in the tumour tissue relative to the surrounding normal colonic tissue, and Grb2 expression was increased resulting in altered ratios of the two alternatively spliced transcripts. Furthermore an inverse correlation between the levels of Grb3-3 and Ras activation was observed. Patients where Grb3-3 expression decreased by greater than 2-fold in the tumour tissue showed a clear increase in Raf1 serine 338 phosphorylation. Whereas patients where Grb3-3 levels remained constant showed no change in Raf1 phosphorylation, indicating that loss of Grb3-3 is sufficient to induce Ras activation and downstream signalling. To further validate these results a larger number of patient samples would be required. Microarrays are a useful approach for high throughput expression analysis. Probes could be developed that target the unique exons or exon-exon junctions in the two transcripts. In addition, probes that analyse expression of Ras-responsive elements would also be useful to provide further evidence that changes in *GRB2* splicing alter Ras signalling. In order to study the levels of Raf-1 phosphorylation in a greater cohort of patient

samples, tissue microarrays could be used. TMAs are produced from small tissue cores of around 0.6mm in diameter from multiple patients printed in a specific array pattern, allowing staining in several different samples to be analysed at the same time. While TMAs are useful for collecting large quantities of data and use a lot less reagents, they provide a lot less information about tissue heterogeneity since the tissue size is much smaller. The IHC data presented here showed that Ras activation was localised predominantly to the colonic cyrpts and in particular the lower third, where the cells are known to be proliferating. Nevertheless, preliminary data suggests that changes in Grb2 splicing that reduces the levels of Grb3-3 can induce Ras activation.

A combination of biophysical, molecular and cellular biology experiments provide further evidence that Grb3-3 negatively regulates Ras activation in cells and highlight a novel molecular mechanism. Exogenous expression of Grb3-3 inhibits Ras activation and downstream Erk signalling in response to growth factor stimulation in HEK293T and Caco2 cell lines resulting in inhibition of cell proliferation. This inhibition was, at least in part, a consequence of the non-functional SH2 domain Exogenous expression of Grb2 did not alter cell proliferation, whereas the Grb2 R82L mutant (where the SH2 domain is non-functional) did, although to a lesser extent. Characterisation of the interaction between Grb3-3 and Sos unveiled the molecular mechanism for inhibition of Ras activation. Pull-down and FRET experiments showed that Grb3-3 is constitutively associated with Sos. This differs from Grb2, where binding to Sos is regulated by dimerization, which largely prevents binding to Sos in the absence of growth factor stimulation. MST experiments using proline-rich peptides derived from Sos, allowed the binding sites for Grb3-3 on Sos to be identified. Grb3-3 binds to three proline-rich motifs in the C-terminus of Sos, previously reported to form the

Grb2 binding site via its N-terminal SH3. A novel proline-rich binding site for Grb3-3 was identified in the catalytic CDC25 domain. This binding site impinges on the allosteric Ras binding site in Sos, however the function of this interaction has not been interrogated further in this study. The ability of both Grb2 and Grb3-3 to bind to proline-rich motifs in the C-terminus of Sos suggests that, in cells where both isoforms exist, two protein complexes can form: Grb2:Sos and Grb3-3:Sos. The concentration of the two isoforms will dictate which complex prevails. Pull-down and MST experiments confirmed that the two isoforms do indeed compete for binding to Sos, and binding of Grb3-3 directly inhibits Grb2 binding. The interaction between Grb2 and Sos is well known to be required to mediate recruitment of Sos to the plasma membrane, whereas binding of Grb3-3 was shown to result in sequestering of Sos in the cytoplasm. Therefore a novel molecular mechanism whereby the two protein isoforms compete for binding to the proline-rich C-terminus of Sos which regulates its cellular localisation can be proposed. When Grb3-3 prevails Sos is sequestered in the cytoplasm away from its substrate Ras thus preventing nucleotide exchange. When Grb2 prevails Sos is recruited to the membrane allowing Ras activation and Erk signalling.

Data shown here highlights that the splice variants Grb2 and Grb3-3 have opposing functions in Ras signalling, thus changes in the expression level of the two isoforms can fine tune activation of Ras and subsequent cellular proliferation. Furthermore, it suggests that the adaptor proteins, not only play a passive role in signalling, but have key regulatory functions. This cellular system is reminiscent of regulation of apoptosis by Bclx. Alternate spicing of Bclx produces two transcripts through the use of an alternate 5’ splice site, the larger transcript denoted Bclx-L and a shorter transcript Bclx-s. Bclx-l is anti-apoptotic whereas Bclx-s promotes

apoptosis. Splicing of Bclx is altered in many cancers to favour the production of the longer transcript contributing to evasion of apoptosis.

One mechanism that can lead to changes in alternate splicing is altered expression of the splicing regulatory proteins. Splicing factors bind to cis-regulatory sequences found in both the exons and introns and act to supress or enhance the inclusion of alternate exons. Intronic splicing silencers are much less well characterised than exonic splicing silencers, therefore the binding of splicing factors to *GRB2* exon 4 was investigated. An *in silico* approach was first adopted to identify known binding sites for splicing factors within this sequence. This identified 20 unique splicing factor binding sites, predominantly localised in the first 50 bases of the exon. RNA:Protein pull down experiments using synthetic desthiobiotin-labelled RNA spanning this exon, followed by mass spectrometry to identify bound proteins, confirmed the binding of 11 of these splicing factors. The heterogeneous nuclear ribonuclear proteins were the major family of splicing factors identified as binding to this exon. This family of splicing factors are typically reported to function as exonic suppressors suggesting binding may inhibit inclusion of exon 4. The effect of the splicing factors on the inclusion of exon 4 was investigated using publicly available RNAseq datasets, where each of the splicing factors had been knocked down using Crispr/Cas9 followed by sequencing of the entire transcriptome. rMATS analysis was used to identify significantly altered splicing events between the knockdown cells and control cells, revealing that knockdown of hnRNPC significantly altered exon 4 inclusion. Given that this splicing event was shown to be altered in colorectal cancer, expression of hnRNPC was analysed in the matched tumour tissue samples. The results showed hnRNPC expression is down regulated in colorectal cancer which has previously been associated with a worse overall prognosis.

To conclude, this thesis highlights a novel mechanism whereby splicing of the adaptor protein *GRB2* can fine tune Ras signalling. Inclusion of exon 4 regulated by hnRNPC produces Grb2 which mediates Ras activation whereas exclusion produces Grb3-3 which inhibits Ras activation. Preliminary data suggests that this regulatory mechanism is important in colorectal cancer. In non-cancerous colon tissue high levels of hnRNPC are expressed which promotes production of the inhibitory isoform Grb3-3 preventing Ras signalling and proliferation. In colorectal cancer, hnRNPC expression is downregulated resulting in down regulation of Grb3-3 and upregulation of Ras signalling. Elevated Ras signalling is a common feature of many cancers and hnRNPC expression has also been shown to be downregulated in thyroid, breast and ovarian cancer, therefore it would be interesting to see if *GRB2* splicing is also dysregulated in these tumours, or if this mechanism is unique to colorectal cancer.

6.2 Future Perspectives

This thesis highlights a novel molecular mechanism for the regulation of Ras signalling in cancer by the opposing action of the splice variants Grb2 and Grb3-3. Knowledge of this regulatory mechanism reveals potential novel strategies that could be exploited therapeutically to inhibit Ras activation. Inhibition of Ras signalling has been at the heart of anti-cancer drug development for several decades. Despite significant efforts there are still no FDA-approved Ras inhibitors. Further studies would allow two potential therapeutic strategies to explored:

1. Developing small molecule inhibitors of the Grb2-Sos complex
2. Correction of *GRB2* splicing in cancer

Data presented in Chapter 3 shows that Grb3-3 competitively inhibits Grb2 from binding to proline-rich motifs in the C-terminus of Sos which prevents Sos

membrane-localisation and Ras activation. Therefore, the design of small molecules that bind to the same site on Sos as Grb3-3 will theoretically inhibit Ras activation. Although this study identifies the specific proline-rich motifs that bind to Sos using short peptides structure based drug design would require a high resolution structure of the Grb3-3 Sos complex. Sos1 is a relatively large protein with a molecular weight of 152kDa, the complex with Grb3-3 would be 172kDa, making it suitable for analysis via electron microscopy. Solving a high-resolution structure of the Grb3-3 Sos complex would enable the binding site to be more precisely mapped and would reveal information about the types of interactions between different atoms. This structure could then be used in docking studies to screen for small molecules that can occupy the same site. Electron microscopy structures can now be resolved at high enough resolution to enable binding of small molecules to be seen. The most promising drugs from docking screens could then be taken forward into structure-based drug design studies to generate a compound that mimics the Grb3-3-Sos interaction. These compounds could then be tested in cell lines for their ability to inhibit Ras activation and Erk signalling. The ability of Grb3-3 to inhibit Ras activation is not only due to the competitive inhibition of Grb2 binding but also because binding sequesters Sos in the cytoplasm away from its substrate Ras. To mimic this exogenously cytoplasmic retention sequences could be added to the small molecules, so they not only prevent Grb2 binding, but also upon binding to Sos sequester it in the cytoplasm.

Another strategy to inhibit Ras activation would be to manipulate *GRB2* splicing to increase production of the inhibitory isoform. This may be particularly important in colorectal cancer where preliminary data suggests that this dysregulation of this splicing facilitates upregulation of Ras signalling in the absence of mutations.

Splice switching oligonucleotides (SSOs) have been developed which can alter splicing events to favour the production of isoforms with anti-tumour properties. The development of SSOs that enhance exon 4 exclusion would increase the production of the truncated inhibitory isoform Grb3-3 which could potentially down-regulate Ras signalling in cancer by stimulating the endogenous regulatory mechanism.

Finally, this thesis identified a novel interaction between Grb3-3 and the catalytic domain of Sos which has not been previously reported on. However, the function of the interaction with this site on Sos was not explored in this study due to time constraints. Further work would enable the function of this interaction in cells to be characterised which may reveal an additional or complementary mechanism for regulation of Sos by Grb3-3. Interestingly, the binding site in the catalytic domain identified in this study appears to impinge on the allosteric Ras binding site which is required for positive feedback activation of Sos. Furthermore, proline 924 and 925 which form part of the proline-rich motif shown to bind Grb3-3 have been reported to be required to correctly orientate residues in Sos to allow binding to allosteric Ras-GTP. Therefore, binding of Grb3-3 to this site may directly block Ras binding, or may alter the conformation of Sos to an orientation that is not conducive for binding. The effect of this interaction on Sos activity could be investigated using an *in vitro* guanine nucleotide exchange assay. These assays measure the exchange of fluorescently labelled GDP Ras for unlabelled GTP in the presence of the recombinantly expressed catalytic domain. Performing these assays in the presence and absence of Grb3-3 would reveal insights into the function of the Grb3-3 interaction with the catalytic domain of Sos.

References

1. Harris H, Watts JW: **The relationship between nuclear and cytoplasmic ribonucleic acid**. *Proceedings of the Royal Society of London Series B, Biological sciences* 1962, **156**:109-121.
2. Chow LT, Gelinas RE, Broker TR, Roberts RJ: **An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA**. *Cell* 1977, **12**(1):1-8.
3. Berget SM, Moore C, Sharp PA: **Spliced segments at the 5' terminus of adenovirus 2 late mRNA. 1977**. *Reviews in medical virology* 2000, **10**(6):356-362; discussion 355-356.
4. Mandel JL, Breathnach R, Gerlinger P, Le Meur M, Gannon F, Chambon P: **Organization of coding and intervening sequences in the chicken ovalbumin split gene**. *Cell* 1978, **14**(3):641-653.
5. Dvinge H, Kim E, Abdel-Wahab O, Bradley RK: **RNA splicing factors as oncoproteins and tumour suppressors**. *Nature reviews Cancer* 2016, **16**(7):413-430.
6. Langford CJ, Klinz FJ, Donath C, Gallwitz D: **Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast**. *Cell* 1984, **36**(3):645-653.
7. Black DL: **Mechanisms of alternative pre-messenger RNA splicing**. *Annu Rev Biochem* 2003, **72**:291-336.
8. Ruskin B, Krainer AR, Maniatis T, Green MR: **Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro**. *Cell* 1984, **38**(1):317-331.
9. Moore MJ, Sharp PA: **Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing**. *Nature* 1993, **365**(6444):364-368.
10. Konarska MM, Grabowski PJ, Padgett RA, Sharp PA: **Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors**. *Nature* 1985, **313**(6003):552-557.
11. Will CL, Luhrmann R: **Spliceosome structure and function**. *Cold Spring Harbor perspectives in biology* 2011, **3**(7).
12. Jurica MS, Moore MJ: **Pre-mRNA splicing: awash in a sea of proteins**. *Molecular cell* 2003, **12**(1):5-14.
13. Matlin AJ, Moore MJ: **Spliceosome assembly and composition**. *Advances in experimental medicine and biology* 2007, **623**:14-35.
14. Will CL, Luhrmann R: **Spliceosomal UsnRNP biogenesis, structure and function**. *Current opinion in cell biology* 2001, **13**(3):290-301.
15. Chiou NT, Lynch KW: **Mechanisms of spliceosomal assembly**. *Methods in molecular biology (Clifton, NJ)* 2014, **1126**:35-43.
16. Staley JP, Guthrie C: **Mechanical devices of the spliceosome: motors, clocks, springs, and things**. *Cell* 1998, **92**(3):315-326.
17. Scotti MM, Swanson MS: **RNA mis-splicing in disease**. *Nature reviews Genetics* 2016, **17**(1):19-32.
18. Baralle D, Buratti E: **RNA splicing in human disease and in the clinic**. *Clinical science (London, England : 1979)* 2017, **131**(5):355-368.
19. Wang Z, Burge CB: **Splicing regulation: from a parts list of regulatory elements to an integrated splicing code**. *RNA* 2008, **14**(5):802-813.

20. McVety S, Li L, Gordon PH, Chong G, Foulkes WD: **Disruption of an exon splicing enhancer in exon 3 of MLH1 is the cause of HNPCC in a Quebec family**. *J Med Genet* 2006, **43**(2):153-156.
21. Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB: **Systematic identification and analysis of exonic splicing silencers**. *Cell* 2004, **119**(6):831-845.
22. Giulietti M, Piva F, D'Antonio M, D'Onorio De Meo P, Paoletti D, Castrignano T, D'Erchia AM, Picardi E, Zambelli F, Principato G *et al*: **SpliceAid-F: a database of human splicing factors and their RNA-binding sites**. *Nucleic acids research* 2013, **41**(Database issue):D125-131.
23. Piva F, Giulietti M, Burini AB, Principato G: **SpliceAid 2: a database of human splicing factors expression data and RNA target motifs**. *Hum Mutat* 2012, **33**(1):81-85.
24. Piva F, Giulietti M, Nocchi L, Principato G: **SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans**. *Bioinformatics* 2009, **25**(9):1211-1213.
25. Shepard PJ, Hertel KJ: **The SR protein family**. *Genome biology* 2009, **10**(10):242.
26. Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, Duncan PI: **The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution**. *EMBO J* 1996, **15**(2):265-275.
27. Gui JF, Lane WS, Fu XD: **A serine kinase regulates intracellular localization of splicing factors in the cell cycle**. *Nature* 1994, **369**(6482):678-682.
28. Graveley BR: **Sorting out the complexity of SR protein functions**. *Rna* 2000, **6**(9):1197-1211.
29. Graveley BR, Maniatis T: **Arginine/serine-rich domains of SR proteins can function as activators of pre-mRNA splicing**. *Molecular cell* 1998, **1**(5):765-771.
30. Geuens T, Bouhy D, Timmerman V: **The hnRNP family: insights into their role in health and disease**. *Hum Genet* 2016, **135**(8):851-867.
31. Han SP, Tang YH, Smith R: **Functional diversity of the hnRNPs: past, present and perspectives**. *The Biochemical journal* 2010, **430**(3):379-392.
32. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ: **Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing**. *Nat Genet* 2008, **40**(12):1413-1415.
33. de la Grange P, Gratadou L, Delord M, Dutertre M, Auboeuf D: **Splicing factor and exon profiling across human tissues**. *Nucleic acids research* 2010, **38**(9):2825-2838.
34. Kahles A, Lehmann KV, Toussaint NC, Huser M, Stark SG, Sachsenberg T, Stegle O, Kohlbacher O, Sander C, Ratsch G: **Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients**. *Cancer cell* 2018, **34**(2):211-224.e216.
35. Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI: **Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes**. *Oncogene* 2016, **35**(19):2413-2427.
36. Ellis JD, Barrios-Rodiles M, Colak R, Irimia M, Kim T, Calarco JA, Wang X, Pan Q, O'Hanlon D, Kim PM *et al*: **Tissue-specific alternative splicing remodels protein-protein interaction networks**. *Molecular cell* 2012, **46**(6):884-892.
37. Ladd AN, Cooper TA: **Finding signals that regulate alternative splicing in the post-genomic era**. *Genome biology* 2002, **3**(11):reviews0008.

38. Wang Y, Liu J, Huang BO, Xu YM, Li J, Huang LF, Lin J, Zhang J, Min QH, Yang WM *et al*: **Mechanism of alternative splicing and its regulation**. *Biomed Rep* 2015, **3**(2):152-158.
39. Snezhkina AV, Krasnov GS, Zaretsky AR, Zhavoronkov A, Nyushko KM, Moskalev AA, Karpova IY, Afremova AI, Lipatova AV, Kochetkov DV *et al*: **Differential expression of alternatively spliced transcripts related to energy metabolism in colorectal cancer**. *BMC Genomics* 2016, **17**(Suppl 14):1011.
40. Bisognin A, Pizzini S, Perilli L, Esposito G, Mocellin S, Nitti D, Zanovello P, Bortoluzzi S, Mandruzzato S: **An integrative framework identifies alternative splicing events in colorectal cancer development**. *Mol Oncol* 2014, **8**(1):129-141.
41. Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, Veitch J, Schweitzer A, Awad T, Sugnet C, Dee S *et al*: **Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array**. *BMC Genomics* 2006, **7**:325.
42. Cheng C, Sharp PA: **Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion**. *Molecular and cellular biology* 2006, **26**(1):362-370.
43. Moore MJ, Wang Q, Kennedy CJ, Silver PA: **An alternative splicing network links cell-cycle control to apoptosis**. *Cell* 2010, **142**(4):625-636.
44. Schwerk C, Schulze-Osthoff K: **Regulation of apoptosis by alternative pre-mRNA splicing**. *Molecular cell* 2005, **19**(1):1-13.
45. Ferreira PG, Jares P, Rico D, Gomez-Lopez G, Martinez-Trillos A, Villamor N, Ecker S, Gonzalez-Perez A, Knowles DG, Monlong J *et al*: **Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia**. *Genome research* 2014, **24**(2):212-226.
46. Martinez-Montiel N, Rosas-Murrieta NH, Anaya Ruiz M, Monjaraz-Guzman E, Martinez-Contreras R: **Alternative Splicing as a Target for Cancer Treatment**. *International journal of molecular sciences* 2018, **19**(2).
47. Srebrow A, Kornblihtt AR: **The connection between splicing and cancer**. *Journal of cell science* 2006, **119**(Pt 13):2635-2641.
48. Grosso AR, Martins S, Carmo-Fonseca M: **The emerging role of splicing factors in cancer**. *EMBO reports* 2008, **9**(11):1087-1093.
49. Wan L, Yu W, Shen E, Sun W, Liu Y, Kong J, Wu Y, Han F, Zhang L, Yu T *et al*: **SRSF6-regulated alternative splicing that promotes tumour progression offers a therapy target for colorectal cancer**. *Gut* 2019, **68**(1):118-129.
50. Piekielko-Witkowska A, Wiszomirska H, Wojcicka A, Poplawski P, Boguslawska J, Tanski Z, Nauman A: **Disturbed expression of splicing factors in renal cancer affects alternative splicing of apoptosis regulators, oncogenes, and tumor suppressors**. *PloS one* 2010, **5**(10):e13690.
51. Jeong HM, Han J, Lee SH, Park HJ, Lee HJ, Choi JS, Lee YM, Choi YL, Shin YK, Kwon MJ: **ESRP1 is overexpressed in ovarian cancer and promotes switching from mesenchymal to epithelial phenotype in ovarian cancer cells**. *Oncogenesis* 2017, **6**(10):e389.
52. Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR: **The gene encoding the splicing factor SF2/ASF is a proto-oncogene**. *Nature structural & molecular biology* 2007, **14**(3):185-193.

53. Anczukow O, Rosenberg AZ, Akerman M, Das S, Zhan L, Karni R, Muthuswamy SK, Krainer AR: **The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation**. *Nature structural & molecular biology* 2012, **19**(2):220-228.
54. Moon H, Cho S, Loh TJ, Oh HK, Jang HN, Zhou J, Kwon YS, Liao DJ, Jun Y, Eom S *et al*: **SRSF2 promotes splicing and transcription of exon 11 included isoform in Ron proto-oncogene**. *Biochimica et biophysica acta* 2014, **1839**(11):1132-1140.
55. Chen M, Zhang J, Manley JL: **Turning on a fuel switch of cancer: hnRNP proteins regulate alternative splicing of pyruvate kinase mRNA**. *Cancer research* 2010, **70**(22):8977-8980.
56. Sakuma K, Sasaki E, Kimura K, Komori K, Shimizu Y, Yatabe Y, Aoki M: **HNRNPLL, a newly identified colorectal cancer metastasis suppressor, modulates alternative splicing of CD44 during epithelial-mesenchymal transition**. *Gut* 2018, **67**(6):1103-1111.
57. Bejar R: **Splicing Factor Mutations in Cancer**. *Advances in experimental medicine and biology* 2016, **907**:215-228.
58. Wan Y, Wu CJ: **SF3B1 mutations in chronic lymphocytic leukemia**. *Blood* 2013, **121**(23):4627-4634.
59. Brooks AN, Choi PS, de Waal L, Sharifnia T, Imielinski M, Saksena G, Pedamallu CS, Sivachenko A, Rosenberg M, Chmielecki J *et al*: **A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events**. *PloS one* 2014, **9**(1):e87361.
60. Lin JC: **Therapeutic Applications of Targeted Alternative Splicing to Cancer Treatment**. *International journal of molecular sciences* 2017, **19**(1).
61. Stahel RA, Zangemeister-Wittke U: **Antisense oligonucleotides for cancer therapy-an overview**. *Lung cancer (Amsterdam, Netherlands)* 2003, **41 Suppl 1**:S81-88.
62. Sazani P, Kole R: **Therapeutic potential of antisense oligonucleotides as modulators of alternative splicing**. *The Journal of clinical investigation* 2003, **112**(4):481-486.
63. Mercatante DR, Bortner CD, Cidlowski JA, Kole R: **Modification of alternative splicing of Bcl-x pre-mRNA in prostate and breast cancer cells. analysis of apoptosis and cell death**. *The Journal of biological chemistry* 2001, **276**(19):16411-16417.
64. Havens MA, Hastings ML: **Splice-switching antisense oligonucleotides as therapeutic drugs**. *Nucleic acids research* 2016, **44**(14):6549-6563.
65. Li Z, Li Q, Han L, Tian N, Liang Q, Li Y, Zhao X, Du C, Tian Y: **Pro-apoptotic effects of splice-switching oligonucleotides targeting Bcl-x pre-mRNA in human glioma cell lines**. *Oncology reports* 2016, **35**(2):1013-1019.
66. Taylor JK, Zhang QQ, Wyatt JR, Dean NM: **Induction of endogenous Bcl-xS through the control of Bcl-x pre-mRNA splicing by antisense oligonucleotides**. *Nature biotechnology* 1999, **17**(11):1097-1100.
67. Gembarska A, Luciani F, Fedele C, Russell EA, Dewaele M, Villar S, Zwolinska A, Haupt S, de Lange J, Yip D *et al*: **MDM4 is a key therapeutic target in cutaneous melanoma**. *Nature medicine* 2012, **18**(8):1239-1247.
68. Haupt S, Mejia-Hernandez JO, Vijayakumaran R, Keam SP, Haupt Y: **The long and the short of it: the MDM4 tail so far**. *Journal of molecular cell biology* 2019, **11**(3):231-244.

69. Dewaele M, Tabaglio T, Willekens K, Bezzi M, Teo SX, Low DH, Koh CM, Rambow F, Fiers M, Rogiers A *et al*: **Antisense oligonucleotide-mediated MDM4 exon 6 skipping impairs tumor growth**. *The Journal of clinical investigation* 2016, **126**(1):68-84.
70. Nancy MM, Nora RM, Rebeca MC: **Peptidic tools applied to redirect alternative splicing events**. *Peptides* 2015, **67**:1-11.
71. Gammons MV, Lucas R, Dean R, Coupland SE, Oltean S, Bates DO: **Targeting SRPK1 to control VEGF-mediated tumour angiogenesis in metastatic melanoma**. *British journal of cancer* 2014, **111**(3):477-485.
72. Siqueira RP, Barbosa Ede A, Poleto MD, Righetto GL, Seraphim TV, Salgado RL, Ferreira JG, Barros MV, de Oliveira LL, Laranjeira AB *et al*: **Potential Antileukemia Effect and Structural Analyses of SRPK Inhibition by N-(2-(Piperidin-1-yl)-5-(Trifluoromethyl)Phenyl)Isonicotinamide (SRPIN340)**. *PloS one* 2015, **10**(8):e0134882.
73. Araki S, Dairiki R, Nakayama Y, Murai A, Miyashita R, Iwatani M, Nomura T, Nakanishi O: **Inhibitors of CLK protein kinases suppress cell growth and induce apoptosis by modulating pre-mRNA splicing**. *PloS one* 2015, **10**(1):e0116929.
74. Sako Y, Ninomiya K, Okuno Y, Toyomoto M, Nishida A, Koike Y, Ohe K, Kii I, Yoshida S, Hashimoto N *et al*: **Development of an orally available inhibitor of CLK1 for skipping a mutated dystrophin exon in Duchenne muscular dystrophy**. *Scientific reports* 2017, **7**:46126.
75. Martinez-Montiel N, Rosas-Murrieta NH, Martinez-Montiel M, Gaspariano-Cholula MP, Martinez-Contreras RD: **Microbial and Natural Metabolites That Inhibit Splicing: A Powerful Alternative for Cancer Treatment**. *BioMed research international* 2016, **2016**:3681094.
76. Corrionero A, Minana B, Valcarcel J: **Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A**. *Genes & development* 2011, **25**(5):445-459.
77. O'Brien K, Matlin AJ, Lowell AM, Moore MJ: **The biflavonoid isoginkgetin is a general inhibitor of Pre-mRNA splicing**. *The Journal of biological chemistry* 2008, **283**(48):33147-33154.
78. He RQ, Zhou XG, Yi QY, Deng CW, Gao JM, Chen G, Wang QY: **Prognostic Signature of Alternative Splicing Events in Bladder Urothelial Carcinoma Based on Spliceseq Data from 317 Cases**. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 2018, **48**(3):1355-1368.
79. Zhang Z, Pal S, Bi Y, Tchou J, Davuluri RV: **Isoform level expression profiles provide better cancer signatures than gene level expression profiles**. *Genome Med* 2013, **5**(4):33.
80. Trincado JL, Sebestyen E, Pages A, Eyras E: **The prognostic potential of alternative transcript isoforms across human tumors**. *Genome Med* 2016, **8**(1):85.
81. Birzele F, Voss E, Nopora A, Honold K, Heil F, Lohmann S, Verheul H, Le Tourneau C, Delord JP, van Herpen C *et al*: **CD44 Isoform Status Predicts Response to Treatment with Anti-CD44 Antibody in Cancer Patients**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2015, **21**(12):2753-2762.
82. Lowenstein EJ, Daly RJ, Batzer AG, Li W, Margolis B, Lammers R, Ullrich A, Skolnik EY, Bar-Sagi D, Schlessinger J: **The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling**. *Cell* 1992, **70**(3):431-442.

83. Suen KL, Bustelo XR, Pawson T, Barbacid M: **Molecular cloning of the mouse grb2 gene: differential interaction of the Grb2 adaptor protein with epidermal growth factor and nerve growth factor receptors**. *Molecular and cellular biology* 1993, **13**(9):5500-5512.
84. Downward J: **The GRB2/Sem-5 adaptor protein**. *FEBS letters* 1994, **338**(2):113-117.
85. Ladbury JE, Arold S: **Searching for specificity in SH domains**. *Chem Biol* 2000, **7**(1):R3-8.
86. Cantley LC, Songyang Z: **Specificity in recognition of phosphopeptides by src-homology 2 domains**. *Journal of cell science Supplement* 1994, **18**:121-126.
87. Sparks AB, Rider JE, Hoffman NG, Fowlkes DM, Quillam LA, Kay BK: **Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCgamma, Crk, and Grb2**. *Proc Natl Acad Sci U S A* 1996, **93**(4):1540-1544.
88. Lewitzky M, Kardinal C, Gehring NH, Schmidt EK, Konkol B, Eulitz M, Birchmeier W, Schaeper U, Feller SM: **The C-terminal SH3 domain of the adapter protein Grb2 binds with high affinity to sequences in Gab1 and SLP-76 which lack the SH3-typical P-x-x-P core motif**. *Oncogene* 2001, **20**(9):1052-1062.
89. Fath I, Schweighoffer F, Rey I, Multon MC, Boiziau J, Duchesne M, Tocque B: **Cloning of a Grb2 isoform with apoptotic properties**. *Science (New York, NY)* 1994, **264**(5161):971-974.
90. Li N, Batzer A, Daly R, Yajnik V, Skolnik E, Chardin P, Bar-Sagi D, Margolis B, Schlessinger J: **Guanine-nucleotide-releasing factor hSos1 binds to Grb2 and links receptor tyrosine kinases to Ras signalling**. *Nature* 1993, **363**(6424):85-88.
91. Plotnikov A, Zehorai E, Procaccia S, Seger R: **The MAPK cascades: signaling components, nuclear roles and mechanisms of nuclear translocation**. *Biochimica et biophysica acta* 2011, **1813**(9):1619-1633.
92. McKay MM, Morrison DK: **Integrating signals from RTKs to ERK/MAPK**. *Oncogene* 2007, **26**(22):3113-3121.
93. Lemmon MA, Schlessinger J: **Cell signaling by receptor tyrosine kinases**. *Cell* 2010, **141**(7):1117-1134.
94. Lin CC, Melo FA, Ghosh R, Suen KM, Stagg LJ, Kirkpatrick J, Arold ST, Ahmed Z, Ladbury JE: **Inhibition of basal FGF receptor signaling by dimeric Grb2**. *Cell* 2012, **149**(7):1514-1524.
95. Egan SE, Giddings BW, Brooks MW, Buday L, Sizeland AM, Weinberg RA: **Association of Sos Ras exchange protein with Grb2 is implicated in tyrosine kinase signal transduction and transformation**. *Nature* 1993, **363**(6424):45-51.
96. Rozakis-Adcock M, Fernley R, Wade J, Pawson T, Bowtell D: **The SH2 and SH3 domains of mammalian Grb2 couple the EGF receptor to the Ras activator mSos1**. *Nature* 1993, **363**(6424):83-85.
97. Gale NW, Kaplan S, Lowenstein EJ, Schlessinger J, Bar-Sagi D: **Grb2 mediates the EGF-dependent activation of guanine nucleotide exchange on Ras**. *Nature* 1993, **363**(6424):88-92.
98. Skolnik EY, Batzer A, Li N, Lee CH, Lowenstein E, Mohammadi M, Margolis B, Schlessinger J: **The function of GRB2 in linking the insulin receptor to Ras signaling pathways**. *Science (New York, NY)* 1993, **260**(5116):1953-1955.

99. Chardin P, Camonis JH, Gale NW, van Aelst L, Schlessinger J, Wigler MH, Bar-Sagi D: **Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2**. *Science (New York, NY)* 1993, **260**(5112):1338-1343.
100. Sastry L, Lin W, Wong WT, Di Fiore PP, Scoppa CA, King CR: **Quantitative analysis of Grb2-Sos1 interaction: the N-terminal SH3 domain of Grb2 mediates affinity**. *Oncogene* 1995, **11**(6):1107-1112.
101. Simon JA, Schreiber SL: **Grb2 SH3 binding to peptides from Sos: evaluation of a general model for SH3-ligand interactions**. *Chem Biol* 1995, **2**(1):53-60.
102. McDonald CB, Seldeen KL, Deegan BJ, Farooq A: **Structural basis of the differential binding of the SH3 domains of Grb2 adaptor to the guanine nucleotide exchange factor Sos1**. *Arch Biochem Biophys* 2008, **479**(1):52-62.
103. Ahmed Z, Timsah Z, Suen KM, Cook NP, Lee GRt, Lin CC, Gagea M, Marti AA, Ladbury JE: **Grb2 monomer-dimer equilibrium determines normal versus oncogenic function**. *Nature communications* 2015, **6**:7354.
104. Rojas JM, Oliva JL, Santos E: **Mammalian son of sevenless Guanine nucleotide exchange factors: old concepts and new perspectives**. *Genes & cancer* 2011, **2**(3):298-305.
105. Boriack-Sjodin PA, Margarit SM, Bar-Sagi D, Kuriyan J: **The structural basis of the activation of Ras by Sos**. *Nature* 1998, **394**(6691):337-343.
106. Santos E, Nebreda AR: **Structural and functional properties of ras proteins**. *Faseb j* 1989, **3**(10):2151-2163.
107. Bos JL, Rehmann H, Wittinghofer A: **GEFs and GAPs: critical elements in the control of small G proteins**. *Cell* 2007, **129**(5):865-877.
108. Findlay GM, Pawson T: **How is SOS activated? Let us count the ways**. *Nature structural & molecular biology* 2008, **15**(6):538-540.
109. Margarit SM, Sondermann H, Hall BE, Nagar B, Hoelz A, Pirruccello M, Bar-Sagi D, Kuriyan J: **Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS**. *Cell* 2003, **112**(5):685-695.
110. Vo U, Vajpai N, Flavell L, Bobby R, Breeze AL, Embrey KJ, Golovanov AP: **Monitoring Ras Interactions with the Nucleotide Exchange Factor Son of Sevenless (Sos) Using Site-specific NMR Reporter Signals and Intrinsic Fluorescence**. *The Journal of biological chemistry* 2016, **291**(4):1703-1718.
111. Dumaz N, Marais R: **Raf phosphorylation: one step forward and two steps back**. *Molecular cell* 2005, **17**(2):164-166.
112. Nan X, Tamguney TM, Collisson EA, Lin LJ, Pitt C, Galeas J, Lewis S, Gray JW, McCormick F, Chu S: **Ras-GTP dimers activate the Mitogen-Activated Protein Kinase (MAPK) pathway**. *Proc Natl Acad Sci U S A* 2015, **112**(26):7996-8001.
113. Chambard JC, Lefloch R, Pouyssegur J, Lenormand P: **ERK implication in cell cycle regulation**. *Biochimica et biophysica acta* 2007, **1773**(8):1299-1310.
114. Burotto M, Chiou VL, Lee JM, Kohn EC: **The MAPK pathway across different malignancies: a new perspective**. *Cancer* 2014, **120**(22):3446-3456.
115. Kohno M, Pouyssegur J: **Targeting the ERK signaling pathway in cancer therapy**. *Ann Med* 2006, **38**(3):200-211.
116. Samatar AA, Poulikakos PI: **Targeting RAS-ERK signalling in cancer: promises and challenges**. *Nat Rev Drug Discov* 2014, **13**(12):928-942.

117. Dhillon AS, Hagan S, Rath O, Kolch W: **MAP kinase signalling pathways in cancer**. *Oncogene* 2007, **26**(22):3279-3290.
118. Yu GZ, Chen Y, Wang JJ: **Overexpression of Grb2/HER2 signaling in Chinese gastric cancer: their relationship with clinicopathological parameters and prognostic significance**. *J Cancer Res Clin Oncol* 2009, **135**(10):1331-1339.
119. Watanabe T, Shinohara N, Moriya K, Sazawa A, Kobayashi Y, Ogiso Y, Takiguchi M, Yasuda J, Koyanagi T, Kuzumaki N *et al*: **Significance of the Grb2 and son of sevenless (Sos) proteins in human bladder cancer cell lines**. *IUBMB Life* 2000, **49**(4):317-320.
120. Pendergast AM, Quilliam LA, Cripe LD, Bassing CH, Dai Z, Li N, Batzer A, Rabun KM, Der CJ, Schlessinger J *et al*: **BCR-ABL-induced oncogenesis is mediated by direct interaction with the SH2 domain of the GRB-2 adaptor protein**. *Cell* 1993, **75**(1):175-185.
121. Dharmawardana PG, Peruzzi B, Giubellino A, Burke TR, Jr., Bottaro DP: **Molecular targeting of growth factor receptor-bound 2 (Grb2) as an anti-cancer strategy**. *Anti-cancer drugs* 2006, **17**(1):13-20.
122. Tari AM, Hung MC, Li K, Lopez-Berestein G: **Growth inhibition of breast cancer cells by Grb2 downregulation is correlated with inactivation of mitogen-activated protein kinase in EGFR, but not in ErbB2, cells**. *Oncogene* 1999, **18**(6):1325-1332.
123. Thomas X, Paubelle E: **Grb2 inhibition: a new potential targeted therapy for myeloid malignancies?** *The Lancet Haematology* 2018, **5**(4):e128-e129.
124. Dong C, Waters SB, Holt KH, Pessin JE: **SOS phosphorylation and disassociation of the Grb2-SOS complex by the ERK and JNK signaling pathways**. *The Journal of biological chemistry* 1996, **271**(11):6328-6332.
125. Waters SB, Yamauchi K, Pessin JE: **Insulin-stimulated disassociation of the SOS-Grb2 complex**. *Molecular and cellular biology* 1995, **15**(5):2791-2799.
126. Cherniack AD, Klarlund JK, Czech MP: **Phosphorylation of the Ras nucleotide exchange factor son of sevenless by mitogen-activated protein kinase**. *The Journal of biological chemistry* 1994, **269**(7):4717-4720.
127. Yu Y, Nie Y, Feng Q, Qu J, Wang R, Bian L, Xia J: **Targeted Covalent Inhibition of Grb2-Sos1 Interaction through Proximity-Induced Conjugation in Breast Cancer Cells**. *Molecular pharmaceutics* 2017, **14**(5):1548-1557.
128. Burke TR: **Development of Grb2 SH2 Domain Signaling Antagonists: A Potential New Class of Antiproliferative Agents**. *International journal of peptide research and therapeutics* 2006, **12**(1):33-48.
129. Gay B, Suarez S, Caravatti G, Furet P, Meyer T, Schoepfer J: **Selective GRB2 SH2 inhibitors as anti-Ras therapy**. *International journal of cancer* 1999, **83**(2):235-241.
130. Gay B, Suarez S, Weber C, Rahuel J, Fabbro D, Furet P, Caravatti G, Schoepfer J: **Effect of potent and selective inhibitors of the Grb2 SH2 domain on cell motility**. *The Journal of biological chemistry* 1999, **274**(33):23311-23315.
131. Soriano JV, Liu N, Gao Y, Yao ZJ, Ishibashi T, Underhill C, Burke TR, Jr., Bottaro DP: **Inhibition of angiogenesis by growth factor receptor bound protein 2-Src homology 2 domain bound antagonists**. *Molecular cancer therapeutics* 2004, **3**(10):1289-1299.

132. Giubellino A, Gao Y, Lee S, Lee MJ, Vasselli JR, Medepalli S, Trepel JB, Burke TR, Jr., Bottaro DP: **Inhibition of tumor metastasis by a growth factor receptor bound protein 2 Src homology 2 domain-binding antagonist**. *Cancer research* 2007, **67**(13):6012-6016.
133. CancerResearchUK: **https://**http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer.
134. CancerResearchUK: **Bowel cancer statistics Available Online at** http://info.cancerresearchuk.org/cancerstats/types/bowel/?script=true. 2010.
135. Santini D, Loupakis F, Vincenzi B, Floriani I, Stasi I, Canestrari E, Rulli E, Maltese PE, Andreoni F, Masi G *et al*: **High concordance of KRAS status between primary colorectal tumors and related metastatic sites: implications for clinical practice**. *The oncologist* 2008, **13**(12):1270-1275.
136. Takayama T, Ohi M, Hayashi T, Miyanishi K, Nobuoka A, Nakajima T, Satoh T, Takimoto R, Kato J, Sakamaki S *et al*: **Analysis of K-ras, APC, and beta-catenin in aberrant crypt foci in sporadic adenoma, cancer, and familial adenomatous polyposis**. *Gastroenterology* 2001, **121**(3):599-611.
137. Pretlow TP, Pretlow TG: **Mutant KRAS in aberrant crypt foci (ACF): initiation of colorectal cancer?** *Biochimica et biophysica acta* 2005, **1756**(2):83-96.
138. Cohen RB: **Epidermal growth factor receptor as a therapeutic target in colorectal cancer**. *Clin Colorectal Cancer* 2003, **2**(4):246-251.
139. Pabla B, Bissonnette M, Konda VJ: **Colon cancer and the epidermal growth factor receptor: Current treatment paradigms, the importance of diet, and the role of chemoprevention**. *World journal of clinical oncology* 2015, **6**(5):133-141.
140. Markman B, Javier Ramos F, Capdevila J, Tabernero J: **EGFR and KRAS in colorectal cancer**. *Adv Clin Chem* 2010, **51**:71-119.
141. Spano JP, Fagard R, Soria JC, Rixe O, Khayat D, Milano G: **Epidermal growth factor receptor signaling in colorectal cancer: preclinical data and therapeutic perspectives**. *Annals of oncology : official journal of the European Society for Medical Oncology* 2005, **16**(2):189-194.
142. Krasinskas AM: **EGFR Signaling in Colorectal Carcinoma**. *Pathology research international* 2011, **2011**:932932.
143. Singh B, Carpenter G, Coffey RJ: **EGF receptor ligands: recent advances**. *F1000Research* 2016, **5**.
144. Amador ML, Hidalgo M: **Epidermal growth factor receptor as a therapeutic target for the treatment of colorectal cancer**. *Clin Colorectal Cancer* 2004, **4**(1):51-62.
145. Heinemann V, Stintzing S, Kirchner T, Boeck S, Jung A: **Clinical relevance of EGFR- and KRAS-status in colorectal cancer patients treated with monoclonal antibodies directed against the EGFR**. *Cancer Treat Rev* 2009, **35**(3):262-271.
146. Oh BY, Lee RA, Chung SS, Kim KH: **Epidermal growth factor receptor mutations in colorectal cancer patients**. *J Korean Soc Coloproctol* 2011, **27**(3):127-132.
147. Krens LL, Baas JM, Gelderblom H, Guchelaar HJ: **Therapeutic modulation of k-ras signaling in colorectal cancer**. *Drug Discov Today* 2010, **15**(13-14):502-516.
148. Porru M, Pompili L, Caruso C, Biroccio A, Leonetti C: **Targeting KRAS in metastatic colorectal cancer: current strategies and emerging**

opportunities. *Journal of experimental & clinical cancer research : CR* 2018, **37**(1):57.

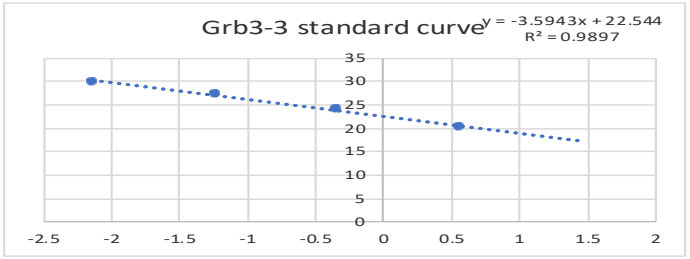
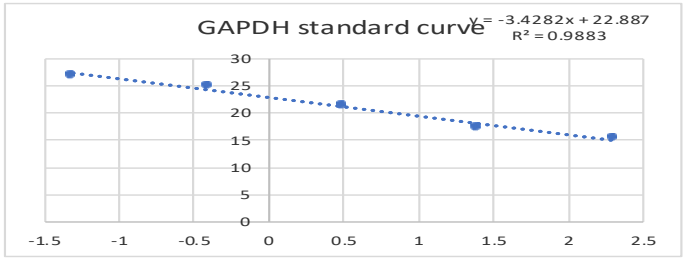
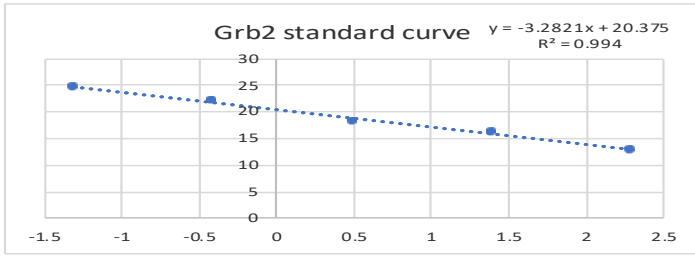
149. Guo F, Gong H, Zhao H, Chen J, Zhang Y, Zhang L, Shi X, Zhang A, Jin H, Zhang J *et al*: **Mutation status and prognostic values of KRAS, NRAS, BRAF and PIK3CA in 353 Chinese colorectal cancer patients**. *Scientific reports* 2018, **8**(1):6076.
150. Zenonos K, Kyprianou K: **RAS signaling pathways, mutations and their role in colorectal cancer**. *World J Gastrointest Oncol* 2013, **5**(5):97-101.
151. Barras D: **BRAF Mutation in Colorectal Cancer: An Update**. *Biomark Cancer* 2015, **7**(Suppl 1):9-12.
152. Lavoie H, Therrien M: **Regulation of RAF protein kinases in ERK signalling**. *Nat Rev Mol Cell Biol* 2015, **16**(5):281-298.
153. Chong H, Guan KL: **Regulation of Raf through phosphorylation and N terminus-C terminus interaction**. *The Journal of biological chemistry* 2003, **278**(38):36269-36276.
154. Cutler RE, Jr., Stephens RM, Saracino MR, Morrison DK: **Autoregulation of the Raf-1 serine/threonine kinase**. *Proc Natl Acad Sci U S A* 1998, **95**(16):9214-9219.
155. Hey F, Pritchard C: **A new mode of RAF autoregulation: a further complication in the inhibitor paradox**. *Cancer cell* 2013, **23**(5):561-563.
156. Tran NH, Wu X, Frost JA: **B-Raf and Raf-1 are regulated by distinct autoregulatory mechanisms**. *The Journal of biological chemistry* 2005, **280**(16):16244-16253.
157. Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, Valtorta E, Schiavo R, Buscarino M, Siravegna G *et al*: **Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer**. *Nature* 2012, **486**(7404):532-536.
158. Hollander D, Donyo M, Atias N, Mekahel K, Melamed Z, Yannai S, Lev-Maor G, Shilo A, Schwartz S, Barshack I *et al*: **A network-based analysis of colon cancer splicing changes reveals a tumorigenesis-favoring regulatory pathway emanating from ELK1**. *Genome research* 2016, **26**(4):541-553.
159. Liu J, Li H, Shen S, Sun L, Yuan Y, Xing C: **Alternative splicing events implicated in carcinogenesis and prognosis of colorectal cancer**. *Journal of Cancer* 2018, **9**(10):1754-1764.
160. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation**. *Cell* 2011, **144**(5):646-674.
161. Oltean S, Bates DO: **Hallmarks of alternative splicing in cancer**. *Oncogene* 2014, **33**(46):5311-5318.
162. Ladomery MR, Harper SJ, Bates DO: **Alternative splicing in angiogenesis: the vascular endothelial growth factor paradigm**. *Cancer letters* 2007, **249**(2):133-142.
163. Boise LH, Gonzalez-Garcia M, Postema CE, Ding L, Lindsten T, Turka LA, Mao X, Nunez G, Thompson CB: **bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death**. *Cell* 1993, **74**(4):597-608.
164. Matos P, Jordan P: **Increased Rac1b expression sustains colorectal tumor cell survival**. *Molecular cancer research : MCR* 2008, **6**(7):1178-1184.
165. Bracco L, Throo E, Cochet O, Einstein R, Maurier F: **Methods and platforms for the quantification of splice variants' expression**. *Progress in molecular and subcellular biology* 2006, **44**:1-25.

166. Lee C, Roy M: **Analysis of alternative splicing with microarrays: successes and challenges**. *Genome biology* 2004, **5**(7):231.
167. Srinivasan K, Shiue L, Hayes JD, Centers R, Fitzwater S, Loewen R, Edmondson LR, Bryant J, Smith M, Rommelfanger C *et al*: **Detection and measurement of alternative splicing using splicing-sensitive microarrays**. *Methods (San Diego, Calif)* 2005, **37**(4):345-359.
168. Harvey SE, Cheng C: **Methods for Characterization of Alternative RNA Splicing**. *Methods in molecular biology (Clifton, NJ)* 2016, **1402**:229-241.
169. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL *et al*: **The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments**. *Clinical chemistry* 2009, **55**(4):611-622.
170. Delire B, Starkel P: **The Ras/MAPK pathway and hepatocarcinoma: pathogenesis and therapeutic implications**. *European journal of clinical investigation* 2015, **45**(6):609-623.
171. Li L, Zhao GD, Shi Z, Qi LL, Zhou LY, Fu ZX: **The Ras/Raf/MEK/ERK signaling pathway and its role in the occurrence and development of HCC**. *Oncology letters* 2016, **12**(5):3045-3050.
172. Newell P, Toffanin S, Villanueva A, Chiang DY, Minguez B, Cabellos L, Savic R, Hoshida Y, Lim KH, Melgar-Lesmes P *et al*: **Ras pathway activation in hepatocellular carcinoma and anti-tumoral effect of combined sorafenib and rapamycin in vivo**. *Journal of hepatology* 2009, **51**(4):725-733.
173. Calvisi DF, Ladu S, Conner EA, Seo D, Hsieh JT, Factor VM, Thorgeirsson SS: **Inactivation of Ras GTPase-activating proteins promotes unrestrained activity of wild-type Ras in human liver cancer**. *Journal of hepatology* 2011, **54**(2):311-319.
174. Zhang X, Li N, Li X, Zhao W, Qiao Y, Liang L, Ding Y: **Low expression of DAB2IP contributes to malignant development and poor prognosis in hepatocellular carcinoma**. *Journal of gastroenterology and hepatology* 2012, **27**(6):1117-1125.
175. Wee P, Wang Z: **Epidermal Growth Factor Receptor Cell Proliferation Signaling Pathways**. *Cancers* 2017, **9**(5).
176. Ramakrishnan P, Baltimore D: **Sam68 is required for both NF-kappaB activation and apoptosis signaling by the TNF receptor**. *Molecular cell* 2011, **43**(2):167-179.
177. Huot ME, Vogel G, Richard S: **Identification of a Sam68 ribonucleoprotein complex regulated by epidermal growth factor**. *The Journal of biological chemistry* 2009, **284**(46):31903-31913.
178. Ramanathan M, Porter DF, Khavari PA: **Methods to study RNA-protein interactions**. *Nature methods* 2019, **16**(3):225-234.
179. Popova VV, Kurshakova MM, Kopytova DV: **[Methods to study the RNA-protein interactions]**. *Molekuliarnaia biologiia* 2015, **49**(3):472-481.
180. Tome JM, Ozer A, Pagano JM, Gheba D, Schroth GP, Lis JT: **Comprehensive analysis of RNA-protein interactions by high-throughput sequencing-RNA affinity profiling**. *Nature methods* 2014, **11**(6):683-688.
181. Lynch KW, Maniatis T: **Assembly of specific SR protein complexes on distinct regulatory elements of the Drosophila doublesex splicing enhancer**. *Genes & development* 1996, **10**(16):2089-2101.
182. **A user's guide to the encyclopedia of DNA elements (ENCODE)**. *PLoS biology* 2011, **9**(4):e1001046.
183. **The ENCODE (ENCyclopedia Of DNA Elements) Project**. *Science (New York, NY)* 2004, **306**(5696):636-640.

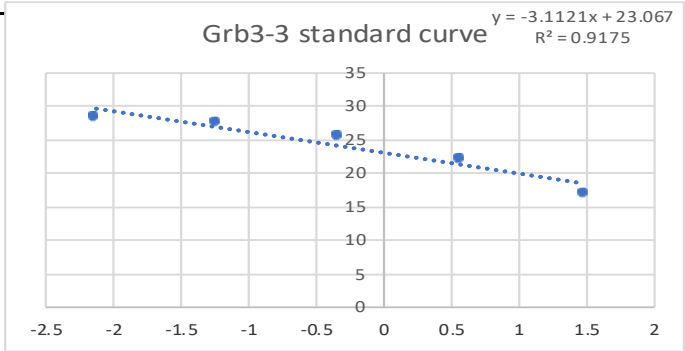
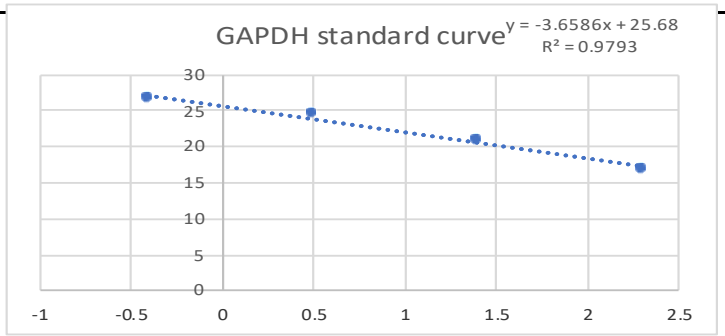
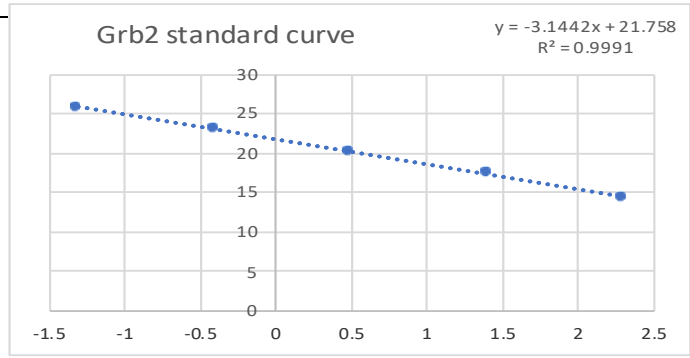
184. Shen S, Park JW, Lu ZX, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y: **rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data**. *Proc Natl Acad Sci U S A* 2014, **111**(51):E5593-5601.
185. Atlas Thp: **https://**http://www.proteinatlas.org/ENSG00000092199-HNRNPC/pathology/colorectal+cancer.
186. Nasrin F, Rahman MA, Masuda A, Ohe K, Takeda J, Ohno K: **HnRNP C, YB-1 and hnRNP L coordinately enhance skipping of human MUSK exon 10 to generate a Wnt-insensitive MuSK isoform**. *Scientific reports* 2014, **4**:6841.
187. Irimura S, Kitamura K, Kato N, Saiki K, Takeuchi A, Gunadi, Matsuo M, Nishio H, Lee MJ: **HnRNP C1/C2 may regulate exon 7 splicing in the spinal muscular atrophy gene SMN1**. *The Kobe journal of medical sciences* 2009, **54**(5):E227-236.

Appendix 1

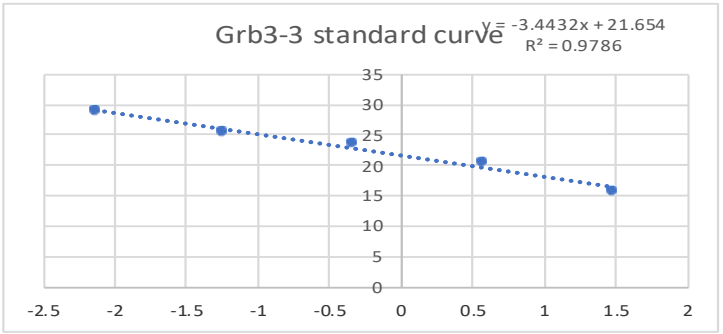
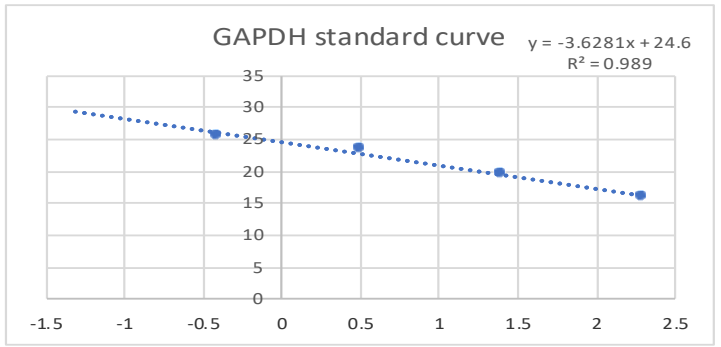
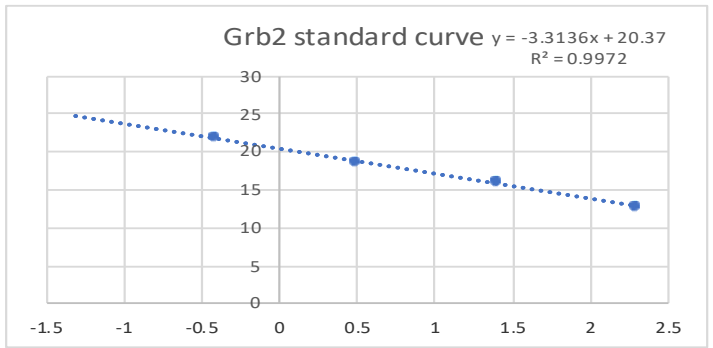
Appendix 1 contains the raw PCR data assessing Grb2 and Grb3-3 expression in colorectal tissue samples, this data corresponds to the processed data shown in figures 3.4, 3.5 and 3.6.

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
1	1	Grb3-3 standard 1	30.00	1.48	15.26	15.29		0.90
1	2				15.32			
1	3	Grb3-3 standard 2	3.75	0.57	20.00	20.09		
1	4				20.18			
1	5	Grb3-3 standard 3	0.47	-0.33	24.17	24.17		
1	6				22.87			
1	7	Grb3-3 standard 4	0.06	-1.23	24.36	27.26		
1	8				27.26			
1	9	Grb3-3 standard 5	0.01	-2.14	27.70	29.88		
1	10				29.88			
1	11	Grb3-3 NTC						
1	12							
1	13	GAPDH standard 1	200.00	2.30	15.29	15.28		0.96
1	14				15.26			
1	15	GAPDH standard 2	25.00	1.40	17.39	17.43		
1	16				17.47			
1	17	GAPDH standard 3	3.13	0.49	21.26	21.30		
1	18				21.34			
1	19	GAPDH standard 4	0.39	-0.41	24.95	24.96		
1	20				24.97			
1	21	GAPDH standard 5	0.05	-1.31	27.06	26.99		
1	22				26.92			
1	23	GAPDH NTC						
1	24							
1	25	Grb2 standard 1	200.00	2.30	12.80	12.84		1.02
1	26				12.88			
1	27	Grb2 standard 2	25.00	1.40	15.93	15.99		
1	28				16.05			
1	29	Grb2 standard 3	3.13	0.49	18.14	18.20		
1	30				18.26			
1	31	Grb2 standard 4	0.39	-0.41	22.05	22.14		
1	32				22.23			
1	33	Grb2 standard 5	0.05	-1.31	24.70	24.59		
1	34				24.47			
1	35	Grb2 NTC						
1	36							

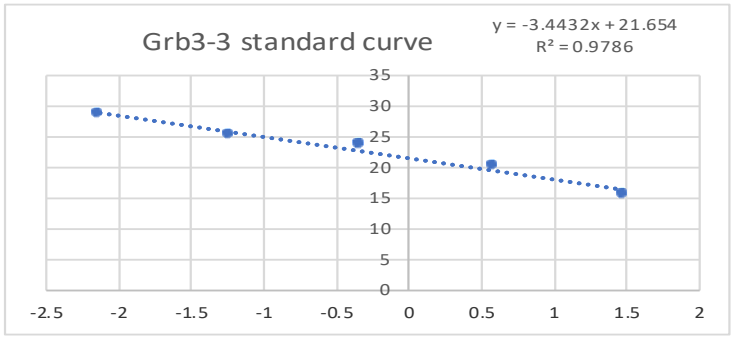
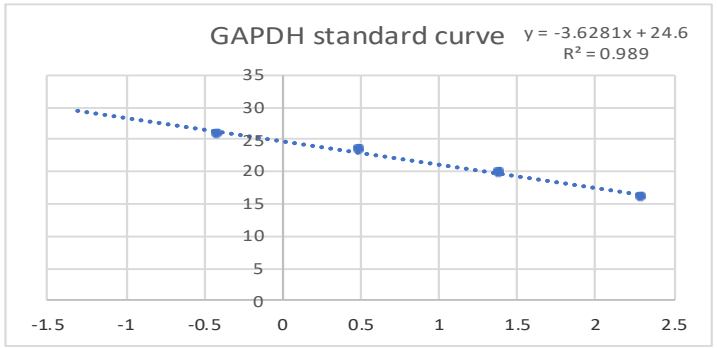
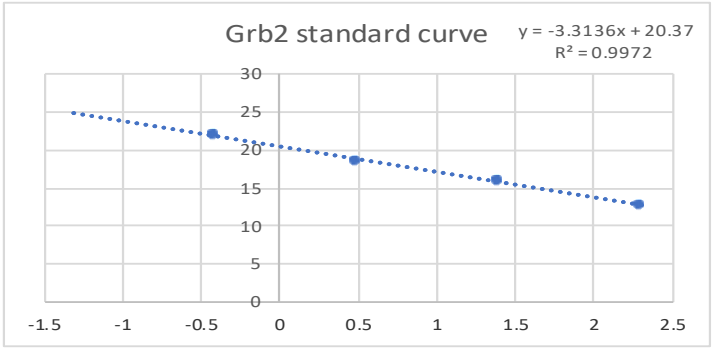
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA (pg)	Normalised amount cDNA (x100) (pg)
1	37	Grb2	92	21.52	21.31	-0.28	0.52	0.04	3.92
1	38			21.10					
1	39	Grb3-3		22.78	22.66	-0.03	0.93	0.07	7.01
1	40			22.54					
1	41	GAPDH		19.02	19.04	1.12	13.25	n/a	n/a
1	42			19.06					
1	43	Grb2	93	22.19	22.32	-0.59	0.26	0.38	38.18
1	44			22.45					
1	45	Grb3-3		26.87	26.52	-1.10	0.08	0.12	11.74
1	46			26.16					
1	47	GAPDH		23.39	23.49	-0.17	0.67	n/a	n/a
1	48			23.58					
1	49	Grb2	106	19.50	19.47	0.28	1.89	0.65	64.56
1	50			19.43					
1	51	Grb3-3		24.26	23.87	-0.37	0.43	0.15	14.58
1	52			23.48					
1	53	GAPDH		20.96	21.29	0.47	2.93	n/a	n/a
1	54			21.61					
1	55	Grb2	107	21.76	22.02	-0.50	0.32	0.26	26.27
1	56			22.28					
1	57	Grb3-3		24.24	24.73	-1.41	0.04	0.03	3.22
1	58			25.21					
1	59	GAPDH		22.62	22.62	0.08	1.20	n/a	n/a
1	60			22.61					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
2	1	Grb3-3 standard 1	30.00	1.48	16.87	16.99		10.00%
2	2				17.10			
2	3	Grb3-3 standard 2	3.75	0.57	21.99	22.11		
2	4				22.22			
2	5	Grb3-3 standard 3	0.47	-0.33	25.74	25.74		
2	6				25.39			
2	7	Grb3-3 standard 4	0.06	-1.23	27.42	27.42		
2	8				27.42			
2	9	Grb3-3 standard 5	0.01	-2.14	28.48	28.28		
2	10				28.28			
2	11	Grb3-3 NTC						
2	12							
2	13	GAPDH standard 1	200.00	2.30	16.82	16.87		88.00%
2	14				16.92			
2	15	GAPDH standard 2	25.00	1.40	20.66	20.76		
2	16				20.86			
2	17	GAPDH standard 3	3.13	0.49	24.88	24.66		
2	18				24.43			
2	19	GAPDH standard 4	0.39	-0.41	26.96	26.59		
2	20				26.21			
2	21	GAPDH standard 5	0.05	-1.31	26.86	26.85		
2	22				26.84			
2	23	GAPDH NTC						
2	24							
2	25	Grb2 standard 1	200.00	2.30	14.35	14.38		108.00%
2	26				14.40			
2	27	Grb2 standard 2	25.00	1.40	17.49	17.58		
2	28				17.66			
2	29	Grb2 standard 3	3.13	0.49	20.55	20.19		
2	30				19.83			
2	31	Grb2 standard 4	0.39	-0.41	22.55	23.02		
2	32				23.49			
2	33	Grb2 standard 5	0.05	-1.31	25.75	25.85		
2	34				25.95			
2	35	Grb2 NTC						
2	36							

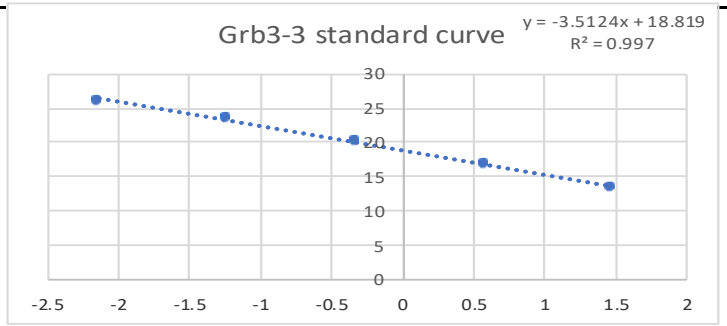
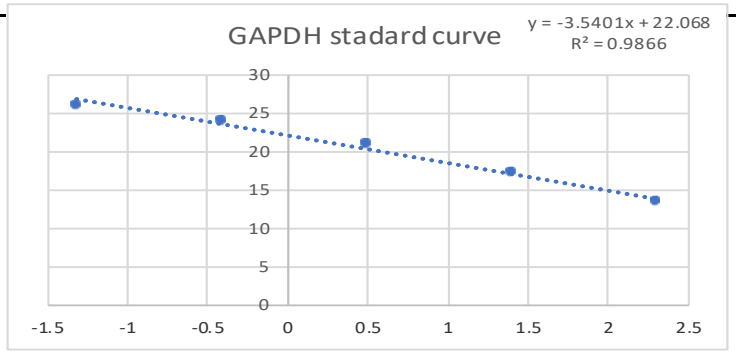
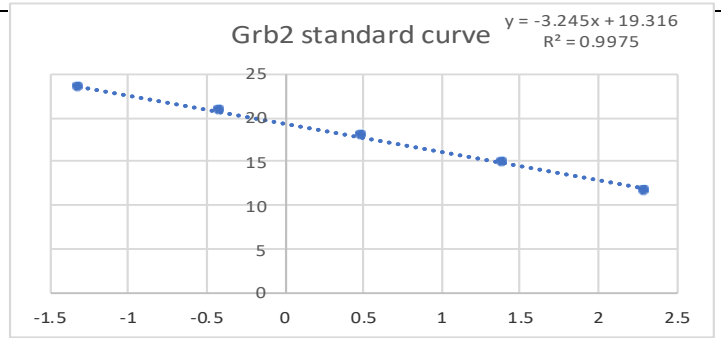
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
2	37	Grb2	85.00	28.47	27.76	-1.91	0.01	0.07	7.12
2	38			27.04					
2	39	Grb3-3		29.73	29.90	-2.19	0.01	0.04	3.68
2	40			30.06					
2	41	GAPDH		28.47	28.46	-0.76	0.17	n/a	n/a
2	42			28.45					
2	43	Grb2	80.00	24.90	24.99	-1.03	0.09	0.06	6.04
2	44			25.08					
2	45	Grb3-3		26.84	27.13	-1.30	0.05	0.03	3.20
2	46			27.41					
2	47	GAPDH		24.69	24.98	0.19	1.55	n/a	n/a
2	48			25.27					
2	49	Grb2	124.00	24.53	24.56	-0.89	0.13	0.08	7.87
2	50			24.58					
2	51	Grb3-3		22.69	22.50	0.18	1.53	0.93	93.16
2	52			22.30					
2	53	GAPDH		24.73	24.90	0.21	1.64	n/a	n/a
2	54			25.06					
2	55	Grb2	125.00	23.83	23.86	-0.67	0.22	0.04	4.17
2	56			23.88					
2	57	Grb3-3		26.03	25.98	-0.93	0.12	0.02	2.25
2	58			25.92					
2	59	GAPDH		23.03	23.07	0.71	5.17	n/a	n/a
2	60			23.11					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
3	1	Grb3-3 standard 1	30.00	1.48	15.86	15.77		95.00%
3	2				15.68			
3	3	Grb3-3 standard 2	3.75	0.57	20.49	20.30		
3	4				20.10			
3	5	Grb3-3 standard 3	0.47	-0.33	23.83	23.83		
3	6				23.55			
3	7	Grb3-3 standard 4	0.06	-1.23	25.77	25.09		
3	8				25.09			
3	9	Grb3-3 standard 5	0.01	-2.14	28.68	28.82		
3	10				28.82			
3	11	Grb3-3 NTC						
3	12							
3	13	GAPDH standard 1	200.00	2.30	15.86	16.00		89.00%
3	14				16.13			
3	15	GAPDH standard 2	25.00	1.40	19.83	19.62		
3	16				19.41			
3	17	GAPDH standard 3	3.13	0.49	23.39	21.65		
3	18				19.91			
3	19	GAPDH standard 4	0.39	-0.41	25.43	25.66		
3	20				25.89			
3	21	GAPDH standard 5	0.05	-1.31	27.21	27.16		
3	22				27.11			
3	23	GAPDH NTC						
3	24							
3	25	Grb2 standard 1	200.00	2.30	12.67	12.69		100.00%
3	26				12.70			
3	27	Grb2 standard 2	25.00	1.40	15.71	15.96		
3	28				16.20			
3	29	Grb2 standard 3	3.13	0.49	18.75	18.48		
3	30				18.20			
3	31	Grb2 standard 4	0.39	-0.41	21.75	21.82		
3	32				21.89			
3	33	Grb2 standard 5	0.05	-1.31	21.62	22.44		
3	34				23.26			
3	35	Grb2 NTC						
3	36							

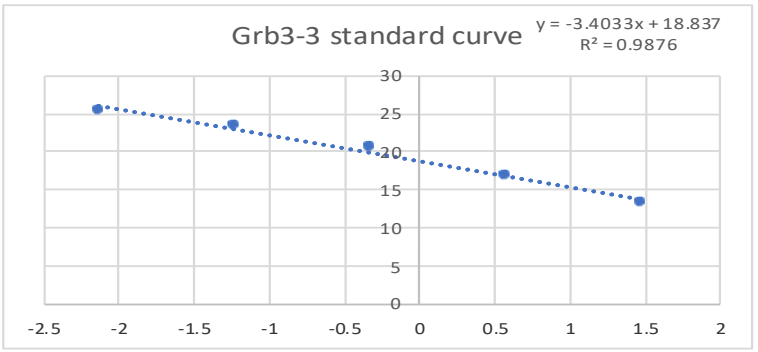
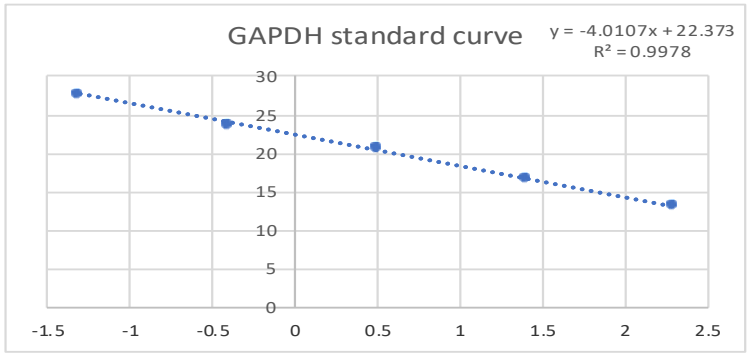
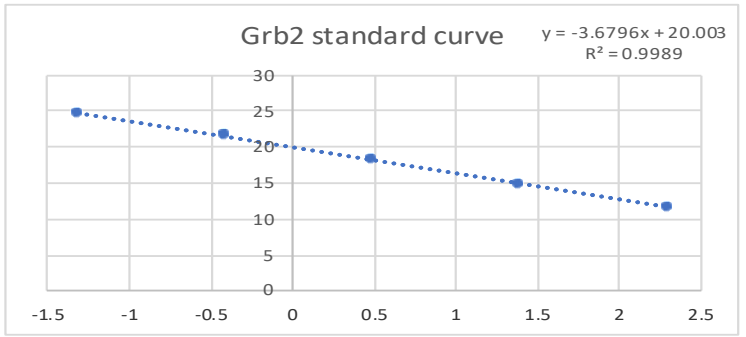
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
3	37	Grb2	126.00	20.61	20.98	-2.34	0.00	0.00	0.42
3	38			21.34					
3	39	Grb3-3		22.33	22.85	-0.35	0.45	0.41	41.18
3	40			23.37					
3	41	GAPDH		24.04	24.47	0.04	1.09	n/a	n/a
3	42			24.90					
3	43	Grb2	127.00	25.81	25.70	-1.61	0.02	0.11	11.30
3	44			25.58					
3	45	Grb3-3		29.34	28.86	-2.09	0.01	0.04	3.70
3	46			28.37					
3	47	GAPDH		26.84	27.00	-0.66	0.22	n/a	n/a
3	48			27.15					
3	49	Grb2	67.00	19.30	19.79	0.18	1.50	0.09	8.97
3	50			20.28					
3	51	Grb3-3		22.26	21.65	0.00	1.01	0.06	6.03
3	52			21.03					
3	53	GAPDH		20.19	20.17	1.22	16.69	n/a	n/a
3	54			20.14					
3	55	Grb2	66.00	23.17	23.52	-0.17	0.68	1.16	115.59
3	56			23.86					
3	57	Grb3-3		25.55	26.10	-1.29	0.05	0.09	8.72
3	58			26.65					
3	59	GAPDH		25.36	25.44	-0.23	0.59	n/a	n/a
3	60			25.52					
3	61	Grb2	64.00	25.98	26.24	-1.77	0.02	0.09	8.67
3	62			26.50					
3	63	Grb3-3		28.84	28.89	-2.10	0.01	0.04	4.06
3	64			28.94					
3	65	GAPDH		27.17	27.18	-0.71	0.20	n/a	n/a
3	66			27.18					
3	67	Grb2	65.00	24.79	24.84	-1.35	0.04	0.07	7.28
3	68			24.88					
3	69	Grb3-3		27.08	27.16	-1.60	0.03	0.04	4.08
3	70			27.24					
3	71	GAPDH		26.16	25.36	-0.21	0.62	n/a	n/a
3	72			24.56					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
4	1	Grb3-3 standard 1	30.00	1.48	15.86	15.77		95.00%
4	2				15.68			
4	3	Grb3-3 standard 2	3.75	0.57	20.49	20.30		
4	4				20.10			
4	5	Grb3-3 standard 3	0.47	-0.33	23.83	23.83		
4	6				23.55			
4	7	Grb3-3 standard 4	0.06	-1.23	25.77	25.09		
4	8				25.09			
4	9	Grb3-3 standard 5	0.01	-2.14	28.68	28.82		
4	10				28.82			
4	11	Grb3-3 NTC						
4	12							
4	13	GAPDH standard 1	200.00	2.30	15.86	16.00		89.00%
4	14				16.13			
4	15	GAPDH standard 2	25.00	1.40	19.83	19.62		
4	16				19.41			
4	17	GAPDH standard 3	3.13	0.49	23.39	21.65		
4	18				19.91			
4	19	GAPDH standard 4	0.39	-0.41	25.43	25.66		
4	20				25.89			
4	21	GAPDH standard 5	0.05	-1.31	27.21	27.16		
4	22				27.11			
4	23	GAPDH NTC						
4	24							
4	25	Grb2 standard 1	200.00	2.30	12.67	12.69		100.00%
4	26				12.70			
4	27	Grb2 standard 2	25.00	1.40	15.71	15.96		
4	28				16.20			
4	29	Grb2 standard 3	3.13	0.49	18.75	18.48		
4	30				18.20			
4	31	Grb2 standard 4	0.39	-0.41	21.75	21.82		
4	32				21.89			
4	33	Grb2 standard 5	0.05	-1.31	21.62	22.44		
4	34				23.26			
4	35	Grb2 NTC						
4	36							

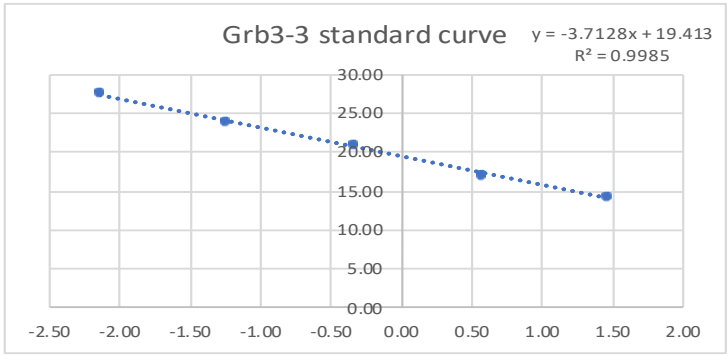
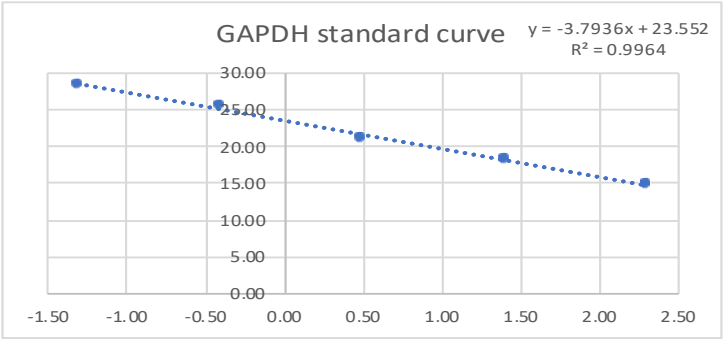
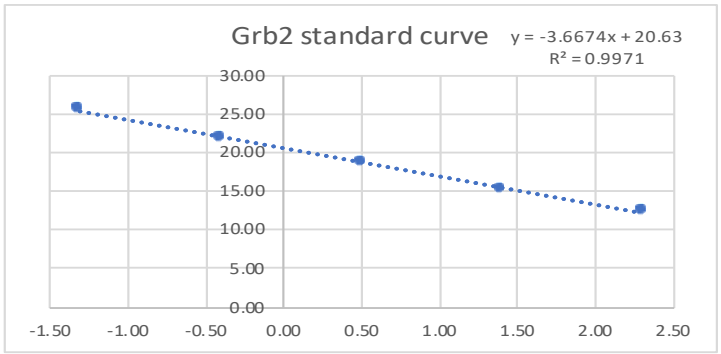
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
4	37	Grb2	126.00	20.61	20.98	-2.34	0.00	0.00	0.42
4	38			21.34					
4	39	Grb3-3		22.33	22.85	-0.35	0.45	0.41	41.18
4	40			23.37					
4	41	GAPDH		24.04	24.47	0.04	1.09	n/a	n/a
4	42			24.90					
4	43	Grb2	127.00	25.81	25.70	-1.61	0.02	0.11	11.30
4	44			25.58					
4	45	Grb3-3		29.34	28.86	-2.09	0.01	0.04	3.70
4	46			28.37					
4	47	GAPDH		26.84	27.00	-0.66	0.22	n/a	n/a
4	48			27.15					
4	49	Grb2	67.00	19.30	19.79	0.18	1.50	0.09	8.97
4	50			20.28					
4	51	Grb3-3		22.26	21.65	0.00	1.01	0.06	6.03
4	52			21.03					
4	53	GAPDH		20.19	20.17	1.22	16.69	n/a	n/a
4	54			20.14					
4	55	Grb2	66.00	23.17	23.52	-0.17	0.68	1.16	115.59
4	56			23.86					
4	57	Grb3-3		25.55	26.10	-1.29	0.05	0.09	8.72
4	58			26.65					
4	59	GAPDH		25.36	25.44	-0.23	0.59	n/a	n/a
4	60			25.52					
4	61	Grb2	64.00	25.98	26.24	-1.77	0.02	0.09	8.67
4	62			26.50					
4	63	Grb3-3		28.84	28.89	-2.10	0.01	0.04	4.06
4	64			28.94					
4	65	GAPDH		27.17	27.18	-0.71	0.20	n/a	n/a
4	66			27.18					
4	67	Grb2	65.00	24.79	24.84	-1.35	0.04	0.07	7.28
4	68			24.88					
4	69	Grb3-3		27.08	27.16	-1.60	0.03	0.04	4.08
4	70			27.24					
4	71	GAPDH		26.16	25.36	-0.21	0.62	n/a	n/a
4	72			24.56					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
5	1	Grb3-3 standard 1	30.00	1.48	13.39	13.47		92.60%
5	2				13.54			
5	3	Grb3-3 standard 2	3.75	0.57	16.72	16.81		
5	4				16.89			
5	5	Grb3-3 standard 3	0.47	-0.33	20.07	20.07		
5	6				20.17			
5	7	Grb3-3 standard 4	0.06	-1.23	23.28	23.75		
5	8				23.75			
5	9	Grb3-3 standard 5	0.01	-2.14	25.98	25.96		
5	10				25.96			
5	11	Grb3-3 NTC						
5	12							
5	13	GAPDH standard 1	200.00	2.30	13.40	13.45		91.5%%
5	14				13.50			
5	15	GAPDH standard 2	25.00	1.40	17.06	17.16		
5	16				17.26			
5	17	GAPDH standard 3	3.13	0.49	20.81	20.94		
5	18				21.07			
5	19	GAPDH standard 4	0.39	-0.41	24.01	24.03		
5	20				24.05			
5	21	GAPDH standard 5	0.05	-1.31	25.79	26.00		
5	22				26.21			
5	23	GAPDH NTC						
5	24							
5	25	Grb2 standard 1	200.00	2.30	11.62	11.66		103%%
5	26				11.70			
5	27	Grb2 standard 2	25.00	1.40	14.72	14.80		
5	28				14.88			
5	29	Grb2 standard 3	3.13	0.49	17.98	17.96		
5	30				17.93			
5	31	Grb2 standard 4	0.39	-0.41	21.30	20.85		
5	32				20.39			
5	33	Grb2 standard 5	0.05	-1.31	23.22	23.29		
5	34				23.36			
5	35	Grb2 NTC						
5	36							

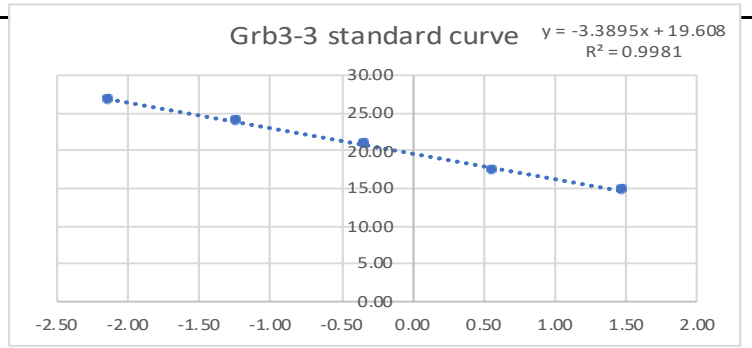
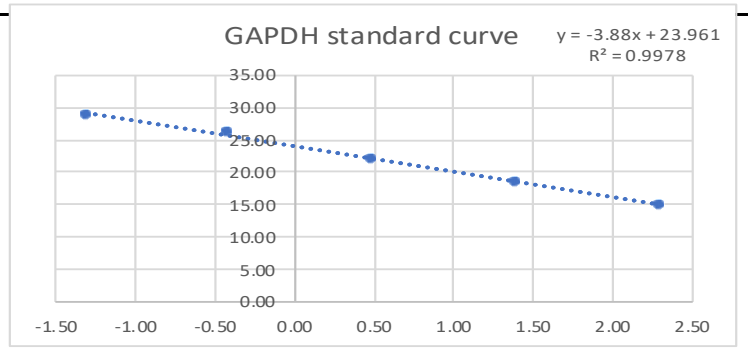
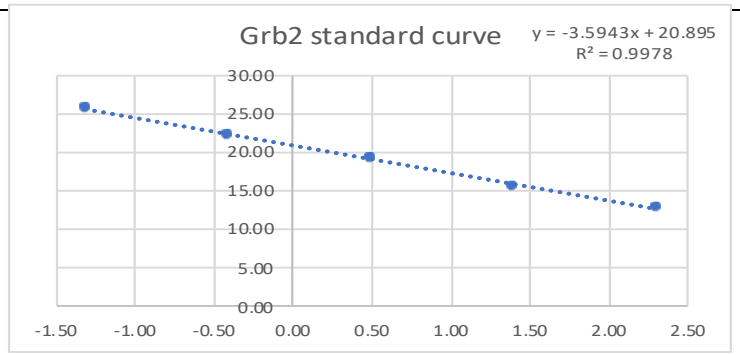
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
5	37	Grb2	114.00	18.79	19.28	0.01	1.03	0.20	19.56
5	38			19.76					
5	39	Grb3-3		20.69	20.83	-0.57	0.27	0.05	5.08
5	40			20.97					
5	41	GAPDH		19.62	19.52	0.72	5.26	n/a	n/a
5	42			19.41					
5	43	Grb2	117.00	18.31	17.83	0.46	2.88	0.70	69.86
5	44			17.34					
5	45	Grb3-3		18.46	18.34	0.14	1.37	0.33	33.20
5	46			18.22					
5	47	GAPDH		20.07	19.89	0.62	4.12	n/a	n/a
5	48			19.71					
5	49	Grb2	110.00	14.87	15.18	1.28	18.89	0.15	15.50
5	50			15.48					
5	51	Grb3-3		16.20	16.18	0.75	5.66	0.05	4.64
5	52			16.15					
5	53	GAPDH		14.66	14.69	2.09	121.84	n/a	n/a
5	54			14.71					
5	55	Grb2	112.00	19.81	19.64	-0.10	0.80	0.42	41.80
5	56			19.46					
5	57	Grb3-3		21.52	21.16	-0.67	0.22	0.11	11.34
5	58			20.79					
5	59	GAPDH		21.10	21.08	0.28	1.91	n/a	n/a
5	60			21.05					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
6	1	Grb3-3 standard 1	30.00	1.48	13.36	13.37		96.70%
6	2				13.37			
6	3	Grb3-3 standard 2	3.75	0.57	16.82	16.89		
6	4				16.95			
6	5	Grb3-3 standard 3	0.47	-0.33	20.61	20.61		
6	6				20.77			
6	7	Grb3-3 standard 4	0.06	-1.23	23.50	23.18		
6	8				23.18			
6	9	Grb3-3 standard 5	0.01	-2.14	25.49	25.52		
6	10				25.52			
6	11	Grb3-3 NTC						
6	12							
6	13	GAPDH standard 1	200.00	2.30	13.03	13.06		77.50%
6	14				13.09			
6	15	GAPDH standard 2	25.00	1.40	16.73	16.75		
6	16				16.77			
6	17	GAPDH standard 3	3.13	0.49	20.68	20.77		
6	18				20.85			
6	19	GAPDH standard 4	0.39	-0.41	23.57	23.64		
6	20				23.71			
6	21	GAPDH standard 5	0.05	-1.31	27.69	27.73		
6	22				27.76			
6	23	GAPDH NTC						
6	24							
6	25	Grb2 standard 1	200.00	2.30	11.53	11.56		87.00%
6	26				11.58			
6	27	Grb2 standard 2	25.00	1.40	14.58	14.68		
6	28				14.78			
6	29	Grb2 standard 3	3.13	0.49	18.30	18.31		
6	30				18.31			
6	31	Grb2 standard 4	0.39	-0.41	21.62	21.72		
6	32				21.82			
6	33	Grb2 standard 5	0.05	-1.31	23.99	24.65		
6	34				25.31			
6	35	Grb2 NTC						
6	36							

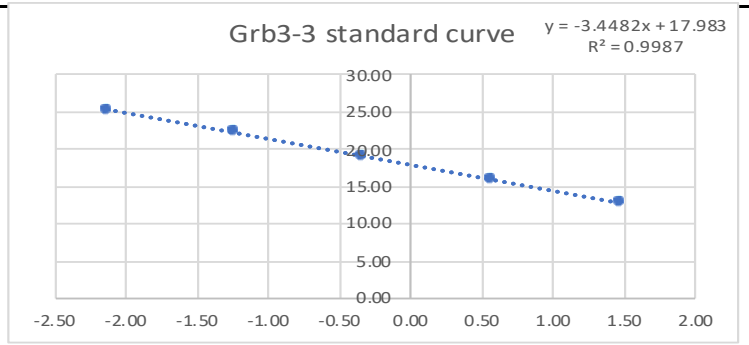
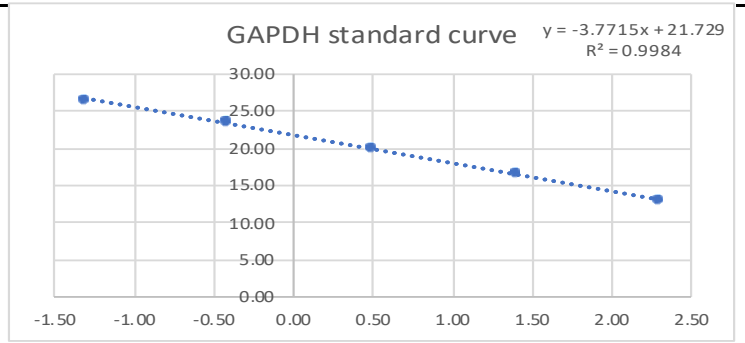
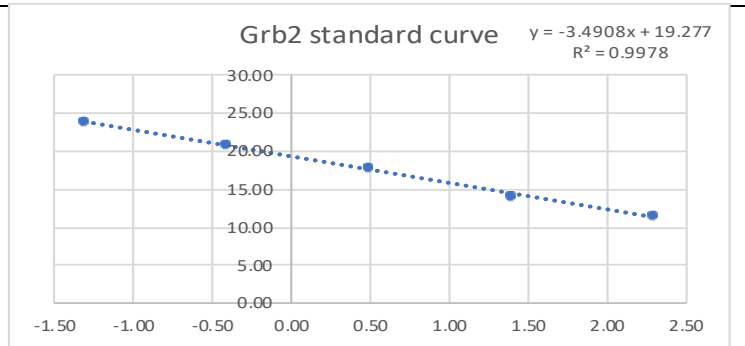
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
6	37	Grb2	128.00	23.25	23.11	-0.84	0.14	0.11	11.16
6	38			22.97					
6	39	Grb3-3		22.96	22.96	-1.21	0.06	0.05	4.81
6	40			22.95					
6	41	GAPDH		21.63	21.94	0.11	1.28	n/a	n/a
6	42			22.25					
6	43	Grb2	94.00	24.08	24.42	-1.20	0.06	0.11	10.77
6	44			24.75					
6	45	Grb3-3		21.96	23.00	-1.22	0.06	0.10	10.18
6	46			24.04					
6	47	GAPDH		23.25	23.30	-0.23	0.59	n/a	n/a
6	48			23.35					
6	49	Grb2	70.00	26.41	26.37	-1.73	0.02	0.18	18.46
6	50			26.32					
6	51	Grb3-3		25.32	25.88	-2.07	0.01	0.08	8.46
6	52			26.43					
6	53	GAPDH		26.33	26.37	-1.00	0.10	n/a	n/a
6	54			26.40					
6	55	Grb2	73.00	23.07	23.18	-0.86	0.14	0.08	8.35
6	56			23.28					
6	57	Grb3-3		23.09	23.17	-1.27	0.05	0.03	3.25
6	58			23.24					
6	59	GAPDH		21.52	21.51	0.22	1.65	n/a	n/a
6	60			21.49					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
7	1	Grb3-3 standard 1	30.00	1.48	14.13	14.12		85.00%
7	2				14.10			
7	3	Grb3-3 standard 2	3.75	0.57	17.05	16.98		
7	4				16.90			
7	5	Grb3-3 standard 3	0.47	-0.33	20.78	20.78		
7	6				20.68			
7	7	Grb3-3 standard 4	0.06	-1.23	23.64	23.88		
7	8				24.11			
7	9	Grb3-3 standard 5	0.01	-2.14	27.48	27.43		
7	10				27.43			
7	11	Grb3-3 NTC						
7	12							
7	13	GAPDH standard 1	200.00	2.30	14.95	14.98		85.00%
7	14				15.00			
7	15	GAPDH standard 2	25.00	1.40	18.05	18.25		
7	16				18.45			
7	17	GAPDH standard 3	3.13	0.49	21.18	21.18		
7	18				21.17			
7	19	GAPDH standard 4	0.39	-0.41	25.27	25.49		
7	20				25.71			
7	21	GAPDH standard 5	0.05	-1.31	28.62	28.49		
7	22				28.35			
7	23	GAPDH NTC						
7	24							
7	25	Grb2 standard 1	200.00	2.30	12.45	12.50		87.00%
7	26				12.54			
7	27	Grb2 standard 2	25.00	1.40	15.15	15.24		
7	28				15.32			
7	29	Grb2 standard 3	3.13	0.49	18.73	18.74		
7	30				18.74			
7	31	Grb2 standard 4	0.39	-0.41	21.82	21.88		
7	32				21.93			
7	33	Grb2 standard 5	0.05	-1.31	25.60	25.74		
7	34				25.87			
7	35	Grb2 NTC						
7	36							

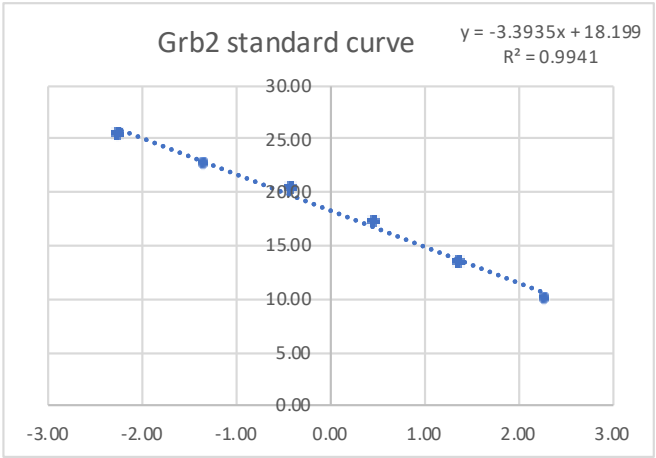
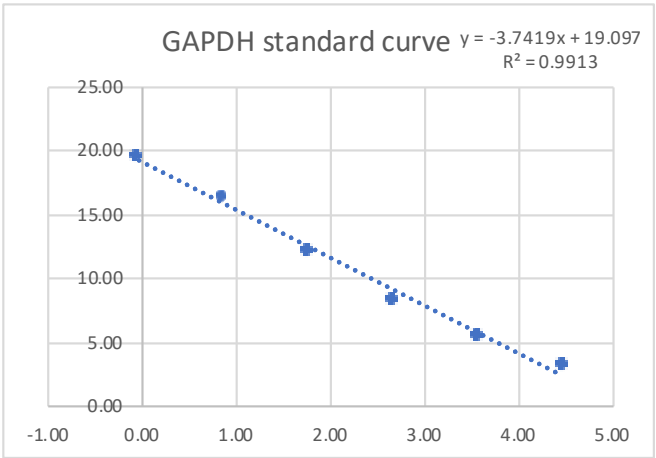
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
7	37	Grb2	115.00	25.43	25.48	-1.32	0.05	0.43	43.18
7	38			25.52					
7	39	Grb3-3		23.19	23.34	-1.04	0.09	0.82	82.50
7	40			23.49					
7	41	GAPDH		27.46	27.18	-0.96	0.11	n/a	n/a
7	42			26.90					
7	43	Grb2	121.00	26.26	26.17	-1.51	0.03	0.29	28.86
7	44			26.08					
7	45	Grb3-3		24.59	25.37	-1.58	0.03	0.24	24.43
7	46			26.15					
7	47	GAPDH		27.03	27.24	-0.97	0.11	n/a	n/a
7	48			27.44					
7	49	Grb2	87.00	26.09	26.09	-1.49	0.03	0.20	19.90
7	50			26.08					
7	51	Grb3-3		28.25	28.44	-2.40	0.00	0.02	2.42
7	52			28.62					
7	53	GAPDH		26.63	26.54	-0.79	0.16	n/a	n/a
7	54			26.44					
7	55	Grb2	95.00	25.19	25.31	-1.27	0.05	0.22	21.96
7	56			25.42					
7	57	Grb3-3		27.27	27.67	-2.20	0.01	0.03	2.62
7	58			28.07					
7	59	GAPDH		26.35	25.89	-0.62	0.24	n/a	n/a
7	60			25.43					
7	61	Grb2	68.00	23.59	25.01	-1.19	0.06	0.78	77.85
7	62			26.42					
7	63	Grb3-3		28.95	29.06	-2.57	0.00	0.03	3.27
7	64			29.17					
7	65	GAPDH		27.64	27.67	-1.08	0.08	n/a	n/a
7	66			27.69					
7	67	Grb2	69.00	27.74	27.36	-1.83	0.01	0.22	21.82
7	68			26.97					
7	69	Grb3-3		28.95	29.45	-2.67	0.00	0.03	3.16
7	70			29.94					
7	71	GAPDH		28.58	28.00	-1.17	0.07	n/a	n/a
7	72			27.42					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
8	1	Grb3-3 standard 1	30.00	1.48	14.63	14.69		97.00%
8	2				14.74			
8	3	Grb3-3 standard 2	3.75	0.57	17.25	17.37		
8	4				17.49			
8	5	Grb3-3 standard 3	0.47	-0.33	20.90	20.90		
8	6				20.75			
8	7	Grb3-3 standard 4	0.06	-1.23	23.90	23.97		
8	8				23.97			
8	9	Grb3-3 standard 5	0.01	-2.14	26.92	26.69		
8	10				26.69			
8	11	Grb3-3 NTC			30.10			
8	12				30.36			
8	13	GAPDH standard 1	200.00	2.30	14.89	14.98		81.00%
8	14				15.06			
8	15	GAPDH standard 2	25.00	1.40	18.46	18.48		
8	16				18.50			
8	17	GAPDH standard 3	3.13	0.49	21.90	22.04		
8	18				22.17			
8	19	GAPDH standard 4	0.39	-0.41	25.89	25.96		
8	20				26.03			
8	21	GAPDH standard 5	0.05	-1.31	28.64	28.76		
8	22				28.87			
8	23	GAPDH NTC						
8	24							
8	25	Grb2 standard 1	200.00	2.30	12.92	12.90		90.00%
8	26				12.87			
8	27	Grb2 standard 2	25.00	1.40	15.51	15.56		
8	28				15.60			
8	29	Grb2 standard 3	3.13	0.49	19.05	19.13		
8	30				19.20			
8	31	Grb2 standard 4	0.39	-0.41	22.10	22.21		
8	32				22.31			
8	33	Grb2 standard 5	0.05	-1.31	25.84	25.80		
8	34				25.76			
8	35	Grb2 NTC			30.62			
8	36				30.58			

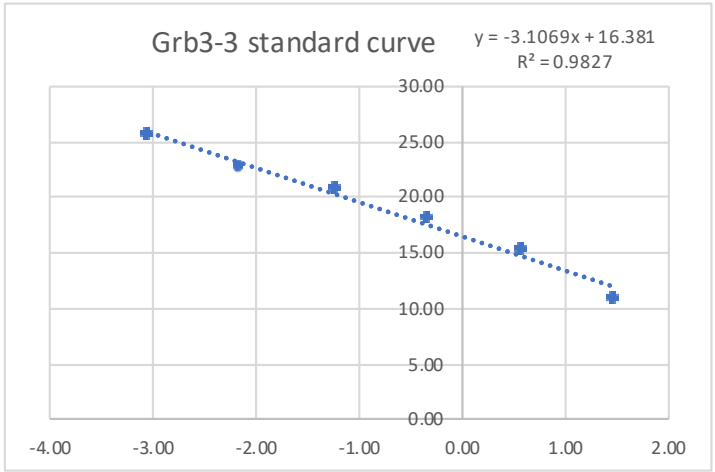
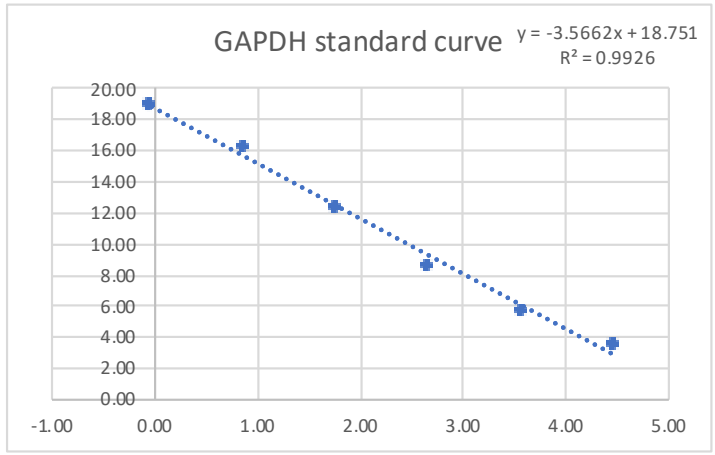
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
8	37	Grb2	99.00	25.47	25.71	-1.34	0.05	0.13	13.07
8	38			25.95					
8	39	Grb3-3		25.68	28.11	-2.51	0.00	0.01	0.89
8	40			30.53					
8	41	GAPDH		24.36	25.73	-0.46	0.35	n/a	n/a
8	42			27.10					
8	43	Grb2	108.00	29.17	29.32	-2.34	0.00	0.09	8.80
8	44			29.46					
8	45	Grb3-3		29.46	29.08	-2.79	0.00	0.03	3.11
8	46			28.70					
8	47	GAPDH		29.23	28.96	-1.29	0.05	n/a	n/a
8	48			28.68					
8	49	Grb2	76.00	30.04	29.84	-2.49	0.00	0.11	11.45
8	50			29.64					
8	51	Grb3-3		31.14	30.81	-3.30	0.00	0.02	1.75
8	52			30.48					
8	53	GAPDH		29.88	29.97	-1.55	0.03	n/a	n/a
8	54			30.05					
8	55	Grb2	77.00	25.66	26.14	-1.46	0.03	0.14	14.17
8	56			26.61					
8	57	Grb3-3		29.69	28.33	-2.57	0.00	0.01	1.09
8	58			26.97					
8	59	GAPDH		26.17	26.33	-0.61	0.25	n/a	n/a
8	60			26.48					
8	61	Grb2	82.00	26.32	26.27	-1.50	0.03	0.14	14.46
8	62			26.22					
8	63	Grb3-3		28.02	28.14	-2.52	0.00	0.01	1.38
8	64			28.26					
8	65	GAPDH		26.70	26.51	-0.66	0.22	n/a	n/a
8	66			26.31					
8	67	Grb2	83.00	24.92	25.79	-1.36	0.04	0.22	21.76
8	68			26.65					
8	69	Grb3-3		28.45	28.53	-2.63	0.00	0.01	1.17
8	70			28.60					
8	71	GAPDH		26.56	26.67	-0.70	0.20	n/a	n/a
8	72			26.78					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
9	1	Grb3-3 standard 1	30.00	1.48	12.81	12.94		95.00%
9	2				13.07			
9	3	Grb3-3 standard 2	3.75	0.57	15.84	15.88		
9	4				15.92			
9	5	Grb3-3 standard 3	0.47	-0.33	19.06	19.06		
9	6				19.26			
9	7	Grb3-3 standard 4	0.06	-1.23	22.10	22.52		
9	8				22.52			
9	9	Grb3-3 standard 5	0.01	-2.14	25.55	25.19		
9	10				25.19			
9	11	Grb3-3 NTC						
9	12							
9	13	GAPDH standard 1	200.00	2.30	12.92	12.95		84.00%
9	14				12.97			
9	15	GAPDH standard 2	25.00	1.40	16.41	16.47		
9	16				16.52			
9	17	GAPDH standard 3	3.13	0.49	19.86	19.91		
9	18				19.95			
9	19	GAPDH standard 4	0.39	-0.41	23.47	23.59		
9	20				23.70			
9	21	GAPDH standard 5	0.05	-1.31	26.28	26.42		
9	22				26.55			
9	23	GAPDH NTC						
9	24							
9	25	Grb2 standard 1	200.00	2.30	11.37	11.43		93.00%
9	26				11.49			
9	27	Grb2 standard 2	25.00	1.40	13.98	14.03		
9	28				14.07			
9	29	Grb2 standard 3	3.13	0.49	17.71	17.76		
9	30				17.81			
9	31	Grb2 standard 4	0.39	-0.41	20.51	20.66		
9	32				20.81			
9	33	Grb2 standard 5	0.05	-1.31	23.70	23.88		
9	34				24.05			
9	35	Grb2 NTC						
9	36							

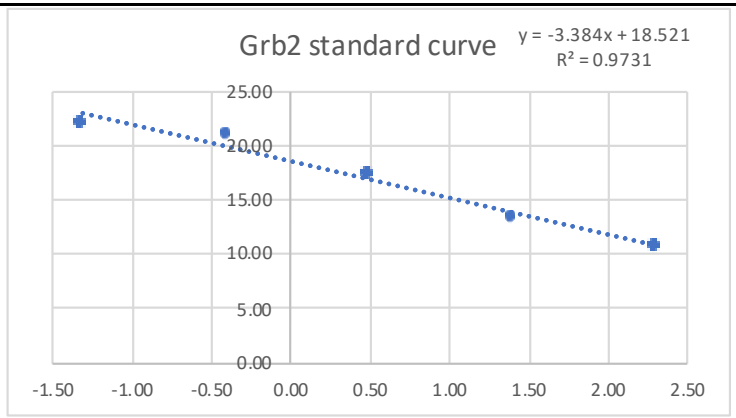
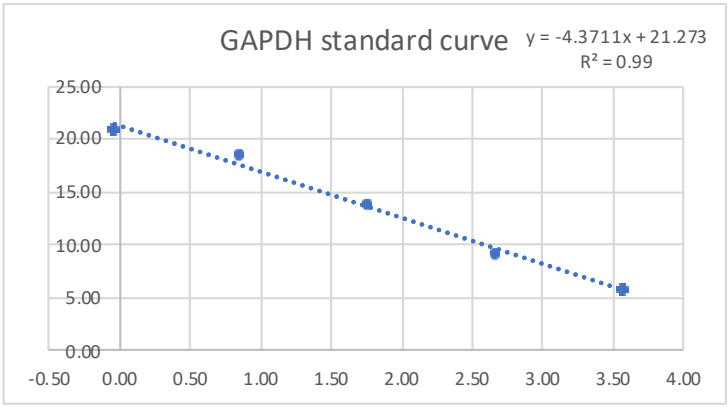
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
9	37	Grb2	96.00	21.12	20.88	-0.46	0.35	0.62	61.51
9	38			20.64					
9	39	Grb3-3		23.26	23.46	-1.59	0.03	0.05	4.58
9	40			23.65					
9	41	GAPDH		22.89	22.67	-0.25	0.56	n/a	n/a
9	42			22.44					
9	43	Grb2	97.00	17.92	17.58	0.49	3.06	0.73	72.55
9	44			17.24					
9	45	Grb3-3		19.92	20.13	-0.62	0.24	0.06	5.67
9	46			20.33					
9	47	GAPDH		19.44	19.37	0.63	4.22	n/a	n/a
9	48			19.30					
9	49	Grb2	89.00	17.78	18.27	0.29	1.94	0.62	62.45
9	50			18.76					
9	51	Grb3-3		19.90	20.46	-0.72	0.19	0.06	6.15
9	52			21.02					
9	53	GAPDH		19.97	19.87	0.49	3.11	n/a	n/a
9	54			19.77					
9	55	Grb2	98.00	18.55	18.33	0.27	1.87	0.32	32.41
9	56			18.10					
9	57	Grb3-3		20.01	20.01	-0.59	0.26	0.04	4.48
9	58			20.00					
9	59	GAPDH		18.97	18.86	0.76	5.78	n/a	n/a
9	60			18.74					
9	61	Grb2	86.00	18.21	18.51	0.22	1.66	0.32	32.02
9	62			18.81					
9	63	Grb3-3		19.21	19.12	-0.33	0.47	0.09	9.04
9	64			19.03					
9	65	GAPDH		19.06	19.04	0.71	5.18	n/a	n/a
9	66			19.01					
9	67	Grb2	88.00	17.38	17.58	0.49	3.07	0.62	62.11
9	68			17.77					
9	69	Grb3-3		19.11	19.13	-0.33	0.47	0.09	9.43
9	70			19.14					
9	71	GAPDH		18.83	19.11	0.69	4.95	n/a	n/a
9	72			19.39					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
10	1	Grb2 Standards 1	200.00	2.30	10.12	10.00		97.00%
10	2				9.88			
10	3	Grb2 Standards 2	25.00	1.40	12.96	13.31		
10	4				13.66			
10	5	Grb2 Standards 3	3.13	0.49	16.56	17.01		
10	6				17.45			
10	7	Grb2 Standards 4	0.39	-0.41	19.93	20.21		
10	8				20.49			
10	9	Grb2 Standards 5	0.05	-1.32	22.47	22.50		
10	10				22.53			
10	11	Grb2 Standards 6	0.01	-2.22	23.47	25.34		
10	12				25.34			
10	13	Grb2 no template control	0.00	n/a				
10	14							
10	15	GADPH sandards 1	30000.00	4.48	3.10	3.12		85.00%
10	16				3.14			
10	17	GAPDH Standards 2	3750.00	3.57	5.37	5.42		
10	18				5.46			
10	19	GAPDH Standards 3	468.75	2.67	8.28	8.31		
10	20				8.34			
10	21	GAPDH Standards 4	58.59	1.77	12.07	12.10		
10	22				12.13			
10	23	GAPDH Standards 5	7.32	0.86	16.28	16.36		
10	24				16.44			
10	25	GAPDH Standards 6	0.92	-0.04	19.40	19.45		
10	26				19.50			
10	27	Grb2 no template control	0.00	n/a				
10	28							

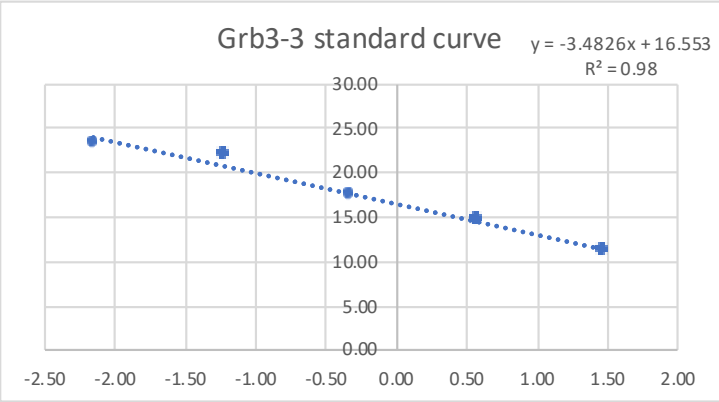
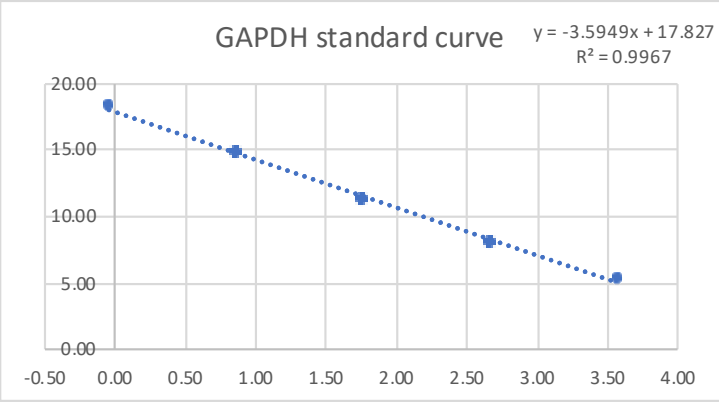
PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
10	29	Grb2	1-T	13.77	13.73	1.32	20.82	0.60	60.05
10	30			13.68					
10	31	GAPDH		13.29	13.34	1.54	34.66	n/a	n/a
10	32			13.38					
10	33	Grb2	2-T	13.02	13.69	1.33	21.32	2.30	230.19
10	34			14.36					
10	35	GAPDH		15.43	15.48	0.97	9.26	n/a	n/a
10	36			15.53					
10	37	Grb2	3-T	14.46	14.27	1.16	14.43	0.48	48.26
10	38			14.07					
10	39	GAPDH		13.76	13.58	1.48	29.90	n/a	n/a
10	40			13.39					
10	41	Grb2	4-T	16.77	16.78	0.42	2.63	0.97	96.56
10	42			16.78					
10	43	GAPDH		17.83	17.47	0.43	2.72	n/a	n/a
10	44			17.11					
10	45	Grb2	5-T	13.44	13.46	1.40	24.92	0.37	36.87
10	46			13.48					
10	47	GAPDH		12.15	12.25	1.83	67.58	n/a	n/a
10	48			12.35					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
11	1	Grb3-3 standard 1	30.00	1.48	10.73	10.75		110.00%
11	2				10.76			
11	3	Grb3-3 standard 2	3.75	0.57	15.32	15.28		
11	4				15.23			
11	5	Grb3-3 standard 3	0.46	-0.34	17.98	18.11		
11	6				18.23			
11	7	Grb3-3 standard 4	0.06	-1.23	20.27	20.62		
11	8				20.96			
11	9	Grb3-3 standard 5	0.01	-2.15	22.42	22.74		
11	10				23.06			
11	11	Grb3-3 standard 6	0.00	-3.05	25.32	25.46		
11	12				25.46			
11	13	Grb3-3 no template control	0.00	n/a	26.85			
11	14				25.42			
11	15	GADPH sandards 1	30000.00	4.48	3.45	3.45		91.00%
11	16							
11	17	GAPDH Standards 2	3750.00	3.57	5.35	5.60		
11	18				5.84			
11	19	GAPDH Standards 3	468.75	2.67	8.53	8.57		
11	20				8.61			
11	21	GAPDH Standards 4	58.59	1.77	12.18	12.32		
11	22				12.46			
11	23	GAPDH Standards 5	7.32	0.86	16.20	16.21		
11	24				16.22			
11	25	GAPDH Standards 6	0.92	-0.04	18.69	18.88		
11	26				19.06			
11	27	Grb2 no template control	0.00	n/a				
11	28							

PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
11	29	Grb3-3	1-T	14.46	14.14	0.72	5.28	0.18	18.45
11	30			13.81					
11	31	GAPDH		13.50	13.56	1.46	28.64	n/a	n/a
11	32			13.61					
11	33	Grb3-3	2-T	17.59	17.16	-0.25	0.56	0.09	9.20
11	34			16.73					
11	35	GAPDH		15.96	15.95	0.79	6.10	n/a	n/a
11	36			15.94					
11	37	Grb3-3	3-T	14.65	14.91	0.48	2.99	0.16	16.17
11	38			15.16					
11	39	GAPDH		14.12	14.24	1.27	18.46	n/a	n/a
11	40			14.35					
11	41	Grb3-3	4-T	16.47	17.13	-0.24	0.57	0.59	58.87
11	42			17.79					
11	43	GAPDH		18.76	18.79	-0.01	0.98	n/a	n/a
11	44			18.82					
11	45	Grb3-3	5-T	12.48	13.16	1.04	10.92	0.21	20.99
11	46			13.83					
11	47	GAPDH		12.60	12.63	1.72	52.05	n/a	n/a
11	48			12.66					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
12	1	Grb2 Standards 1	200.00	2.30	10.45	10.61		110.00%
12	2				10.76			
12	3	Grb2 Standards 2	25.00	1.40	13.22	13.29		
12	4				13.35			
12	5	Grb2 Standards 3	3.13	0.49	18.43	17.31		
12	6				16.18			
12	7	Grb2 Standards 4	0.39	-0.41	20.94	21.02		
12	8				21.09			
12	9	Grb2 Standards 5	0.05	-1.32	21.41	22.05		
12	10				22.69			
12	11	Grb2 no template control	0.00	n/a				
12	12							
12	13	GADPH sandards 1	3750.00	3.57	5.74	5.69		70.00%
12	14				5.63			
12	15	GAPDH Standards 2	468.75	2.67	9.12	9.10		
12	16				9.07			
12	17	GAPDH Standards 3	58.59	1.77	13.75	13.78		
12	18				13.81			
12	19	GAPDH Standards 4	7.32	0.86	18.40	18.40		
12	20				18.40			
12	21	GAPDH Standards 5	0.92	-0.04	21.00	20.77		
12	22				20.54			
12	23	Grb2 no template control	0.00	n/a				
12	24							

PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
12	25	Grb2	1-N	15.37	15.63	0.85	7.15	1.24	124.19
12	26			15.89					
12	27	GAPDH		17.92	17.95	0.76	5.76	n/a	n/a
12	28			17.98					
12	29	Grb2	2-N	12.26	12.33	1.83	67.76	1.06	105.71
12	30			12.39					
12	31	GAPDH		13.13	13.38	1.81	64.10	n/a	n/a
12	32			13.62					
12	33	Grb2	3-N	13.27	13.29	1.55	35.14	0.69	69.48
12	34			13.31					
12	35	GAPDH		13.71	13.83	1.70	50.57	n/a	n/a
12	36			13.94					
12	37	Grb2	4-N	22.15	22.29	-1.11	0.08	2.30	229.69
12	38			22.43					
12	39	GAPDH		27.89	27.72	-1.47	0.03	n/a	n/a
12	40			27.55					
12	41	Grb2	5-N	12.29	12.15	1.88	76.33	0.75	75.30
12	42			12.01					
12	43	GAPDH		12.45	12.51	2.01	101.37	n/a	n/a
12	44			12.56					

PCR Run	PCR Tube No.	Sample ID	Amount cDNA	Log Amount CDNA	CT value	Average CT value	Standard Curves	Amplification Efficiency
13	1	Grb3-3 standard 1	30.00	1.48	11.11	11.15		96.00%
13	2				11.19			
13	3	Grb3-3 standard 2	3.75	0.57	14.67	14.63		
13	4				14.59			
13	5	Grb3-3 standard 3	0.46	-0.34	18.66	17.55		
13	6				16.43			
13	7	Grb3-3 standard 4	0.06	-1.23	21.65	21.98		
13	8				22.31			
13	9	Grb3-3 standard 5	0.01	-2.15	22.30	23.28		
13	10				24.25			
13	11	Grb3-3 no template control	ntc		22.40			
13	12				24.57			
13	13	GADPH sandards 1	3750.00	3.57	5.29	5.32		90.00%
13	14				5.35			
13	15	GAPDH Standards 2	468.75	2.67	7.98	8.02		
13	16				8.05			
13	17	GAPDH Standards 3	58.59	1.77	11.07	11.13		
13	18				11.19			
13	19	GAPDH Standards 4	7.32	0.86	14.63	14.67		
13	20				14.71			
13	21	GAPDH Standards 5	0.92	-0.04	18.04	18.23		
13	22				18.41			
13	23	Grb2 no template control	0.00	n/a				
13	24							

PCR Run	PCR Tube No.	Primer pair	RNA ID	CT value	Average CT value	Log10 amount cDNA (pg)	Amount cDNA (pg)	Normalised amount cDNA	Normalised amount cDNA (x100)
13	25	Grb3-3	1-N	16.35	16.01	0.16	1.43	1.02	101.84
13	26			15.67					
13	27	GAPDH		17.69	17.30	0.15	1.41	n/a	n/a
13	28			16.90					
13	29	Grb3-3	2-N	13.73	13.43	0.90	7.91	0.55	54.83
13	30			13.12					
13	31	GAPDH		13.56	13.66	1.16	14.43	n/a	n/a
13	32			13.76					
13	33	Grb3-3	3-N	14.48	14.33	0.64	4.36	0.46	46.30
13	34			14.17					
13	35	GAPDH		14.27	14.33	0.97	9.42	n/a	n/a
13	36			14.38					
13	37	Grb3-3	4-N	23.05	22.73	-1.77	0.02	15.43	1542.76
13	38			22.41					
13	39	GAPDH		27.95	28.48	-2.96	0.00	n/a	n/a
13	40			29.00					
13	41	Grb3-3	5-N	13.40	13.46	0.89	7.75	0.33	33.36
13	42			13.51					
13	43	GAPDH		12.87	12.92	1.37	23.25	n/a	n/a
13	44			12.96					