Functional Analysis of a Rare Disease Variant in Human *NRAS* Causing Noonan-Like Syndrome

Jessica Cale

(31854091)

Bachelor of Science Honours in Molecular Biology

School of Veterinary and Life Science

Murdoch University

October, 2016

I declare this thesis is my own account of my research and contains, as its main content, work which has not been previously submitted for a degree at any tertiary education institution.

Jessica Cale

ABSTRACT

Background:

Rare diseases are chronic and debilitating and while individually they affect less than 1 in 2,000 people, collectively they have a huge impact affect around 1.8 million Australians. Approximately 80% of these have a genetic origin, however only 30% of patients receive a formal molecular diagnosis. RASopathies are a group of rare disease that are caused by a mutations in the genes involved in the RAS-MAPK pathway, the most common of which is Noonan syndrome, which is characterised by heart defects, short stature, chest deformities, and specific craniofacial features. This study focuses on the effects of a novel c.173C>T (p.Thr58Ile) mutation, found in the NRAS gene of a patient diagnosed with Noonan-like syndrome, on the localisation and function of the NRAS protein. In doing so this study aimed to validate the role of mutations in the patient's disease and provide information toward the creation of an experimental pipeline for the validation of other rare variants.

Methods:

U87-MG cells were transiently transfected with the NRAS constructs tagged with GFP2, and stained with specific antibody markers to determine localisation of the proteins using confocal microscopy. Functional studies included an Annexin V apoptosis assay, using flow cytometry to detect and quantify levels of apoptosis in transfected and untransfected populations, and the prediction of conserved domains using online bioinformatic tools.

Results and Conclusions:

Preliminary results from this study suggest that there is a difference between the localisation and function of the mutant and wild type proteins. The mutant protein was seen to co-localise with the Golgi apparatus as expected if the mutant protein was constitutively active. However the mutant protein was also observed to co-localise with the markers for the plasma membrane, nucleus and endoplasmic reticulum indicating that the mutation affects more than just the activation of the protein.

An increase in apoptosis observed in $NRAS^{MUT}$ transfected population, in comparison to both the wild type and untransfected populations, indicates that the mutation is engaging the pro-apoptotic function of NRAS associated with an increased activation of the RAS–RAF–MAPKK–MAPK pathway. Bioinformatic analyses identified that the mutation is location within a number of motifs involved in the binding of GTP and thus the activation and inactivation of NRAS, directing further studies toward the proliferation, activation and interaction of NRAS with effectors.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Full gene names given in Appendix I

DNA	Deoxyribonucleic acid
SNV	Single nucleotide variant
CNV	Copy number variation
NGS	Next-generation sequencing
WGS	Whole genome sequencing
WES	Whole exome sequencing
RUDDS	Rare and undiagnosed diseases diagnostic service
GSWA	Genetic Services of Western Australia
FORGE	Finding of Rare Disease Genes in Canada
DDD	Deciphering Developmental Disorders project
CS	Costello syndrome
CFC	cardio-facio-cutaneous syndrome
NSML	Noonan syndrome with multiple lentigines
GTP	Guanosine triphosphate
HVR	Hypervariable region
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GAP	GTPase-Activating Proteins
TGN	*Trans*-Golgi network
GFP2	Green fluorescent protein 2
PMCA	Anti-Calcium Pump pan PMCA ATPase antibody
PS	Phosphatidylserine
EF-Tu	Elongation factor thermo unstable
CRISPR	Clustered regularly interspaced short palindromic repeats
Cas	CRISPR associated protein
PCR	Polymerase chain reaction
bp	Base pair

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my four supervisors for their patience and support throughout my honours year. Specifically I would like to thank Professor Jenefer Blackwell and Dr Sarra Jamieson, for their help especially whilst writing my thesis, and Genevieve Syn who, despite doing a PhD, provided me with extensive help and taught me the many lab skills I needed throughout my Honours. Lastly I would like to thank Andrew Currie for his help throughout the honours enrolment process.

Thank you to the Telethon Kids Institute and SPACH who provided me with a vacation scholarship, and Murdoch University for awarding me with a MUAEA completing scholarship. This help has allowed me to study what I love.

My sincere thanks to all the staff at Telethon Kids Institute who helped throughout the year, especially Tamara Abel for teaching me how to use the confocal microscope, and Arunesh Mohandas for teaching me about flow cytometry. I would like to thank the Genetics and Health group in particular, who welcomed me and made Telethon Kids a great place to be every day. Thank you especially to, Emily Kempin for keeping me sane throughout the year, despite going through the same ups and downs as me, and Rachael Lappan for answering my endless questions and always being so willing to help.

Thank you to my friends and family, who have supported me throughout the year and provided me with a much needed social outlet. Last but not least, Mum and Dad, thank you for your support from the first day until the last, you have enabled me to go so much further than I could have ever hoped.

1 LITERATURE REVIEW

1.1 Rare Diseases

Diseases can be devastating and are often a lifelong concern, rare diseases in particular are one of the most difficult types of diseases to diagnose and manage due to their rare nature, heterogeneity, and often unknown cause. In the United Kingdom a rare disease is defined as a debilitating or chronic disease that affects less than 1 in 2,000 people,[1] this is also the definition adopted by Australia, however many rare diseases are much more uncommon.[2] Despite being individually infrequent there are between 5,000 and 8,000 different rare diseases meaning that when assessed collectively they affect around 8% of the population,[1] this equates to approximately 1.9 million Australians making them a serious health concern.[3]

1.1.1 Diagnostic Odyssey

More than 80% of rare diseases are believed to have a genetic origin.[3] This origin can be the result of variations in the deoxyribonucleic acid (DNA) such as coding or splicing single nucleotide variants (SNVs), copy number variants (CNVs) and coding or splicing insertions and deletions (indels).[4] Despite extensive research and numerous discoveries on the causes of genetic diseases via phenotypic diagnosis, up to half of children suffering from such diseases remain without a genetic based, or molecular, diagnosis[4] and only approximately 30% have received an official clinical diagnosis. Of those that have a diagnosis approximately 30% waited five years for the diagnosis and in approximately 50% of cases the first diagnosis was wrong.[5,6]

While some rare diseases are more easily diagnosed; for example those diseases that have a Mendelian inheritance pattern and are monogenic,[7] others present a challenge

for diagnostic tools. The ability to diagnose these rare diseases is challenged by their infrequency and by their heterogeneity; as they often affect few individuals and families per disease[7] and multiple mutations can cause a phenotypically similar set of diseases. This is further enhanced by the fact that these diseases are rare because the mutations causing them are under strong negative selection.[7] This means that transmission to subsequent generations is infrequent, with mutations often being *de novo* in nature thus undiscoverable through non-sequence based techniques such as linkage analysis.[7] The infrequency of these variants; which leads to underpowered analysis, and their other characteristics mean that it is often necessary to use sequencing to directly detect these mutations.[7]

1.1.1.1 Sanger sequencing

Sanger sequencing has been the main technique utilised for the molecular diagnosis of genetic diseases and is still the test used in clinical genetic testing to confirm a diagnosis.[8] Sanger sequencing works by incorporating tagged di-deoxynucleotides to single stranded DNA one base at a time. These di-deoxynucleotides differ from natural nucleotides in two ways; they have a fluorescent tag so they can be visualised and they lack a hydroxyl group meaning that when they are incorporated into the DNA synthesis ceases and the fragment of DNA detaches from the synthesis complex. Using these di-deoxynucleotides the sequence is built up, small fragments of DNA at a time. These fragments are then 'read' on a gel or through capillary electrophoresis to get the final DNA sequence. While Sanger sequencing is exact and accurate it is limited to analysing a single segment of DNA at a time making it neither economical nor efficient.[8] The limitations of Sanger sequencing make it inefficient for the diagnosis of rare diseases when multiple genes need to be sequenced, thus over the past decade,

there has been an increase in the use of next-generation sequencing (NGS) techniques.[8]

1.1.1.2 Next Generation sequencing

NGS is a sequencing platform that allows for the rapid sequencing of multiple genes in one reaction to a single nucleotide degree on the whole genome scale making NGS time and cost efficient.[7-9] There are four main NGS techniques; pyrosequencing, Ion torrent semiconductor sequencing, sequencing by ligation (SOLiD) and sequencing by synthesis. The more commonly applied currently, is sequencing by synthesis which works on the same premise to Sanger sequencing, however all four di-deoxynucleotides, and no nucleotides, are added at the same time each with a different florescent tag. This technique, along with the other three NGS techniques, allow for multiple sequencing runs to happen simultaneously. Because of this advantage NGS has led to the identification of possible pathogenic mutations in multiple genes that are potentially associated with a disorder through whole genome (WGS) or whole exome sequencing (WES; sequencing the protein coding regions and their regulatory sequences).[8,9]

The ability to more easily sequence a large portion of the genome, combined with its consistency and robustness, makes NGS an important tool in diagnosis.[9] In particular one of the advantages of NGS is that there is no requirement for prior knowledge as the whole genome, the exome or a panel of genes can be sequenced rather than a specific target; this makes it the perfect tool to use to diagnose rare diseases when the genetic origin is unknown or when multiple different mutations could be causing the disease.[10] NGS has also allowed for the expansion of phenotype-genotype

association studies.[11,12] This is because NGS allows for the unbiased sequencing of a larger portion of the genome and the relatively inexpensive sequencing of genomes of both healthy and unhealthy individuals. This allows for mutations that would not previously have been sequenced to be identified and compared to healthy individuals thus linking to a particular phenotype.[11,12]

Despite these advantages, WES and WGS remain primarily in the research diagnosis scene, they are still not often used in the clinical setting where Sanger sequencing is still the gold standard.[13] NGS of a target panel of genes is however more frequently used in clinical diagnosis. The advantage of a targeted panel over WES and WGS is its increased speed of execution and reduced cost, while still retaining its robustness and reproducibility.[9] A targeted panel also decreases the likelihood of incidental findings which is one of the major problems associated with WES and WGS.[14]

1.1.1.3 Current diagnostic pipelines

Following thorough phenotyping of each patient the current diagnostic pipeline for an individual with a rare genetic disease involves gene-by-gene diagnostic tests.[5,13] Sanger sequencing is most often used for this stage and is most helpful if the phenotype is highly suggestive of a specific causative mutation. However, in cases where no mutations are found in the suspected gene, other possible genes need to be tested, leading to a long and costly diagnosis.[13]

In Western Australia the rare and undiagnosed diseases diagnostic service (RUDDS) was recently implemented within Genetic Services WA (GSWA).[5] With this pipeline if no molecular diagnosis has been achieved after the use of single gene techniques then

WES data is generated allowing the analysis of multiple bioinformatically selected genes from the WES data[5]. A gene panel rather than the whole exome sequence is analysed currently to reduce the number of incidental findings.[13] This pipeline currently give a diagnosis rate of approximately 30%.[5] There are several studies both internationally: Finding Of Rare Disease Genes (FORGE),[15] Deciphering Developmental Disorders (DDD),[16] and locally such as the collaborative SeqNextGen project based at Telethon Kids Institute, that aim to increase this diagnostic rate by analysing the complete exome. International studies suggest that this should increase the diagnostic rate to between 45 and 50%.

This honours project focuses on a single novel variant identified in the NRAS gene of a patient diagnosed with Noonan-like syndrome through the RUDDS service and attempts to validate the role of this variant in the disease.

1.2 RASopathies

One group of rare diseases are referred to as the RASopathies. RASopathies are caused by mutations in the RAS/mitogen-activated protein kinase (RAS/MAPK) pathway[17,18] (Figure 1.1). RASopathies include Noonan Syndrome (OMIM# 163950), Costello syndrome (CS; OMIM# 218040), Cardio-facio-cutaneous syndrome (CFC; OMIM# 115150) and Noonan syndrome with multiple lentigines (NSML; 151100), each is associated with one or more components of the RAS/MAPK pathway. This diagram shows how mutations within more than one gene can be responsible for the same phenotype and mutations within the same gene can cause a number of different syndromes. One example of this is that mutations in *KRAS* can lead to CFC or Noonan

syndrome, and Noonan syndrome can be caused by mutations in a number of different genes (Figure 1.1).

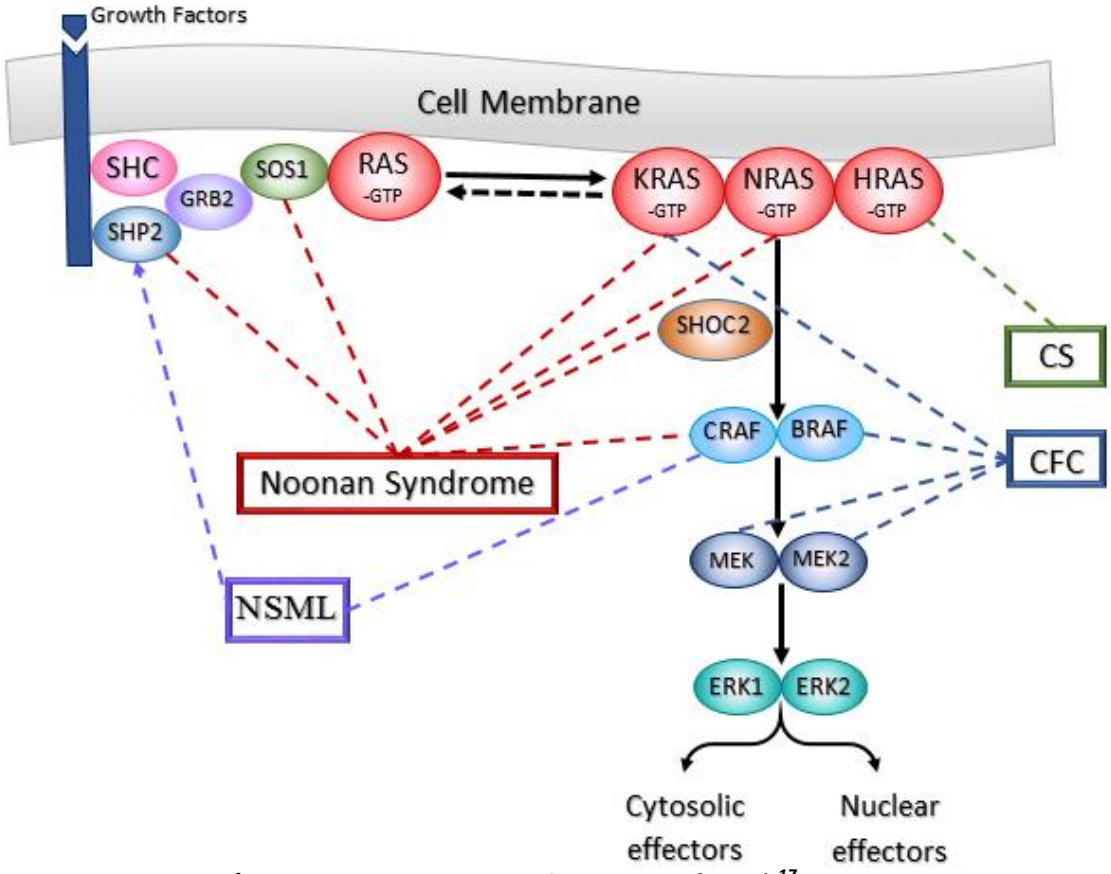


Figure 1.1: The RAS/MAPK pathway adapted from Rauen (2013).[17]

This pathway is involved in cell proliferation, differentiation, apoptosis and senescence[17]*. The boxes and dashed lines indicate RASopathies and their associated genes respectively. These diseases include Noonan Syndrome, Costello syndrome (CS), Cardio-facio-cutaneous syndrome (CFC) and Noonan syndrome with multiple lentigines (NSML). Full gene names in Appendix I*

1.2.1 Noonan Syndrome

Noonan syndrome while a rare disease, is the most common of the RASopathies.[17,18]

There are ten different recognised phenotypes of Noonan syndrome which affect between 1 in 1,000 to 2,500 people. However, because of the range in severity of symptoms this is likely to be an underestimate of the true number of people living with a Noonan syndrome.[19]

1.2.1.1 Characteristics of Noonan syndrome

The ten phenotypes of Noonan syndrome share common clinical signs including heart defects, cognitive deficits, short stature, chest deformities such as pectus excavatum, pterygium colli or a webbed neck (Figure 1.2a) and specific craniofacial features.[19-21] These craniofacial features include low-set ears (Figure 1.2b), ocular ptosis or drooping of one or both eyelids (Figure 1.2c), hypertelorism or wide set eyes (Figure 1.2d) and down-slanting palpebral fissures.[21] These clinical signs however differ with age and decrease in severity between individuals, leading to mild cases often being overlooked.[22]

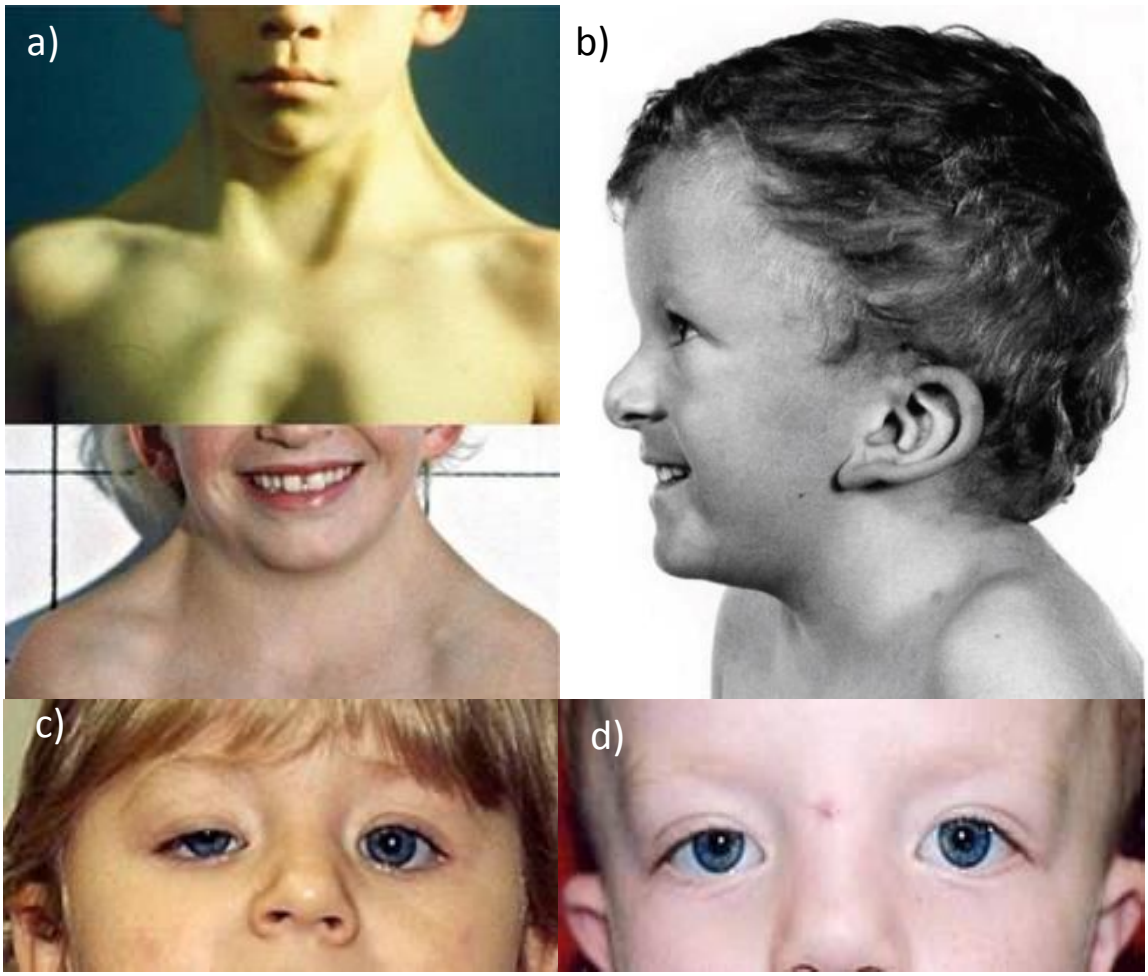


Figure 1.2: Clinical signs of Noonan syndrome

Including a) webbed/broad neck, b) low-set ears,[20] c) ocular ptosis and d) hypertelorism. Images courtesy of National Human Genome Research Institute[23] and Cirstea et al (2010)[20].

A study by Allanson (2010) attempted to define the specific phenotypic traits and link them to specific genotypes of Noonan.[24] In this study the faces of 81 individuals with a: *PTPN11, SOS1, RAF1* or KRAS mutation, were studied by two experts. It was determined that facial features alone were not sufficiently different between genotypes to allow for the patient's genotype to be predicted from phenotype alone. The study did however, conclude that certain features such as face shape and hair type are suggestive of a particular genotype.[24]

The phenotype of Noonan syndrome changes throughout development,[25] in the newborn the most prominent features include hypertelorism, macrocephaly, a deeply grooved philtrum, low set ears and wispy hair. During childhood and into adolescence the face typically becomes more triangular with a wide brow and hair may become curly or woolly.[24] During adolescence, as the neck lengthens, webbing becomes more pronounced.[24]

1.2.1.2 Treatment and management of Noonan syndrome

Noonan syndrome is most often 'treated' through the management of its symptoms. This can include early childhood surgery to correct ocular deformations such as strabismus and ptosis, as well as early evaluation and management of cardiac function and surgery depending on the severity.[19] Growth and motor development are also assessed frequently and speech therapy and physiotherapy are commonly required.[19] By adulthood the majority individuals with Noonan syndrome are able to function normally in the adult world due to this early stage management and counselling.[19] This is further enhanced by the reduction of signs and symptoms with age.

1.2.1.3 Causes of Noonan syndrome

Noonan syndrome has been associated with multiple different genes involved in the RAS/MAPK pathway, including the more frequently associated *PTPN11* which accounts for 50% of cases and *SOS1* (10-13%) as well as less frequent mutations in the *KRAS* (<5%), *MAP2K1* (<2%), and *NRAS* (<1%) genes.[18] The ten phenotypes of Noonan syndrome, while sharing common symptoms, are each associated with mutations in a particular gene. There are also multiple Noonan-*like* syndromes each differing slightly in morphology from Noonan syndrome. Each phenotype of Noonan syndrome, their mode of inheritance, and their associated gene are shown in *Table 1.1*.

In general Noonan syndrome has an autosomal dominant inheritance pattern and, while *de novo* mutations are commonly the cause of Noonan, between 30%-75% of cases have an affected parent.[19] The exception to this is Noonan syndrome 2 which has little known about it including its cause/s and is the term used for the phenotype of Noonan syndrome that is inherited in an autosomal recessive manner.[26-28]

Table 1.1: Noonan and Noonan-Like syndromes, their inheritance and associated genes (OMIM Database)[29]

Name	Mode of inheritance	Associated Gene	OMIM number
Noonan syndrome 1	Autosomal dominant	PTPN11	163950
Noonan syndrome 2	Autosomal recessive	unknown	605275
Noonan syndrome 3	Autosomal dominant	KRAS	609942
Noonan syndrome 4	Autosomal dominant	SOS1	610733
Noonan syndrome 5	Autosomal dominant	RAF1	611553
Noonan syndrome 6	Autosomal dominant	NRAS	613224
Noonan syndrome 7	Autosomal dominant	BRAF	613706
Noonan syndrome 8	Autosomal dominant	RIT1	615355
Noonan syndrome 9	Autosomal dominant	SOS2	616559
Noonan syndrome 10	Autosomal dominant	LZTR1	616564
Noonan syndrome-like disorder with loose anagen hair	Autosomal dominant	SHOC2	607721
Noonan syndrome-like disorder with or without juvenile myelomonocytic leukemia	Autosomal dominant	CBL	613563
neurofibromatosis-Noonan syndrome	Autosomal dominant	NF1	601321

1.2.2 The RAS Family

The RAS proteins, named for their role in forming **ra**t **s**arcoma, are members of a large family of guanosine triphosphate (GTP) binding proteins and a major part of the RAS-MAPK pathway (Figure 1.1). *RAS* has three isoforms; Neuroblastoma *RAS* viral oncogene homolog (*NRAS*), V-KI-RAS2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*) and V-HA-RAS Harvey rat sarcoma viral oncogene homolog (*HRAS*), which are universally expressed and are highly conserved with more than 90% sequence similarity.[20,30] These isoforms differ only in a region of their N-terminal called the hypervariable region (HVR), which is involved in their function and localisation within the cell. Due to this HVR the three isoforms are functionally different and thus not redundant.[30]

The RAS proteins work as molecular switches for signalling pathways including the RAF1/ERK, PI3K/AKT and RalGDS pathways (Figure 1.3).[31,32] Through these pathways the RAS proteins have a role in the regulation of many cellular functions including cell growth, migration, survival, apoptosis and endocytosis (Figure 1.3).[33] As shown in Figure 1.3 this is achieved through their intrinsic GTP*ase* activity by cycling between their inactive guanosine diphosphate (GDP)-bound and active GTP-bound states.[34] This cycle is facilitated by guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs) which aid in the hydrolysis and binding of GDP/GTP respectively.[33,35]

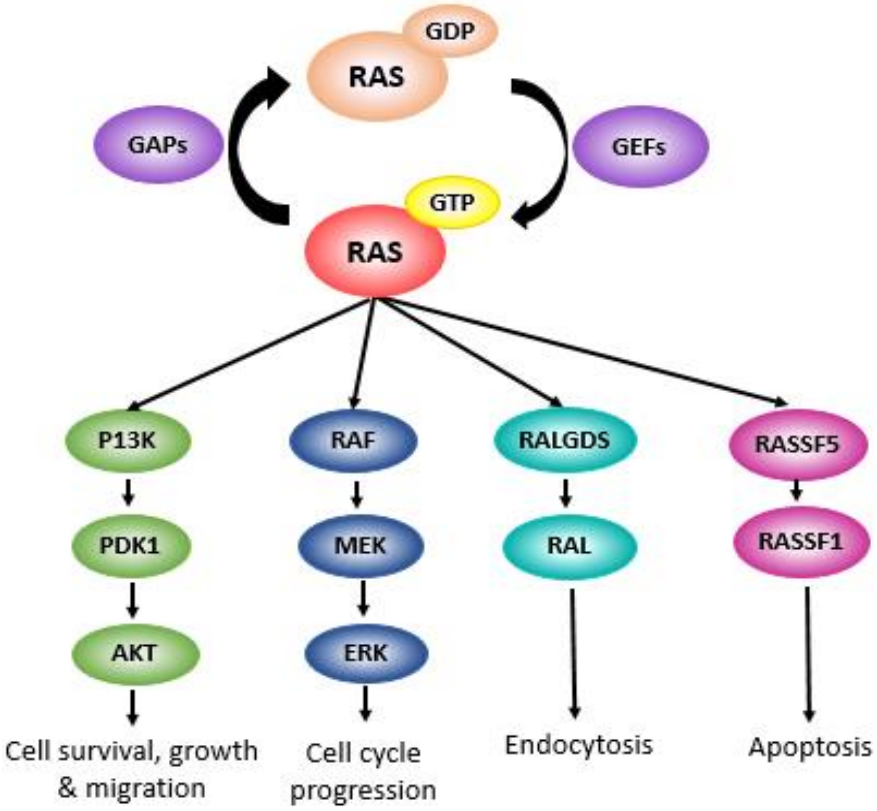


Figure 1.3: Simplified RAS pathway
Showing the activation of RAS through GTP binding and inactivation through hydrolysis of GDP mediated by RAS-GAP and RAS-GEF, as well as the pathways for which this activation/inactivation acts as a molecular switch.

Due to their role in a large number of processes essential for survival and growth, mutations in the RAS proteins can cause serious and debilitating diseases including a large number of different cancers, such as carcinoma[36,37] and hematologic malignancies,[38] as well as *RAS*-associated autoimmune leukoproliferative disorder,[39,40] Noonan syndrome, CS[41,42] and congenital melanocytic nevus syndrome.[43]

1.2.2.1 Structure of the RAS isoforms

The RAS genes are made up of six (*HRAS* and *KRAS*) or seven (*NRAS*) exons[44], four of which code for the protein, shown in blue in Figure 1.4. The RAS proteins have four main regions, a P-loop which is involved in the binding of phosphates and two switch regions (SW I and SW II in Figure 1.4) which are responsible for the change in

conformation resulting from the activation or inactivation of the RAS proteins after binding/hydrolysis of GTP (Figure 1.4).[20] As well as the HVR, which consists of the last 23 (*KRAS*) or 24 (*HRAS* and *NRAS*) amino acids, this region is responsible for the subcellular localisation and function of the RAS proteins.[30]

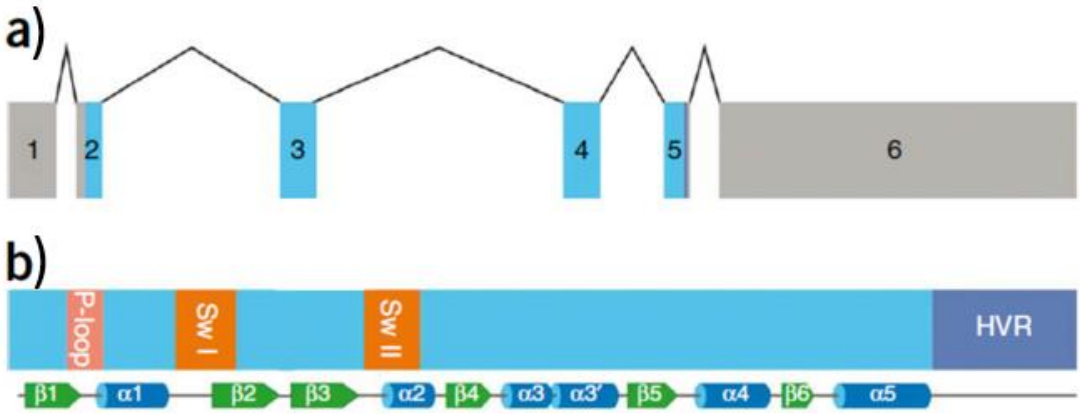


Figure 1.4: Diagram of NRAS structure modified from Cirstea et al. 2010[20]
a) organisation of the NRAS exons, with coding exons in blue and non-coding exons in grey. (b) The four main regions of the RAS proteins including the P-Loop, Switch I (SW I), Switch II (SW II) and the hypervariable region (HVR), as well as secondary structures

The RAS proteins function as a molecular switch by cycling between their active GTP bound and inactive GDP bound states.[20] In the active GTP-bound state the switch I and switch II motifs (Figure 1.4) undergo a conformational change that enables RAS to bind to effector proteins and activate them.[20,45] This interaction is terminated by hydrolysis of GTP to GDP, which promotes a switch toward the inactive conformation of RAS.[20] Mutations in any of these regions or in residues that would cause the disruption of any of these regions can cause serious diseases. For example, mutation in amino acids 12, 13, 59, 61, or 63, lead to impaired GTPase activity,[46] and due to the importance of switch I in the binding of GAPs to RAS, mutations in this region could result in RAS remaining in its active state.[47]

1.2.2.2 Localisation of the RAS isoforms

The RAS proteins are expressed in almost all tissue types with the exception of adipose tissue.[48] Their expression is enhanced in the pancreas, cerebellum and lower in soft tissue.[48] While within the cell, the RAS proteins have been observed in multiple subcellular compartments including the Golgi apparatus, plasma membrane, endoplasmic reticulum and the inner and outer membranes of mitochondria.[30,34,49]

A study by Wolman *et al.* (2006) showed that NRAS has a relatively wide distribution being present on most cellular membrane containing organelles, with at least three distinct subcellular localisations: the plasma membrane, Golgi apparatus and mitochondria.[34] This is supported in studies by Quatela and Philips (2006) which showed that there is a large amount of trafficking of HRAS and NRAS bi-directionally between the Golgi complex and the plasma membrane.[49] Another study showed that tubular post-Golgi carriers transport a number of membranous proteins, including NRAS, from the *trans*-Golgi network (TGN) to the cell surface.[50]

The subcellular distribution of the RAS isoforms has been suggested to be determined by their activation and, in turn, regulate their function. This came to light when studies found that, while active HRAS is associated with both the Golgi apparatus and the endoplasmic reticulum , only active HRAS associated with the endoplasmic reticulum is able to activate the RAF1–ERK pathway.[51] Another example of function differing between subcellular locations is that when KRAS is localised to the plasma membrane it can induce transformation, however when it is associated with the mitochondria it induces apoptosis.[52] Activation of the RAS proteins is also thought to be associated with their location within the cell as they are transiently active on the

plasma membrane while constitutively active on the Golgi apparatus.[51,53] This indicates that, if the localisation of the protein is altered due to the disruption of a localisation signal or an alteration to the proteins folding or structure due to a mutation, the activation and function of the protein could be significantly altered thus leading to disease.

The localisation of the RAS proteins is regulated by the C-terminal HVR.[30] It is because of this HVR that the RAS proteins have become typical examples of isoform and compartment specific signalling, as the three isoforms are differentially distributed on endomembranous compartments.[30] This region is also thought to determine the function and, after multiple post-translational modifications, the localisation of the isoforms.[30] While the post translational modifications of the HVR are responsible for the regulation of RAS protein localisation, the result is dynamic and can be altered by the activation status of RAS and its interacting proteins.[30] Localisation to the Golgi apparatus is more pronounced for NRAS than for HRAS because NRAS is palmitoylated once during post translational modifications causing it to be more likely to lose anchorage in the plasma membrane and thus return to the Golgi apparatus than the dipalmitoylated HRAS.[30]

1.2.2.3 Known mutations of the RAS family members

The role of RAS proteins in a number of different diseases is well documented, a number of examples of which are outlined in Table 1.2 focusing particularly on those mutations that cause Noonan or Noonan-like syndromes. Table 1.2 also contains all known cases of Noonan syndrome caused by mutations in *NRAS*.

The most common of the disease causing mutations in the RAS proteins are those that impair their intrinsic GTP*ase* activity.[33] For example mutations in residue 61 of NRAS affect the GTP binding site and result in NRAS being constitutively activated.[54] Other mutations such as those that affect residues 32-40, which are crucial for the putative effector region of RAS*,* result in the inhibition of biological activity and impair RAS-GAP binding.[46] Such mutation could lead to the protein being transported or 'stuck' in the wrong location and thus not be able to carry out its proper function. This is particularly the case for RAS as studies have demonstrated that the localisation of the RAS proteins is associated with their function for example, transiently activated by growth factors on the cell membrane however they remain activated when on the Golgi apparatus.[51]

Table 1.2: Examples of known mutations in the RAS isoforms associated with diseases including every known case of NRAS associated Noonan syndrome.

Source	Schubbert *et al* (2006)[45]	Oliveira *et al* (2007)[55]	Matsuda *et al* (2007)[56]	Matsuda *et al* (2007)[56]	Zenker *et al* (2007)[57]	Gripp *et al* (2008)[58]	Dessars *et al* (2009)[43]	De Filippi *et al* (2009)[59]
Gene	K*RAS*	*NRAS*	*NRAS*	*NRAS*	*KRAS*	*HRAS*	*NRAS*	*NRAS*
Mutation	Thr58Ile (T58I)	Gly13Asp (G13D)	Gly12Asp (G12D)	Gly13Asp (G13D)	Pro34Leu (G34L)	Thr58Ile (T58I)	Gln61Lys (Q61K)	Gly13Asp (G13D)
Disease	Noonan syndrome -3	RAS-associated autoimmune lymphoproliferative syndrome (RALPS)	Juvenile myelomonocytic leukemia (JMML)	Juvenile myelomonocytic leukemia (JMML)	Noonan syndrome -3	Costello syndrome	congenital melanocytic nevi	Noonan Syndrome/juvenile myelomonocytic leukaemia (JMML)
MIM#	#609942	#614470	#607785	#607785	#609942	#218040	#137550	#613224/ #607785
Type	ND	*De novo*	*De novo*	*De novo*	ND	ND	*De novo*	*De Novo*
Notes			2 individuals	2 unrelated individuals			14 of 27 patients	

Table 1.2: Continued

Source	Cirstea *et al* (2010)[20]	Cirstea *et al*. (2010)[20]	Runtuwene *et al* (2011)[60]	Denayer *et al* (2012)[61]	Denayer *et al* (2012)[61]	Denayer *et al* (2012)[61]	Kraoua *et al* (2012)[62]	Ekvall *et al* (2015)[63]
Gene	*NRAS*	*NRAS*	*NRAS*	*NRAS*	*NRAS*	*NRAS*	*NRAS*	*NRAS*
Mutation	Gly60Glu (G60E)	Thr50Ile (T50I)	Ile24Asn (I24N)	Thr50Ile (T50I)	Ile24Asn (I24N)	Pro34Leu (P34I)	Gly60Glu (G60E)	Gly60Glu (G60E)
Disease	Noonan syndrome-6	Noonan syndrome-6	Noonan Syndrome-6	Noonan syndrome-6	Noonan Syndrome-6	Noonan Syndrome-6	Noonan syndrome-6	Noonan syndrome-6
MIM#	#613224	#613224	#613224	#613224	#613224	#613224	#613224	#613224
Type	1x de novo, 1x inherited from mother, 1x ND	*De novo*	*De novo*	ND	*De novo*	Inherited	1x de novo, 1x ND (probably inherited)	1x inherited, 1x ND (probably inherited)
Notes	3 patients from 2 unrelated families	2 unrelated individuals					2 Unrelated individuals	2 related individuals

ND- Not determined

1.2.3 NRAS

NRAS was the last of the three *RAS* isoforms to be described when it was discovered as a transforming gene in neuroblastoma cells in the early 1980s.[64,65] Despite not having been studied as extensively as *KRAS* and *HRAS* due to their similarity they can be used as models for *NRAS*. The involvement of NRAS in Noonan syndrome was described in 2010 in a study by Cirstea *et al.* who found that of 917 patients, who were negative for previously known Noonan-associated mutations, five had a mutation in *NRAS,*[20] indicating that *NRAS* mutations are a rare cause of Noonan syndrome. As of 2015, eight unrelated probands and three families, 14 individuals in total, have been identified to have *NRAS* mutations causing Noonan syndrome.[63] Of these cases, seven are caused by a p.G60E substitution, a further three are caused by a p.T50I substitution and the remainder arise from one of the following substitutions; p.G13D, p.I24N, or p.P34L.[63]

The study by Cirstea *et al.* (2010) found two different mutations in highly conserved regions of *NRAS* in four unrelated individuals.[20] These c.149C>T and c.179G>A mutations caused p.T50I and p.G60E missense substitution in NRAS, respectively.[20] Mutations of the 60th residue had been previously described in the KRAS gene of a patient with CFC syndrome[66] however this study was the first description of a mutation affecting the 50th amino acid.[20] Further study found that this residue is exposed and located within the β2–β3 loop that connects switch I and II.[20] Both mutations were also found to be activating mutations, causing an increase in the phosphorylation of MEK (MAPK/ERK Kinase) when stimulated with epidermal growth factors or in the presence of serum.[20]

Cirstea *et al.* (2010) also found that a p.T50I substitution identified in two unrelated boys, caused enhanced downstream phosphorylation and through molecular modelling showed that this residue interacts with the polar heads of the membrane phospholipid and thus was an integral part of NRAS membrane orientation.[20] This mutation was hypothesised to cause an increase in the interaction of active RAS with its effectors causing enhanced downstream signalling consistent with a gain-of-function mutation.[20]

1.3 Patient background

The patient who is the focus of this study is under the care of GSWA and was diagnosed with Noonan-like disease due to the phenotype displayed, including craniofacial anomalies such as macrocephaly, high forehead, lowset ears, hypertelorism, strabismus and palpebral ptosis. Musculoskeletal abnormalities included webbed neck, pectus excavatum and hyper-extensible joints. The patient also had gross motor delay. When this patient's DNA was sequenced, no known Noonan associated mutations were found. However, a novel single nucleotide substitution (c.173C>T) in the 3rd exon of *NRAS* causing a missense (p.Thr58Ile) substitution in NRAS, was identified. A search for this mutation in the exome aggregation consortium (ExAC)[67] database returned no result however since this particular mutation has not previously been identified to cause Noonan syndrome its role in this patient's disease needs to be validated.

This variant falls in a highly conserved region flanking the switch II motif that is involved in the activation and inactivation of RAS (Figure 1.5). The equivalent mutation has been observed in *HRAS* of a patient with Costello syndrome[58] and *KRAS*

of a patient with Noonan syndrome-3.[45] All three patients share common characteristics such as failure to thrive leading to short stature, cognitive impairment and a webbed neck.[45,58]

Figure 1.5: Four main regions of the RAS, modified from Cirstea et al. (2010)[20]
The four main regions of NRAS are the P-Loop, Switch I (SW I), Switch II (SW II) and the hypervariable region (HVR), this diagram also shows a partial amino acid sequence of NRAS with the mutation of this study in red, note the proximity of the mutation to the switch II region.

Similar to those mutations found by Cirstea *et al.* (2010) that resulted in NRAS being constitutively active, preliminary analysis of the novel p.T58I substitution in this study shows it to be a gain-of-function mutation that results in an increase in the expression of ERK (extracellular signal–regulated kinases) and AKT causing dysregulation of the RAS-MAPK pathway (M. Tartaglia and G. Baynam, personal communication, 2016). No further analysis of this mutation has been carried out to date.

1.4 Research Aims

This project will address the hypothesis that the c.173C>T (p.Thr58Ile) variant identified in the NRAS gene of a patient with Noonan-like syndrome causes a change in localisation and function of NRAS. These results will contribute to the validation of the mutation's role in Noonan-Like syndrome. To achieve this the study involves several aims:

1. To determine if there is evidence to suggest that the p.T58I substitution in *NRAS* disrupts protein localisation, and/or whether the mutation affects other known or predicted functional domains in the protein using *in silico* bioinformatic tools.
2. To determine the subcellular localisation of the mutant and wild type NRAS proteins, by visualising transiently transfected cells co-stained with specific organelle markers using confocal and epifluorescent microscopy.
3. To determine the effect of the p.T58I variant on downstream cell functions, including cellular proliferation, cytokine production, mitochondrial function, and apoptosis, as time permits.
4. Use the knowledge and techniques optimised in this study to contribute to a validation pipeline that can be adapted for the validation of any rare disease variant discovered in Western Australia.

1.5 Significance of Research

The outputs from this project will provide new knowledge that could contribute to the clinical management of the child with this specific rare disease. As well as providing proof-of-principle for the establishment of a functional pipeline that will permit the rapid and efficient study of protein localisation and function of other RASopathies identified in Western Australia with the potential for adaptability to facilitate the study of other rare variants.

2 MATERIALS AND METHODS

Figure 2.1 provides a flow chart of the pipeline for this study, from designing of constructs to localisation, function and bioinformatic analyses to address the three main aims of this study. Details of these methods are provided in sections below.

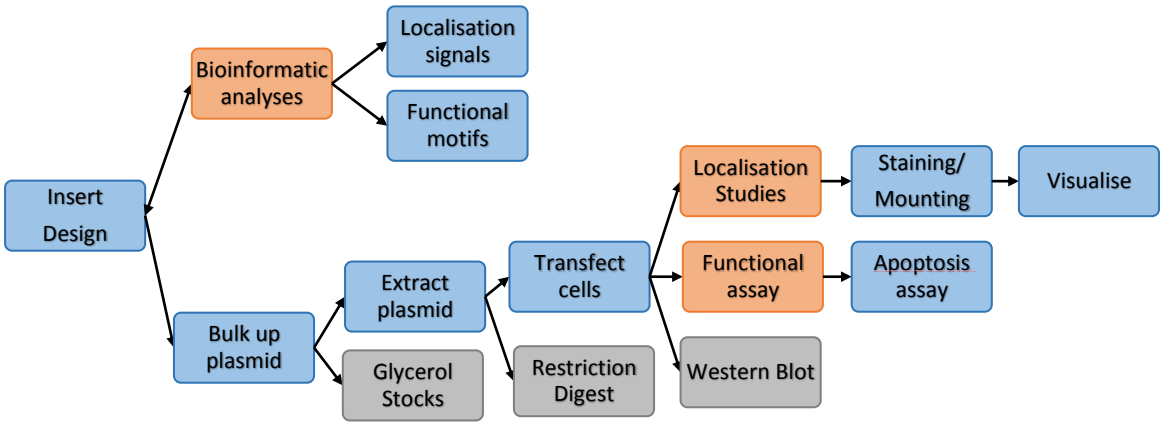


Figure 2.1: Flowchart of the pipeline for this study.
This includes designing of gene inserts to be synthesised and sub-cloned into appropriate vectors by Invitrogen GeneArt® and subsequently used to transfect cells in order to do localisation and functional studies. Orange boxes signify the three main aims of this project.

2.1 Construct Design

The coding DNA sequence for wild type *NRAS* (*NRAS*^{WT}) was obtained from the UCSC Genome Browser.[68] To replicate the mutation found in the patient, the 173rd base of the wild type sequence was altered from a C (cytosine) to a T (thymine) to form the mutant NRAS (*NRAS*^{MUT}) sequence.

From this four constructs were generated with Invitrogen GeneArt gene synthesis (Thermo Fisher Scientific); *NRAS*^{WT} and *NRAS*^{Mut} both tagged with either Green fluorescent protein 2 (GFP2) or 3xFLAG. To design the insert to be synthesised the stop codon was removed from the 5' end of the sequence, then restriction enzyme sites within the sequence were identified using NEBcutter V2.0.[69] Due to the absence

of these restriction sites in both *NRAS* sequences, the restriction site sequences of *Hin*dIII was added to the 5'end and *Bam*HI and *Pst*I were added to the 3' end. Two restriction sites were added to the 3' end of each insert to allow sub-cloning into two different expression vectors: pTagGFP2-N using *Hin*dIII and *Pst*I and p3XFLAG-CMV-14 using *Hin*dIII and *Bam*HI. As codon optimisation by the GeneOptimizer® software in GeneArt™ modifies the input sequence to maximize translation efficacy, the option for codon optimisation was not selected to ensure that the sequence synthesised truly reflected the patient's sequence and mutation.

The synthesised gene inserts were: sub-cloned into pTagGFP2-N and p3XFLAG-CMV-14 vectors, sequence verified, and transformed into *Escherichia coli* by Invitrogen™ GeneArt™ gene synthesis (Thermo Fisher Scientific). A diagrammatic representation of the construct and final translated product are shown in Figure 2.2; full plasmid maps are presented in Appendix II.

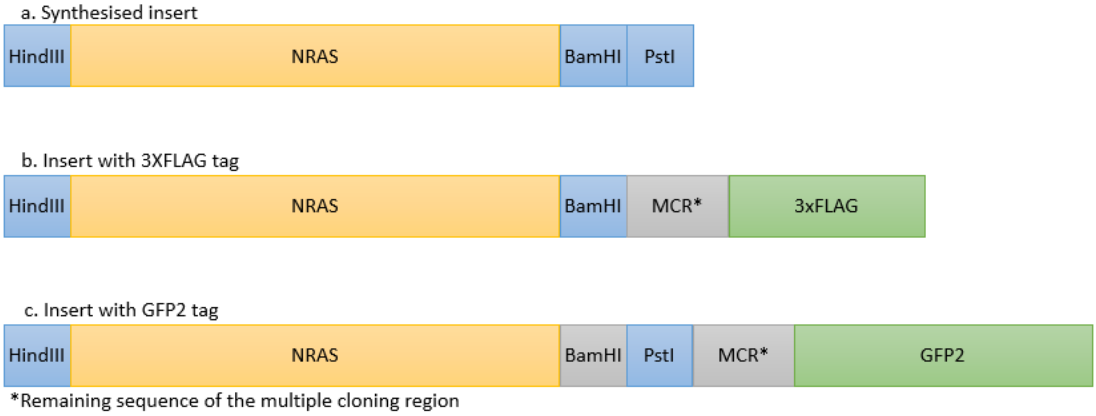


Figure 2.2: Diagrammatic representation of the NRAS insert design
a) NRAS construct with HindIII, BamHI and PstI restriction enzyme sites on the 3 and 5 prime ends to allow for the construct to be sub-cloned into the pTagGFP2-N and p3XFLAG-CMV-14 vectors by Invitrogen™ GeneArt™. b) Construct as it would be when sub-cloned into the p3XFLAG-CMV-14 vector (NRAS-3xFLAG). c) Construct as it would be when sub-cloned into the pTagGFP2-N vector (NRAS-GFP2).

2.1.1 Protein translation

To confirm that the addition of restriction enzyme sites and the GFP2/3xFLAG tags to *NRAS*^{WT} and *NRAS*^{Mut} had no effect on the translation of the proteins and to ensure that the tags would be translated with the protein, the tagged and untagged insert sequences (Appendix III) were translated using the online ExPASy translate tool.[70]

2.1.2 Solvent accessibility

To determine if the mutation or addition of tags were likely to affect the structure of NRAS, the solvent accessibility of each protein sequence was assessed using the online tool ACCpro, which is part of SCRATCH protein predictor.[71] ACCpro predicts the relative solvent accessibility of protein residues and annotates each as either buried or exposed. ACCpro was run with default parameters.

2.1.3 Bacterial Growth and Plasmid extraction

The *NRAS*^{WT} and *NRAS*^{Mut} constructs were provided by GeneArt as lyophilised plasmid DNA and a stab culture of *E. coli* transformed with the plasmid. Stab cultures were streaked onto Lysogeny broth (LB)-agar plates containing 1% Bacto™Tryptone, 5% Bacto™Yeast extract and 1% sodium chloride (NaCl; Biochemicals), with 1.5% Bacto™Agar (Bacto™ reagents from Becton Dickinson (BD)) and supplemented with appropriate antibiotic: kanamycin (50µg/ml; Gibco™/Life Technologies) for GFP2 tagged constructs and ampicillin (100µg/ml; Gibco™/Life Technologies) for 3xFLAG tagged constructs. Plates were incubated at 37^{o}C for 16-18 hours before a single colony was selected and transferred to 1ml of LB medium with appropriate antibiotic and incubated for 7 hours at 37^{o}C with constant agitation at 300rpm using a KS 4000i Control incubator (IKA). This starter culture was then transferred to 50ml of LB medium with appropriate antibiotic to be incubated for a further 16-18 hours at 37^{o}C

with constant agitation at 300rpm. From these bacterial cultures both glycerol stocks and samples for plasmid extraction were taken. Glycerol stocks were made by adding 850ul of bacterial culture to 150ul of 50% filter sterilised glycerol (MERCK) and were kept at -80^{0}C for long term storage.

Plasmid DNA for the GFP2 tagged constructs was initially extracted using two different kits, PureYield™ plasmid miniprep system (Promega) and FavorPrep plasmid extraction mini Kit (FAVORGEN), to determine which kit provided the highest yield. After the first extraction all subsequent extractions were performed using the FavorPrep Kit. Both kits were used as per manufacturer's protocol with the modification of a repeated elution step to ensure complete collection of plasmid DNA. Extracted DNA was quantified on the Nanodrop® ND-1000 spectrophotometer (Thermo Fisher Scientific) using FavorPrep elution buffer as the blank and with duplicate samples of 2µl measured. Plasmid DNA was subsequently stored at -20^{o}C until required.

2.1.4 Restriction enzyme digest

Two restriction enzyme digests were performed on the GFP2 tagged constructs in order to confirm that the insert was of the expected size and that it was inserted correctly. The first was a single digest reaction using *Nhe*I-HF® (New England Biolabs) to linearise the plasmid. This reaction was carried out as per manufacturer's protocol with an extended incubation time of 2 hours to ensure digestion was complete. A 1% agarose gel (Biochemicals) stained with ethidium bromide (final concentration of 1.875µg/ml; Fisher) was loaded with 100ng of digested samples, along with their uncut counterparts, 1:1 with cresol red (Sigma-Aldrich). Hyperladder 1kb (2µl; Bioline)

was used as a molecular weight marker. Samples were electrophoresed for 120 minutes at 130V in 1x Tris/Borate/EDTA (TBE; AMRESCO®) buffer. The gel was then visualised under ultraviolet light using GelDoc™ XR+ (Biorad).

A second restriction digest was performed using the *Hin*dIII and *Pst*I enzymes to cut the gene insert from the plasmid. A sequential restriction digest was used as different buffers were required for the two restriction enzymes used: Buffer E for *Hin*dIII and Buffer H for *Pst*I (Promega, kindly donated by the Cancer Immunology Unit at Telethon Kids Institute). DNA from the pTagGFP2-N vector, $NRAS^{WT}$ and $NRAS^{Mut}$ constructs was incubated for 2 hours at 37^0C in *Hin*dIII reaction mix made as per manufacturer's instructions. Following this samples were purified using an UltraClean™ PCR Clean-up™ kit (MoBio) as per manufacturer's instructions and quantified using Nanodrop® spectrophotometry to allow the volumes in the *Pst*I reaction mix to be adjusted for the concentration of DNA. Samples were then incubated in *Pst*I reaction mix made as per manufacturer's instructions for 2 hours at 37^0C. Samples were electrophoresed for 45 minutes at 120V and visualised as previously described. A restriction digest was not performed on 3xFLAG tagged constructs.

2.2 Cell Culture

Three adherent cells lines were used for the localisation studies; COS-7 (ATCC® CRL-1651™), U87-MG (ATCC® HTB-14™) and HEK293T (ATCC® CRL-3216™; kindly donated by Brain Tumour Research Program at Telethon Kids Institute). All cell culture reagents were sourced from Gibco™/Life Technologies unless otherwise stated. COS-7 cells and HEK293T cells were cultured in Roswell Park Memorial Institute medium (RPMI-1640)

or Dulbecco's Modified Eagle Medium (DMEM) respectively each supplemented with 10% foetal calf serum (FCS, Bovogen biologicals), 100UPenicillin/100µgml^{-1} Streptomycin and 2mM GlutaMAX-1. U87-MG cells were cultured in Minimum Essential Media (MEM) supplemented with 10% FCS, 100UPenicillin/100µg ml^{-1} Streptomycin, GlutaMAX-1, 1% MEM non-essential amino acids, and 1% sodium pyruvate. All cell lines were cultured in 25cm^2 CELLSTAR® cell culture flasks (Greiner Bio-one), maintained horizontally in humid conditions at 37°C and 5% CO_2. Cells were passaged regularly using TryPLE Express (Gibco™/Life Technologies) as per manufacturer's recommendation to detach cells from the flask surface.

Freezer stocks were made by resuspending 1x10^6 cells in freezing mix in cryovials for long-term storage in liquid nitrogen. Freezing mix was prepared as follows: 65% appropriate media, 25% FCS and 10% dimethyl sulphoxide (DMSO; Sigma-Aldrich). Prior to freezing, cell lines were tested for mycoplasma contamination using protocol based from Asarnow *et al. (2010)*[72] with the following modifications. Pellet obtained from centrifugation of 0.5ml culture media sample was resuspended in TE buffer. Polymerase chain reaction (PCR) mastermix was made as follows: 1.35x PCR Buffer (Applied Biosystems), 4mM MgCl (Applied Biosystems), 0.67mM dNTPs (Bioline) and 67mM Betaine (Fluka Biochemika) in water. 3µl of sample was added to 7.45µl of PCR mastermix, 0.5µl of the forward and reverse primers and 0.05µl of AmpliTaqGold (Applied Biosystem). PCR amplification was performed using an S-1000™ Thermocycler (Bio-Rad) set to cycling conditions previously optimised for AmpliTaqGold as follows: 10 minute denaturing step at 94^{o}C, 14 cycles of 20 seconds at 94^{o}C, 60 seconds of 63-56 using 0.5^{o}C per cycle decrements and 60 seconds at 72^{o}C. Then 20 cycles of 94^{o}C for 20 seconds, 56^{o}C for 60 seconds and 72^{o}C for 60seconds.

Followed by 5 minutes at 72^{o}C as a final extension step before samples were maintained at 4^{o}C until electrophoresed on a 2% agarose gel for 40 minutes at 115V and visualised as previously described.

2.3 Transfection

When required for a transfection experiment, cells were counted manually using a Neubauer improved bright light counting chamber (Marienfeld) and an inverted microscope. Trypan blue (Sigma-Aldrich) exclusion was used to assess the viability of cells. U87-MG and COS-7/HEK293T cells were seeded into a tissue culture treated Greiner CELLSTAR® 24 well plate (Sigma-Aldrich) lined with a number 1, 13mm round coverslips (ProSciTech); or left unlined for flow cytometry related experiments, at a density of 7.5x10^{5} and 5x10^{4} cells/well respectively. Cells were then incubated in humid conditions at 37°C and 5% CO_2 for 16-18 hours prior to transfection.

Initially cells were transfected using Lipofectamine® LTX with Plus™ Reagent (Invitrogen™/Life Technologies), however after the completion of this product a similar, more efficient product Lipofectamine® 3000 (Invitrogen™/Life Technologies) was used. This alteration was unlikely to have any effect on the results obtained as both products use the same principle. In both cases reagent concentrations were as the manufacturer's protocol. Briefly plasmid DNA (0.5 µg/well) was incubated in humid conditions at 37^{o}C and 5% CO_2 with Plus™ /3000 reagent in OPTI-MEM (Gibco™/Life Technologies) for 15 minutes. The mixture of plasmid DNA and Plus/3000 was then added drop-wise into a mix of diluted Lipofectamine™ in OPTI-MEM so as to promote the formation of miscelles and incubated for a further 30 minutes. During the incubation of the Lipofectamine®-DNA mixture, cells were washed twice with

serum free media (appropriate media + 2mM GlutaMAX-1) then incubated at 37°C and 5% CO_2 for 15 minutes in serum free media. Following which 50µl of the Lipofectamine® mixture was added directly to the serum free media into each well to be transfected and incubated for 3 hours at 37°C and 5% CO_2. The Lipofectamine®/media mixture was then removed from wells and replaced with complete media without antibiotics (appropriate media, 10%FCS, 2mM GlutaMAX-1) and incubated at 37°C and 5% CO_2 for 27 hours.

2.3.1 Transfection efficiency

Multiple transfection efficiency experiments were performed to ensure that: the two Lipofectamine® reagents gave efficient levels of transfection, all three cell types had sufficient levels of transfection and to determine the most efficient length of time for which the cells should be transfected. Three time points were used; 24, 30 and 48 hours, after the addition of Lipofectamine® reagent to the wells.

To determine transfection efficiency cells were transfected as previously described in a 24 well plate without coverslip lining. After the appropriate incubation time, cells were detached using TryPLE Express, harvested into individual round bottom tubes (BD falcon™) and fixed with 2-4% paraformaldehyde (Sigma-Aldrich) for 15 minutes. After incubation cells were washed twice with FACSWash containing 10% Bovine serum albumin (Sigma-Aldrich) and 10% EDTA (Sigma®) in PBS with cells being pelleted by centrifugation for 5 minutes at 201g between each step. Cells were then resuspended in FACSWash and stored at 4^{o}C until analysed using the LSRII flow cytometer (BD). Samples were analysed in duplicate collecting a minimum of 10,000 events per run. GFP2 fluorescence was measured using the 488nm laser. Cells that

had been through the transfection process with no DNA (no DNA control) were used to determine the gating of GFP2 negative (non-transfected) and positive (transfected) cell populations as shown in Figure 2.3. The 30 hour time point was used in all subsequent experiments (see section 3.4.1).

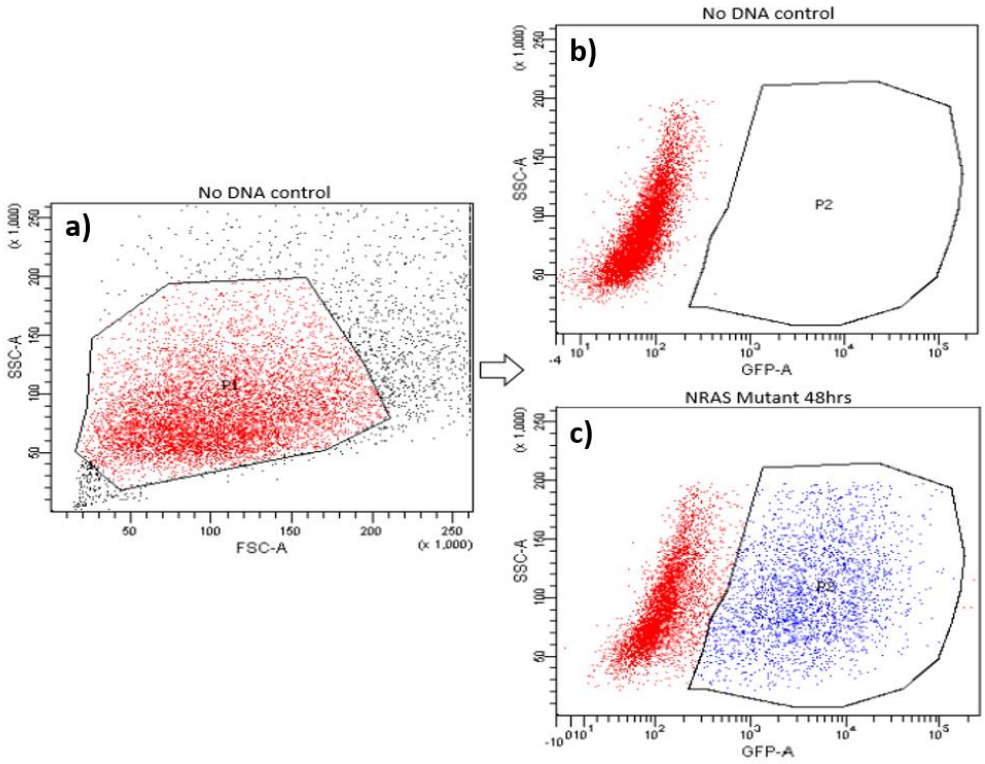


Figure 2.3: Gating strategy for transfection efficiency experiments.
The cells of interest were selected as seen in a. this formed population 1(P1), from this a polygonal gate was drawn excluding all events in the no DNA control as shown in b. this forms population 2(P2). P2 correlates to those cells that are GFP2 positive as seen in c.

2.4 Localisation studies

2.4.1 Bioinformatic analysis

There were two main aims for the bioinformatic analysis section of this study. Firstly to determine the predicted localisation of $NRAS^{WT}$ and if the mutation harboured by the patient was predicted to alter this localisation. This allowed for the organelle markers used in the localisation studies to be selected. Secondly to identify if there was any bioinformatic evidence to suggest that the addition of the GFP2 or 3xFLAG tags, especially the large GFP2 tag, might influence the localisation of $NRAS^{WT}$ and/or $NRAS^{MUT}$ constructs. To address these aims the online localisation tools described below were used. Full sequences for the tagged and untagged $NRAS^{WT}$ and $NRAS^{MUT}$ proteins analysed with these tools are provided in Appendix IV. All tools were used on default settings.

2.4.1.1 cNLS Mapper

cNLS mapper[73] is an online tool which identifies importin α-dependent nuclear localization signals. This tool annotates the input sequence with the localisation signal and its score, this score relates to the level of activity of the NLS. Higher scores indicate stronger NLS activities. A score of greater than 8 indicates exclusive localisation to the nucleus, 7 or 8 indicates partial localisation to the nucleus, between 7 and 3 indicates localisation to both the nucleus and the cytoplasm, and a score below 3 indicates cytoplasmic localisation.[73] By default this tool restricts its analysing to the terminal 60 amino acid regions, the option to search the entire sequence was chosen to ensure no signals were overlooked. This tool was developed based on budding yeast however the importin α/β pathway is highly conserved in all eukaryotes.[73]

2.4.1.2 Hum-mPLoc

Hum-mPLoc 2.0[74] is an adapted version of Hum-PLoc which can determine the predicted localisation of human proteins which may have more than one location in a cell at one time or that move between multiple locations. This tool predicts if the protein of interest localises to any of 14 different subcellular locations including cytoplasm, cytoskeleton, endoplasmic reticulum, endosome, Golgi apparatus, lysosome, and plasma membrane. These predictions are based on a dataset that includes 3,681 protein sequences amassed from the SWISS-PROT[75] database.

2.4.1.3 LocSigDB

LocSigDB[76] is a database of 533 experimentally-validated localisation signals for eight different subcellular locations that have been manually curated from the literature. This tool annotates the input sequence with the localisation signal, the source of this signal, PubMed references and the protein in which the signal was identified.[77]

2.4.1.4 WoLF PSORT

WoLF-PSORT[78] predicts the subcellular localisation of proteins from their amino acid sequence based on amino acid sequence information and a set of rules derived from experimental observations. The animal specific database was used for this prediction. This database used was curated from UniProt[79] and Gene Ontology[80] and contains known sorting signals, amino acid composition and functional motifs for 12,771 animal proteins.[78]

2.4.1.5 BaCelLo

BaCelLo[81] is a prediction tool for the subcellular localization of eukaryotic proteins. This can classify proteins as localising to one of four locations: secretory pathway, cytoplasm, nucleus or mitochondrion based on the whole protein sequence and its N- and C-terminal regions.[81] The animal option was selected when running this tool, therefore the prediction was made using an animal specific database derived from SWISS-PROT.[75]

2.4.2 Immunofluorescence

For localisation studies the media was removed from the wells 30 hours post transfection (Section 2.3) and cells were fixed with 2-4% paraformaldehyde for 15 minutes. The cells were washed thrice with 1x Phosphate Buffered Saline (PBS;

Biochemicals) and subsequently incubated in permeabilisation/blocking solution (PBT) containing 1% BSA and 0.25% Triton-X 100 (ICN Biomedicals) in PBS for 30 minutes at room temperature. Cells were then incubated overnight at $4^{O}C$ with primary antibodies diluted in PBT to their optimal concentration as described below (Section 2.4.2.1).

2.4.2.1 Antibody optimisation

Optimal primary antibody concentration/dilution was determined through an antibody titration experiment in which COS-7 cells were seeded, grown, fixed and permeabilised/blocked. These cells were then incubated with varying concentrations/dilutions of primary antibody in PBT as determined by the manufacturer's recommendations. Information on the primary antibodies used in this study is summarised in Table 2.1. Cells were then incubated with the appropriate secondary antibody (Section 2.4.3), counterstained with Counterstained with NucBlue® fixed cell stain, as described in section 2.4.3, and visualised using the C2+ confocal microscope system (Nikon; Section 2.4.6). From this the lowest concentration or dilution, as indicated, that gave a clear picture with staining specific to the respective sub-cellular location and low levels of background was selected for ongoing experiments. Controls included; no primary antibody, no secondary antibody and no antibodies, as well as cells transfected with the ptagGFP2-N vector alone and those transfected with no-DNA added.

Table 2.1: Information on the source, host, clonality and localisation for each primary antibody used in this study and the concentrations/dilutions used in the antibody optimisation experiment.

Antibody	Reference Number	Company	Clonality	Marker Localisation	Host	Concentration/Dilutions tested			Optimal Concentration/Dilution
						1	2	3	
Glut1 antibody	NB110-39113	Novus Biologicals	Polyclonal	Plasma membrane	Rabbit	2 µg/ml	1 µg/ml	0.5 µg/ml	0.5µg/ml
TGN38 antibody [2F7.1]†	NB300-575	Novus Biologicals	Monoclonal	Trans-Golgi network	Mouse	1:50	1:100	1:200	1:100
GM130/GOLGA2 antibody [EP892Y] †	NB110-57012	Novus Biologicals	Monoclonal	Golgi apparatus	Rabbit	1:100	1:150	1:250*	1:500*
Anti-Calcium Pump pan PMCA ATPase antibody [5F10]†	ab2825	abcam	Monoclonal	Plasma membrane	Mouse	1:200	1:500	1:1000	1:200
Anti-LAMP1 antibody	ab24170	abcam	Polyclonal	Lysosomes	Rabbit	Optimised by Emily Kempin, Genetics and health			0.5µg/ml
Anti-PDI antibody†	ab3672	abcam	Polyclonal	Endoplasmic reticulum	Rabbit				1:800
Anti-EEA1 antibody	Ab2900	abcam	Polyclonal	Early endosome	Rabbit				0.5µg/ml

*† Antibody unpurified thus unable to be quantified, therefore dilutions shown are based off of specific product. *later altered to 1:500 after this dilution was found to produced images of a higher quality*

2.4.3 Secondary antibodies

Following overnight incubation at 4^{o}C with primary antibodies, coverslips were incubated at room temperature in the dark for 1 hour with 1 drop (~25µl) of Alexa Fluor® 594 conjugated goat anti-rabbit IgG (R37117; Molecular Probes®/Life technologies™) or Alexa Fluor® 594 conjugated goat anti-mouse IgG (R37121; Molecular Probes®/Life technologies™) secondary antibodies in 1ml of PBT. Cells were then counter stained for 5 minutes with 1 drop (~25µl) of NucBlue® fixed cell stain ReadyProbes™ reagent (R37606; Molecular Probes®/Life technologies™) in 1ml of PBT. Between all staining steps cells were washed with 1x PBS thrice for 5 minutes each to reduce non-specific binding.

2.4.4 Mitotracker

A selected number of coverslips with transfected U87-MG cells were stained with MitoTracker® Orange CMTMRos (M7510; Molecular Probes®/Life technologies™). This staining protocol was very different from those using the other antibodies as MitoTracker has its own conjugate. Media was removed from the wells 30 hours post transfection (Section 2.3) and cells were then incubated in MitoTracker diluted to 500nM in Hank's Balanced Salt Solution (Sigma-Aldrich) warmed to 37^{o}C for 45 minutes in humid environment at 37^{o}C and 5% CO_2. Following incubation cells were counterstained with 1 drop (~25µl) of NucBlue® Live ReadyProbes® Reagent (R37605; Molecular Probes®/Life technologies™) in 1ml of PBT for 5 minutes. Cells were washed thrice with sterile PBS warmed to 37^{o}C, then fixed in 2-4% paraformaldehyde for 15 minutes and washed again with PBS prior to mounting.

2.4.5 Mounting of coverslips

After the final wash coverslips were mounted onto SUPERFROST®PLUS microscope slides (LabServ). Two different mountants were used; VECTASHIELD® Hard set™ mounting medium (Vector Laboratories) for COS-7 and HEK293T cell experiments and Immu-Mount™ (Thermo Scientific™ Shandon™) for all U87-MG cell experiments. Both mountants were used as per manufacturer's recommendations. VECTASHIELD® and Immu-Mount™ have a refractive index of 1.46 and 1.586 respectively. When using Immu-Mount™ slides were left to dry for 2-4 hrs and then coverslip edges sealed with clear nail varnish (Cutex). All slides were stored in the dark at 4^{O}C until and after visualisation using the C2+ confocal microscope system (Nikon).

2.4.6 Epifluorescent and Confocal imaging

Cells were initially imaged using epifluorescence, a technique which utilises a broad light source passing through specific filter cubes. This technique was used for antibody optimisation. The cubes used were mCherry B NTE (641nm; Semrock) for the Alexa fluor® 594 stained organelles, UV-2A (440nm; Nikon) for the DAPI stained nucleus and GFP-3035D-NTE(510nm; Semrock) for the GFP2 and 3xFLAG constructs. Epifluorescent images were taken with the Nikon DS-U3 driver of the NIS-Elements-C software provided with the C2+ confocal microscope. Images were taken using a Plan Apo 20x DIC M NA0.75 objective (Nikon).

Localisation study images were taken using confocal microscopy. The following lasers were used: 561nm, 488nm and 405nm lasers for Alexa Fluor® 594, FITC/GFP2 and DAPI respectively. Z-Series images were taken sequentially with the Nikon confocal driver of the NIS-Elements-C software. Images were taken using Plan Apo VC 60x NA1.4 oil immersion objective (Nikon) with a step of 0.5µm between images and

pinhole setting of 1.2AU. Laser power, gain and offset were adjusted using control slides to eliminate background fluorescence and were kept constant within an experiment. The top and bottom sections of a z-series were chosen arbitrarily based on what a particular cell or group of cells dictated. Individual images from a stack were chosen based on quality.

All Images were saved in the program's default format of ND2 as well as single images saved as TIFF (converted to 8bit RBG) or JPEG. Images were altered post acquisition using look up tables which were set to show no background staining using control slides and kept constant for all other slides pertaining to the control. Scale bars were added using ImageJ.[82]

2.5 Western blot

2.5.1 Sample preparation

Protein was extracted from HEK293Tand U87-MG cells that were transfected as previously described. The extraction process involved washing the cell monolayer in sterile PBS twice before lysing cells with radioimmunoprecipitation assay (RIPA) buffer, containing 50mM Tris (AMRESCO), 150mM NaCl, 0.5% sodium deoxycholate (Sigma-Aldrich), 0.1% sodium docecyl sulphate (Sigma) and 1% IGEPAL (Sigma), and then incubating at 4^{o}C for 10 minutes before using a 25cm cell scraper (Sarstedt) to collect the cell lysate and debris. This was incubated on ice for 10minutes then centrifuged at 15,700g for 20 minutes at 4^{o}C (Eppendorf 5415 R) to collect the supernatant. Protein concentration was determined by a direct-detect assay-free card and direct detect machine (Millipore) on default settings prior to being stored at -80^{o}C until required.

In preparation for western blotting 5.3µl of NuPAGE® 4x LDS sample buffer (Novex®/Life technologies™) and 1.75µl NuPAGE® 10x sample reducing agent (Novex®/Life technologies™) were added to 50µg of protein samples and the volume made up to 20µl using RPIA buffer. Samples were heated at 70^{o}C for 10 minutes in Thermomixer Comfort (Eppendorf) heating block and quenched on ice for at 2 minutes before loading onto gel.

2.5.2 Sample Loading

NuPAGE 4 –12% Bis-Tris gels (Invitrogen®/Thermo Fisher Scientific) were washed thoroughly with distilled water before being loaded into a gel tank. Running buffers: 200ml of inner chamber running buffer (5% NuPAGE® 20x MES SDS running buffer(Novex®/Life technologies™) and 0.15% NuPAGE® antioxidant (Novex®/Life technologies™) in MilliQ water) and 500ml of outer chamber running buffer (5% MES buffer in MilliQ water), were added ensuring no mixing. Wells were rinsed gently with inner chamber running buffer and 20µl samples loaded into each well. 7µl of Precision Plus Protein™ Kaleidoscope™ (Bio-Rad) was used as a size standard. The gel was run for 70 minutes at 140V in an XCell SureLock™ Mini-Cell Electrophoresis System (Thermo Fisher Scientific). The gel was then placed between the bottom and top sections of a Trans-Blot® Turbo™ mini nitrocellulose transfer pack (Bio-Rad). Protein was transferred to the PVDF membrane using Trans-Blot® Turbo™ transfer system (Bio-Rad). The pre-programmed Bio-Rad protocol for mixed molecular weight were used.

2.5.3 Blocking and staining of membrane

Post transfer, the membrane was incubated for 2 hours on platform mixer (Ratek) at 80rpm in blocking solution (5% skim milk powder in 1x TBST (1x Tris-buffered saline, 0.1% Tween 20); Sigma-Aldrich), then washed thrice for 5 minutes each with 1x TBST at 80rpm. The membrane was cut in half with one half incubated on rotator overnight at 4°C in anti-GFP antibody (G10362;Life technologies™) diluted 1:1000 in antibody diluent (1% skim milk powder in TBST) and the remaining half incubated in diluent alone as the no primary antibody control. Both membrane halves were washed thrice for 5 minutes each with 1x TBST before being incubated on rotator at room temperature for 2 hours in Amersham ECL anti-rabbit IgG, horseradish peroxidase-linked whole Ab (0.19µg/ml; NA934; GE Healthcare). Protein was detected using the Amersham ECL plus chemiluminescence system (GE Healthcare) and visualised using GelDoc™ MG (Bio-Rad) with an exposure time of 180 seconds.

2.6 Functional analysis

2.6.1 Apoptosis assay

To determine the effects of the mutation in NRAS on apoptosis, U87-MG cells transfected with either $NRAS^{MUT}$ or $NRAS^{WT}$ were compared in assays using Andy Fluor™ 594 conjugated Annexin V (Applied BioProbes). This product binds to phosphatidylserine labelling apoptotic cells which can then be detected using flow cytometry.

This product was used as per manufacturer's protocol, with the addition of LIVE/DEAD fixable Near-IR dead cell stain (Molecular Probes®/Life Technologies™) to distinguish dead cells from apoptosing cells. Briefly: $1x10^6$ U87-MG cells were transfected as previously described (Section 2.3). An apoptosis positive control was prepared by

incubating one well of untransfected cells in 5mM sodium azide for 5 minutes after which cells were washed and incubated in appropriate media for a further 24 hours. Post transfection period cells were harvested using TryPLE express, washed with cold PBS and stained with 5 ng/µl LIVE/DEAD stain in the dark at room temperature for 30 minutes. Cells were then centrifuged for 5 minutes at 201g (Eppendorf 5810R) and resuspended in annexin-binding buffer containing 10 mM HEPES (ICN Biomedicals), 140 mM NaCl and 2.5 mM $CaCl_2.2H_2O$ (ICN Biomedicals). To this 0.625 µg/ml Andy Fluor™ 594 Annexin V was added and samples were incubated in the dark at room temperature for a further 15 minutes before being placed on ice until analysis using flow cytometry.

Samples were analysed using the LSRFortessa (BD), collecting a minimum of 50,000 events per run. An initial gate was set using a no-DNA control, these were unstained cells that had been through the transfection process but with no DNA added. The population of GFP2 positive (transfected cells) could then be identified, and within this population the proportions of live, dead or apoptosing cells determined. The following lasers were used: 488nm, 561nm and 640nm for GFP2, Andy Fluor™ 594 Annexin V and Live/Dead stain respectively. Further controls for this experiment included, cells stained with each stain individually, a sample of apoptosis positive cells and cells fixed with 2% paraformaldehyde for 15 minutes as a control for the LIVE/DEAD stain.

2.6.2 Functional motifs

A bioinformatic analysis of the insert sequences was performed to determine if there was the potential for any other functions to be disrupted by the mutation. The tagged and untagged insert sequences were analysed using MOTIF search

(http://www.genome.jp/tools/motif/), which annotates the amino acid sequence of interest with putative functional motifs and binding domains present based on a database of known motifs. This tool was used on the default Pfam[83] database. The same sequences were also analysed using the CD-Search tool[84] which searches the conserved domains database[84] and annotates the input sequence with known protein domains.

3 RESULTS

Previous studies have observed NRAS interacting with multiple subcellular locations including the Golgi apparatus,[50,51,85] plasma membrane,[30] endoplasmic reticulum[50,51] and the inner and outer membranes of the mitochondria.[34,86] There are also studies which suggest that the activation of the RAS isoforms is associated with their localisation and in turn their function. For example the RAS proteins are known to be transiently active on the plasma membrane while constitutively active on the Golgi apparatus.[51] Therefore it was hypothesised that if the localisation of NRAS was altered due to the disruption of a localisation signal or it's normal folding/structure, the activation and function of the protein could be significantly affected leading to disease.

In turn the function of the RAS proteins has been closely linked to their localisation. For example it is known that HRAS is only able to activate the RAF1–ERK pathway while it maintains an association with the endoplasmic reticulum.[51] Similar location-specific functions have been observed from the KRAS isoform as well; KRAS associated with the plasma membrane can induce transformation whereas mitochondria associated KRAS leads to apoptosis.[52] Due to its dependence on localisation and activation, the function of NRAS, which has a major role in many pathways, is likely to be severely disrupted in the instance of altered localisation and or activation.

Preliminary analysis of the novel p.T58I substitution by another group showed it to be a gain-of-function mutation meaning that NRAS was active more so than naturally (M. Tartaglia and G. Baynam, personal communication, 2016). This preliminary data also

identified that this gain-in-function results in increased expression of ERK and AKT. AKT has a major role in regulating cell survival[87] and an increase in active AKT has been shown to protect a number of cell lines, including COS-7[88] and MDCK[89] epithelial cells, from apoptosis induced in a number of ways. Therefore, it was hypothesised that an increase in AKT caused by the mutation would result in lower levels of apoptosis. Since apoptosis is a programmed form of cell death that plays an important role in tissue homeostasis an imbalance in this process is likely to have serious consequences.

The importance and connection of activation, localisation and function means that knowing the effects of the mutation on NRAS localisation and function is vital. A combination of online bioinformatic tools and a literature search were used to predict the effect of the mutation on the protein's localisation. This was then complemented by *in vitro* localisation studies in which cells were transfected with $NRAS^{WT}$ or $NRAS^{MUT}$ constructs tagged with GFP2 or 3xFLAG tags (Section 2.3) and stained with organelle markers (Section 2.4.2) to determine experimentally where the proteins were localising. To identify the mutation's effect on apoptosis an assay to detect and quantify apoptosis was performed. Other potential effects of the mutation on function were predicted using bioinformatics tools to identify the presence of functional motifs and binding domains that could potentially be disrupted by the mutation.

3.1 Aims

This chapter relates to the results for four main aims:

1. To determine if there is evidence to suggest that the p.T58I substitution identified in NRAS disrupts its localisation using *in silico* bioinformatic tools and literature.
2. To determine the subcellular localisation of $NRAS^{WT}$ and $NRAS^{MUT}$, by visualising transiently transfected cells, co-stained with specific organelle markers using confocal and epifluorescent microscopy.
3. To determine if the p.T58I substitution leads to a decrease in apoptosis through the use of an apoptosis assay
4. To determine if there is any evidence to suggest that any other functions of NRAS would be disrupted by the mutation through the use of bioinformatic tools

3.2 Bioinformatic analysis

3.2.1 Translation and solvent accessibility

The translation of the $NRAS^{WT}$ and $NRAS^{MUT}$ inserts was examined using the online ExPASy translate tool.[70] This tool predicted that the mutation would have no effect on the translation of the full protein and that the only difference between the wild type and mutant proteins was the 58^{th} amino acid which was predicted to be an isoleucine in $NRAS^{MUT}$ and a threonine in $NRAS^{WT}$. The tool also predicted that the tagged inserts; *NRAS-GFP2* and *NRAS-3xFLAG,* would be translated into a single uninterrupted protein. The predicted amino acid sequences for the tagged and untagged inserts are provided in Appendix IV.

The predicted structure of these translated sequences was explored using AccPro[71] which annotates each residue as either exposed (e) or buried (-) based on the amount of surface area of the residue that is accessible to a solvent. In all of the sequences that were studied: $NRAS^{WT}$ and $NRAS^{MUT}$ tagged and untagged, the only residue that was predicted to change was the last amino acid prior to remainder of the restriction enzyme site used to sub-clone the insert into expression vectors, this residue was predicted to be buried in the constructs rather than exposed, highlighted in Figure 3.1. The addition of the 3xFLAG and GFP2 tags was not predicted to alter the structure of NRAS. The mutation site, shown in red, was predicted to be buried in both the wild type and mutant protein forms (data not shown).

```
  MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDIL
a)eee-e------ee----------eeeeeeeee-eeee-eee-e-eeee------
b)eee-e------ee----------eeeeeeeee-eeee-eee-e-eeee------

DTAGQEEYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLV
----eeeeee-ee--ee-e--------eeee--e--ee--e---e-eeee------
----eeeeee-ee--ee-e--------eeee--e--ee--e---e-eeee------

GNKCDLPTRTVDTKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLN
--e-eeeee-eeee-ee--ee-e-e------eeee--ee------e--eeeeeeee
--e-eeeee-eeee-ee--ee-e-e------eeee--ee------e--eeeeeeee

SSDDGTQGCMGLPCVVMGSLQSTVPRARDPPVATMSGGEELFAGIVPVLIELDGDVHG
eeeeeee-ee-e---e
eeeeeeeeee-e---eeeeeeeeeeeeeeeeeee-eee-eee-e-e-e-e---ee

HKFSVRGEGEGDADYGKLEIKFICTTGKLPVPWPTLVTTLCYGIQCFARYPEHMKMND

ee-----eee---ee-e-e-e---eeee-------------------e--ee-ee--

FFKSAMPEGYIQERTIQFQDDGKYKTRGEVKFEGDTLVNRIELKGKDFKEDGNILGHK

------ee----e-e-e-ee----e-e-e-e-eee----e-e-e-ee-eee----eee

LEYSFNSHNVYIRPDKANNGLEANFKTRHNIEGGGVQLADHYQTNVPLGDGPVLIPIN

-eee-ee-e-e-ee-eeee--e-e-e-e---eeeee---e-e-e-e-eeeee-ee-ee

HYLSTQTKISKDRNEARDHMVLLESFSACCHTHGMDELYR

-e-e-e-e-eeeee-eee--e-e-e-e-ee-eeeeeeeee
```

Figure 3.1: Solvent accessibility of tagged protein sequence as predicted by AccPro.
a) and b) show residues predicted to be exposed (e) and buried () for the NRAS and GFP2 tagged protein sequences respectively. The mutation site is shown in red and the only residue predicted to have different accessibility is highlighted in blue. The protein sequence for the remainder of the restriction enzyme site, multiple cloning region and the GFP2 tag is highlighted in green. The same result was seen for all constructs (data not shown).

3.2.2 Prediction of subcellular localisation of $NRAS^{WT}$ and $NRAS^{MUT}$

To determine if the p.T58I substitution in NRAS caused disruption of any localisation signals or caused the protein to localise differently the tagged and untagged protein sequences as predicted by the ExPASy translate tool (Section 3.2.1) were examined for localisation signals curated from literature as well as using online bioinformatic tools. These tools provided a predicted localisation based on their databases as well as manually from signals curated from literature.

3.2.2.1 Nuclear localisation

A review of literature revealed that a number of different signals are known to be associated with nuclear localisation, these fall under the classical or non-classical categories.[90-92] Such signals include monopartite signals like PKKKRKV[91] and bipartite signals which consist of two groups of basic amino acids separated by a group of approximately 10 amino acids, for example KR[PAATKKAGQA]KKKK which is one of the most well characterised bipartite signals.[90,92] No such sequences were detected within the wild type or mutant sequences nor in the tagged protein sequences.

In addition to manual analysis of the protein using signals curated from literature, the insert sequences were analysed with the online tool cNLS Mapper[73,93] which highlights importin α-dependant nuclear localisation signals. This analysis showed no predicted signals when using a mid-range cut-off of a score greater than 5, which predicts nuclear specific localisation signals. When the threshold was decreased to a score of 3, which indicates nuclear/cytoplasmic localisation, a number of signals were found Figure 3.2. This tool predicts that there are weak nuclear and cytoplasmic localisation signals within both of the GFP2 (Figure 3.2.b) and 3xFLAG (Figure 3.2.c) tagged NRAS inserts therefore potentially directing NRAS to these locations. While there were also NLS sequences identified in the untagged inserts (Figure 3.2.a), each tag resulted an additional localisation signal. The mutation site, however, is not within any of these localisation signals and was not found to introduce a previously unseen NLS or abolish one (Figure 3.2).

a) NRAS

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDT/IAGQE
EYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVD
TKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQGCMGLPCVVM

b) NRAS-GFP

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDT/IAGQE
EYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVD
TKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQGCMGLPCVVMG
SLQSTVPRARDPPVATMSGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKFIC
TTGKLPVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKYKTRGEV
KFEGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFKTRHNIEGGGV
QLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDHMVLLESFSACCHTHGMDELYR

c) NRAS-3XFLAG

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDT/IAGQE
EYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVD
TKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQGCMGLPCVVMG
SRADYKDHDGDYKDHDIDYKDDDDK

Signal	Score
REQIKRVKD...	3.8
ETSAKTRQG...	4.5
DDGKYKTRG...	3.3
DDGTQGCMG...	3.2

Figure 3.2: Nuclear localisation signals as predicted by cNLS Mapper.
The a) NRAS, b) GFP2 tagged NRAS and c) 3xFLAG tagged NRAS, protein sequences were analysed using cNLS Mapper using a cut-off score of 3.0. This score indicates the strength of the NLS sequence with 3, 4 and 5 indicating nuclear and cytoplasmic and 7-10 indicating varying levels of nuclear localisation. The mutation site is underlined and the GFP2 and 3xFLAG tag sequences are highlighted in green.

3.2.2.2 Other subcellular compartments

While NRAS has been shown in multiple studies to localise to a number of organelles, the effect of the mutation and addition of tags were not previously known, therefore all sequences were analysed using a number of tools. The results of these analyses led to the general consensus that both the wild type and mutant, tagged and untagged proteins would likely localise to the cytoplasm, plasma membrane and/or nucleus. A full summary of the prediction from these tools is shown in Table 3.1. Specifically, the online tools HumPLoc, PSORT II, WoLF PSORT and BaCeLo predict that the proteins are most likely to be cytoplasmic (Table 3.1). HumPLoc and PSORT II also predict a plasma membrane localisation of all sequences analysed (Table 3.1). These tools indicate that there is no difference in localisation between the wild type and mutant forms of NRAS, however WoLF PSORT and PSORTII predict that the addition of the tags could result in altered localisation. WoLF PSORT predicted that the GFP2 tagged constructs potentially localise to the mitochondria, therefore this needs to be carefully interpreted in the localisation studies.

Table 3.1: Predicted localisation of the tagged and untagged NRAS^{WT} and NRAS^{MUT} proteins determine using a number of online bioinformatics tools.

Tool	NRAS^{WT}	NRAS^{MUT}	NRAS^{WT}-GFP2	NRAS^{MUT}-GFP2	NRAS^{WT}-3xFLAG	NRAS^{MUT}-3xFLAG
Hum-mPLoc	- Plasma membrane - Cytoplasm	- Plasma membrane - Cytoplasm	- Plasma membrane - Cytoplasm	- Plasma membrane - Cytoplasm	- Plasma membrane - Cytoplasm	- Plasma membrane - Cytoplasm
LocSigDB	- Lysosome - Endoplasmic reticulum	- Lysosome - Endoplasmic reticulum	- Lysosome/ melanosome, - Lysosome - Endoplasmic reticulum	- Lysosome/ melanosome, - Lysosome - Endoplasmic reticulum	- Lysosome - Endoplasmic reticulum	- Lysosome - Endoplasmic reticulum
WoLF PSORT	- Cytoplasm - Nucleus	- Cytoplasm - Nucleus	- Cytoplasm - Mitochondria	- Cytoplasm, - Nucleus - Mitochondria	- Cytoplasm, - Nucleus	- Cytoplasm - Nucleus
PSORT II	- Cytoplasm - Nucleus - Vesicles of secretory system - Plasma membrane - Golgi apparatus	- Cytoplasm - Vesicles of secretory system - Golgi apparatus - Plasma membrane - Nucleus	- Cytoplasm - Nucleus - Vesicles of secretory system - Endoplasmic reticulum	- Cytoplasm - Nucleus - Vesicles of secretory system	- Cytoplasm - Nucleus - Vesicles of secretory system - Golgi apparatus -Mitochondria	- Cytoplasm - Nucleus - Vesicles of secretory system - Endoplasmic reticulum
BaCelLo	- Cytoplasm	- Cytoplasm	- Cytoplasm	- Cytoplasm	- Cytoplasm	- Cytoplasm

LocSigDB, which provides an annotated report of predicted signals for localisation rather than a location itself, predicted that the proteins would localise to the lysosomes or endoplasmic reticulum. This tool highlighted a number of sequences which are known to be associated with localisation to the lysosome and endoplasmic reticulum as shown in Figure 3.3. These signals do not differ between $NRAS^{WT}$ and $NRAS^{MUT}$, however, there were additional localisation signals identified in the GFP2 tagged constructs. These are shown in red and blue below and indicate that the GFP2 tag may dictate the localisation of NRAS to the melanosome, lysosome and/or endoplasmic reticulum. Full results from LocSigDB are shown in Appendix

MTE**YKLV**VVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVI**DGE**TCLLDILD**T/I**AGQE
E**YSAM**RDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVD
TKQAHELAKSYGIPFIETSAKTRQGVEDAF**YTLV**REIRQYRMKKLNSSDDGTQGCMGLPCVVMG
SLQSTVPRARDPPVATMSGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDAD**YGKL**EIKFIC
TTGKLPVPWPTLVTTLCYGIQCFARYPE**HMK**MNDFFKSAMPEGYIQERTIQFQDDG**KYK**TRGEV
KFEGDTLVNRIEL**KGK**DFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFKTRHNIEGGGV
QLADHYQTNVPLG**DGPVLI**PINHYLSTQTKISKDRNEAR**DHMVLL**ESFSACCHTHGMDELYR

Signal	Coordinate(s)	Localization
[DE]x{3}L[LI]	395-401, 421-427	**Lysosome|melanosome**
Yx{2}[VILFWCM]	3-7, 63-67, 156-160, 244-248	**Lysosome**
Dx{1}E	46-49	**Endoplasmic reticulum**
[HK]x{1}K	282-285, 310-313, 331-334	**Endoplasmic reticulum**

Figure 3.3: Localisation signals identified in the NRAS-GFP2 using LocSigDB.
The protein sequence of the GFP2 tagged NRAS proteins were analysed, there was no difference between the wild type and mutant. The mutation site is in bold and underlined, no localisation signal was found to overlap with this residue. The residues highlighted refer to the localisation signals shown in the same colour below. The signals in red and blue are unique to the GFP2 tagged constructs while the other signals were found in the NRAS and NRAS-3xFLAG sequences (data not shown). The GFP2 tag is highlighted in green.

3.2.2.3 Post-translational modifications

A review of the literature found that all three RAS isoforms undergo post-translational modifications to their hyervariable region (HVR), a process which is involved in their ability to stably interact with membranes.[30] The first step is the farnesylation of the cysteine of the C terminal CAXX motif,[30] highlighted in green in Figure 3.4 This is followed by further methylation which allows the RAS proteins to interact with the endoplasmic reticulum.[30] This binding is then supported by a second motif. For NRAS and HRAS the second motif consists of one or two amino acids respectively shown in red in Figure 3.4 which are palmitoylated allowing plasma membrane localisation.[30] NRAS then has a third motif of basic/hydrophobic amino acids shown in yellow in Figure 3.4 which is required for stable plasma membrane localisation.[30]

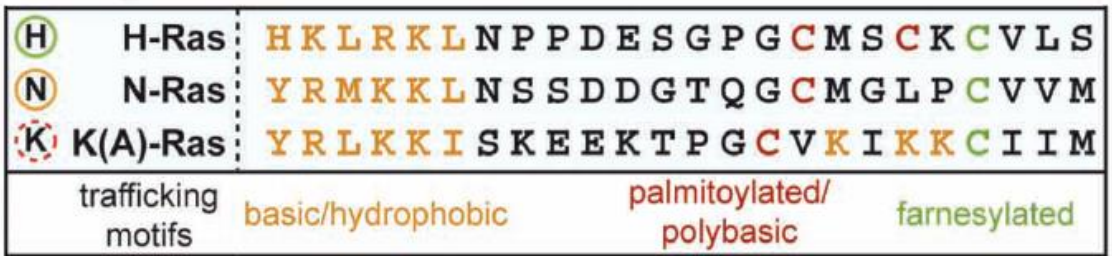


Figure 3.4: Diagram of the HVR of RAS modified from Henis et al (2010)[30]
This hypervariable region contains three motifs which are modified post translation to allow that proteins to localisation to the necessary sub-cellular locations.

3.3 Quality Control

Prior to localisation studies, the constructs were studied to ensure that the insert was present and that the NRAS-GFP2 protein was expressed. This was done to ensure that the constructs were designed and synthesised as expected and to validate the quality of the results obtained.

3.3.1 Restriction digest

To confirm that the gene insert was sub-cloned into the vectors and that the sequence was approximately the correct length two restriction digest experiments were performed. The first was a single digest using *Nhe*I to linearise the constructs and allow the larger size of the NRAS constructs compared to the vector alone to be observed. This confirms the presence of the insert within each construct and its absence from the vector alone (Figure 3.5). The uncut controls in lanes 3, 5 and 7 of Figure 3.5 confirm that the cut samples were appropriately cleaved as the uncut circular DNA migrated further through the gel as expected of supercoiled DNA.

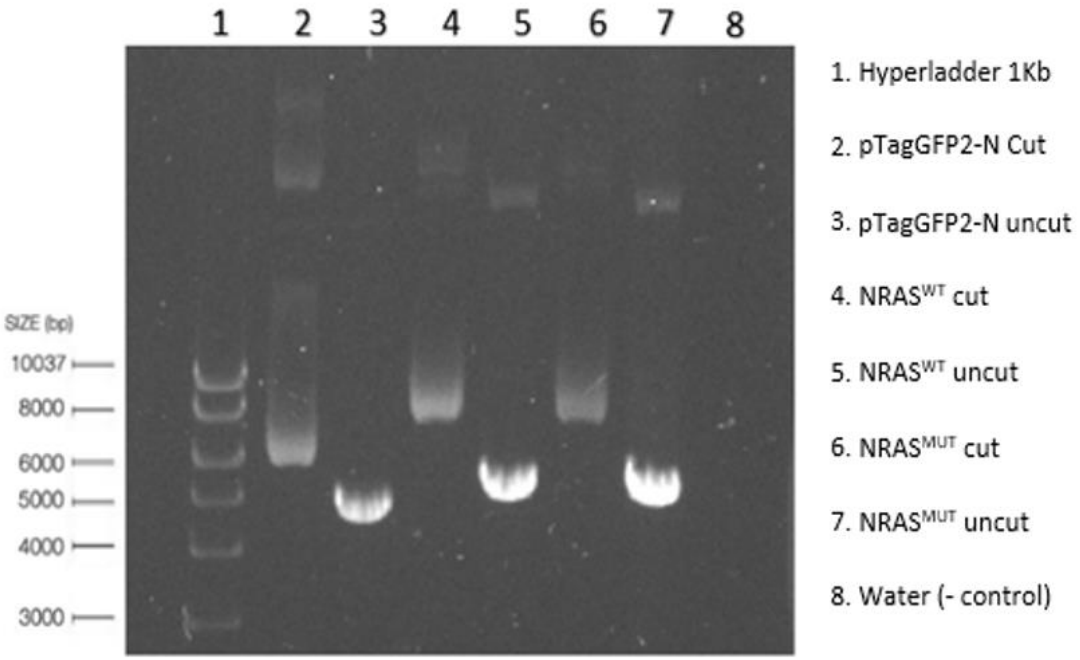


Figure 3.5: Restriction digestion of GFP2 tagged constructs and pTagGFP2-N vector
These plasmids were linearised with NheI and electrophoresed alongside their uncut counterparts. Note that the larger size of the NRAS constructs than the vector alone which confirms the presence of the insert.

A second restriction digest was then performed on the GFP2 tagged constructs to determine the size of the insert itself and to confirm that it was indeed sub-cloned into the expression vector. This restriction digest involved the sequential cutting of the plasmid DNA with first *Hin*dIII then *Pst*I (Section 2.1.4) so as to result in two

fragments of DNA: the NRAS construct which is 579bp, and the remainder of the plasmid which is 4717bp. This digest showed that there were two fragments of the expected length in both the $NRAS^{WT}$ and $NRAS^{MUT}$ samples. The excised insert in each of these samples is indicated by an arrow in Figure 3.6, these bands are weak due to a deficit of sample loaded into the gel. The band correlating to the insert was not present in the pTagGFP2-N sample as expected (Figure 3.6). An uncut pTagGFP2-N vector sample was used as a control for the restriction enzymes to ensure that they worked, this sample showed the expected supercoiled DNA band seen at approximately 4000bp however there was also an additional band at approximately 600bp which was not expected. Since this band was not seen in the cut vector control it is possible that it is the result of contamination with and partial digest of the vector by the restriction enzymes. It is possible this is a result of non-standard conditions which can cause restriction enzymes to cleave at non-conocal sites, a phenomenon known as star activity.

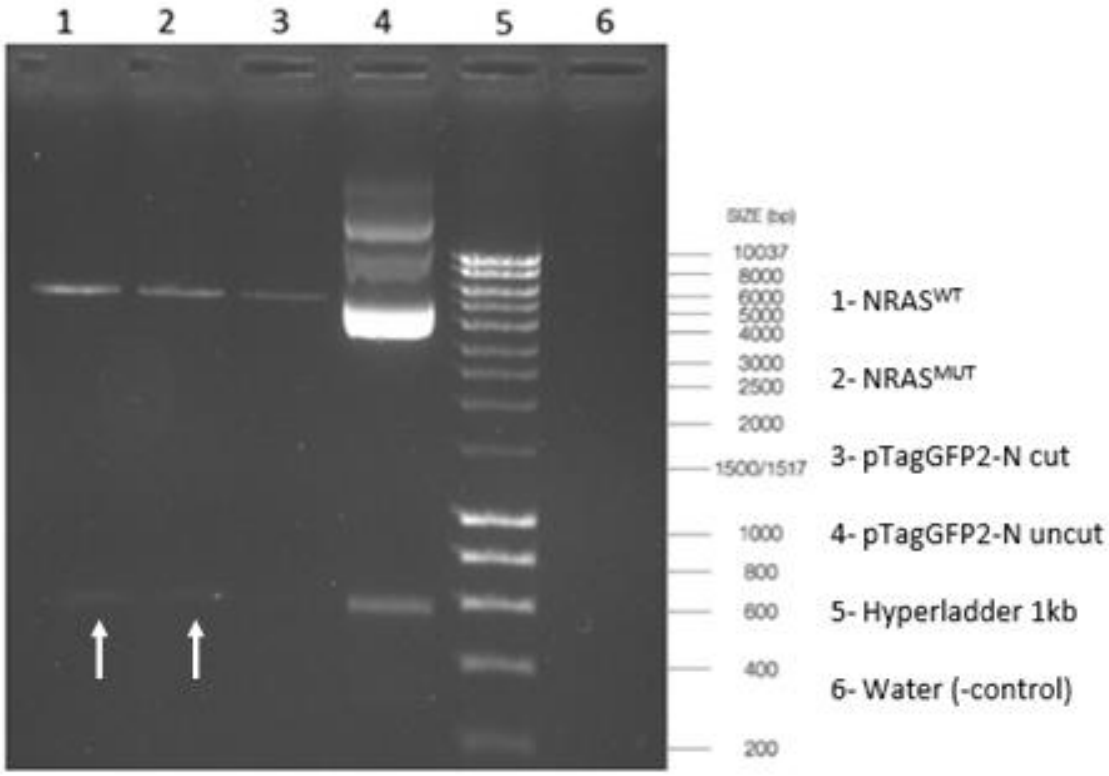


Figure 3.6: Sequential restriction digest on GFP2 tagged constructs.
These plasmids were digested using HindIII and PstI to excise the insert from the vector. The arrows indicate bands at approximately 579bp which correlate to the bands expected of the excised NRAS insert.

3.3.2 Western Blot

To further confirm that *NRAS*^{WT} and *NRAS*^{MUT} were not only within the plasmid but also that the NRAS-GFP2 protein was being expressed, a western blot was performed. This utilised protein extracted from U87-MG cells transfected with GFP2 tagged constructs, the pTagGFP2-N vector or left untransfected for the negative control. Both the NRAS^{WT}-GFP2 and NRAS^{MUT}-GFP2 proteins are expressed, as shown by the bands in lanes 3 and 4 respectively, which are ~49.83 kilodaltons (kDa) matching the predicted size of the NRAS and GFP2 proteins (Figure 3.7.a). The pTagGFP2-N vector produced a band which is approximately 28.6kDa in size, this correlates to the size of the GFP2 protein. The primary only control showed no bands as expected (Figure 3.7.b)

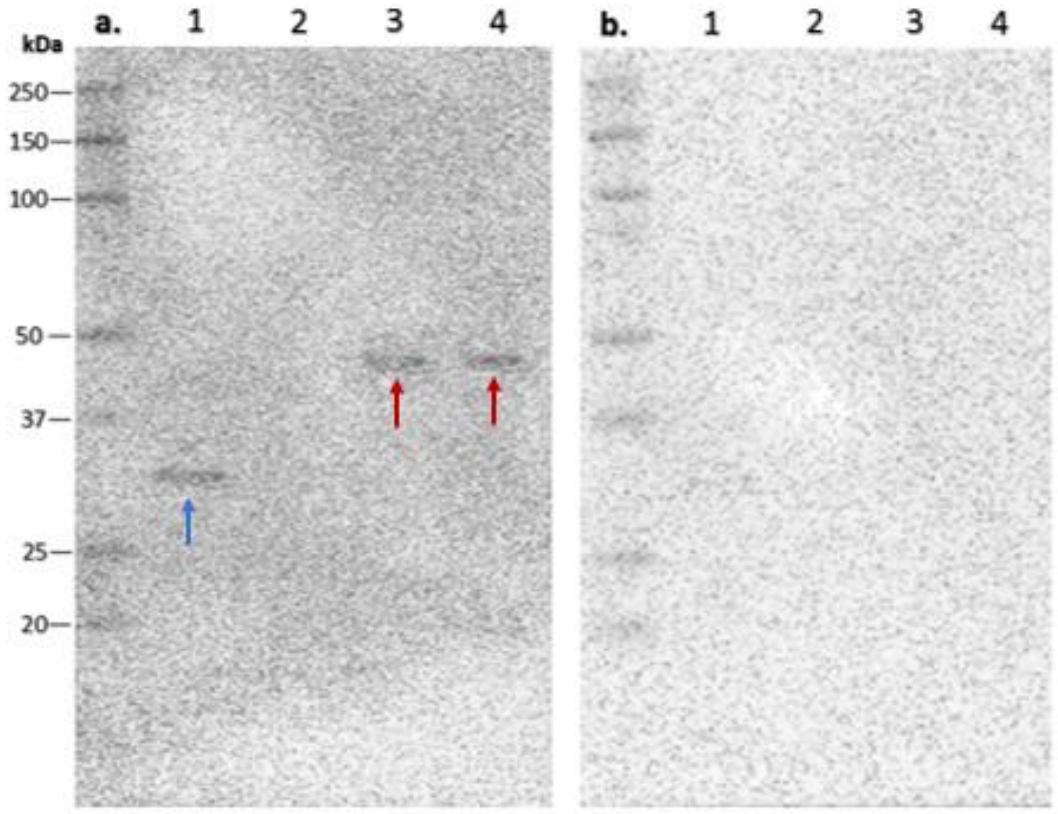


Figure 3.7: Image of Western blot membranes taken at 180seconds of exposure.
Bands of approximately ~49.83 kDa were observed in the the NRAS^{WT}-GFP2 and NRAS^{MUT}-GFP2 samples (red arrows) and 28.6kDa in the vector only sample(blue arrow). a) membrane stained with both the anti-GFP primary antibody and the appropriate secondary antibody. b) No primary control which was incubated in the secondary only. Lane details are as follows: 1. pTagGFP2-N vector, 2. No DNA(negative control), 3. NRAS^{WT} and 4. NRAS^{MUT}, with kaleidoscope ladder (Bio-Rad) used as a protein size standard.

3.4 Optimisation of Methods

Optimisation of the protocols to be used for transfection studies was undertaken. This included optimising the transfection incubation period, the cell line used, antibody concentrations and antibodies used.

3.4.1 Transfection efficiency

The efficiency of transfection using two Lipofectamine® reagents in HEK293T and U87-MG cells was determined at 24, 30 and 48 hours. Cells were transfected and then analysed using flow cytometry (Section 2.3.1). The time points originally chosen were 24 and 48 hours as a standard length of time for transfection. The results of this

experiment, which was carried out in HEK293T cells with Lipofectamine® LTX with plus reagent, found that there was no noteworthy difference between the two time points with approximately 38% transfection at 24 hours and 32% at 48 hours. This lead to the decision that either time point was acceptable (Table 3.2). A third time point of 30 hours was subsequently trialled; this was chosen to match work that had previously been done on the same *NRAS* mutation. The 30 hour transfection optimisation was similarly performed in HEK293T cells with Lipofectamine® LTX and plus with transfection rates similar to those at 24 and 48 hours observed (Table 3.2). Since there was very little difference between the three times points it was determined that a 30 hour incubation would be used for all experiments to provide conditions comparable to those used by collaborators also working on this mutation (M. Tartaglia and G. Baynam, personal communication, 2016).

An additional efficiency experiment was performed when a new Lipofectamine® reagent (3000) and cell line (U87-MG) were being used. This experiment used the same 30 hour incubation period in U87-MG cells transfected using Lipofectamine® 3000. The results of this are similar to those of the previous efficiency experiments (Table 3.2).

Table 3.2: Transfection efficiency results.
HEK293T and U87-MG cells were transfected with either Lipofectamine® LTX with plus reagent or Lipofectamine® 3000 and incubated for 24, 30 or 48 hours then analysed using flow cytometry.

Time (hr)	Cell type	Lipofectimine Reagent	Plasmid DNA	Average rate of Transfection (%)
24	HEK293T	LTX with Plus	$NRAS^{WT}$	38.7
24	HEK293T	LTX with Plus	$NRAS^{MUT}$	38.6
48	HEK293T	LTX with Plus	$NRAS^{WT}$	30.5
48	HEK293T	LTX with Plus	$NRAS^{MUT}$	33.6
30	HEK293T	LTX with Plus	$NRAS^{WT}$	39.4
30	HEK293T	LTX with Plus	$NRAS^{MUT}$	40.2
30	U87	3000	$NRAS^{WT}$	41.5
30	U87	3000	$NRAS^{MUT}$	43.7

3.4.2 Antibody optimisation

Optimisation of the dilution/concentration of primary antibody to be used is essential in order to get good quality images. The dilutions tested (Section 2.4.2.1) were based on the recommended dilution provided by the manufacturer. The aim of this was to determine the lowest concentration that would provide clear images with little background and specific staining of the organelle of interest. The results of this optimisation for each antibody are shown in Figure 3.8. The dilution of GOLGA2 was later changed to 1:500 as the images produced using a dilution of 1:250 were too bright causing images to look overexposed. All images shown in this study used primary antibodies as follows; 0.5ug/ml for Glut1 and dilutions of 1:500, 1:100 and 1:200 for GOLGA2, TGN38 and Anti-calcium pump pan PMCA ATPase (PMCA) respectively. As seen in Figure 3.8 the TGN38 antibody gave very low levels of fluorescence, the 1:100 dilution was chosen for this antibody rather than the 1:50 due to the amount of antibody available for the project. The antibody also had a staining pattern broader than was expected with a high level of background.

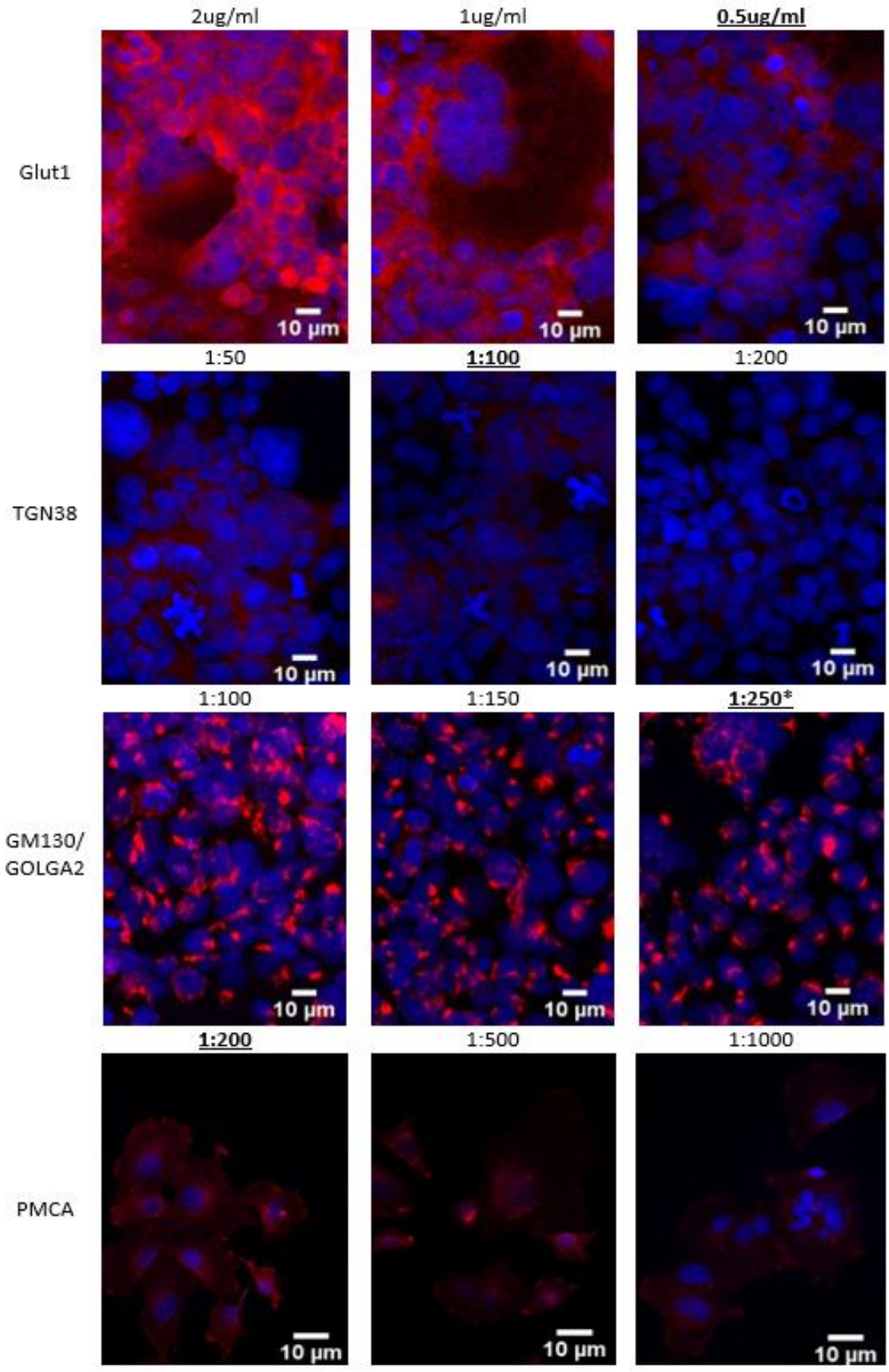


Figure 3.8: Epifluorescent images from antibody optimisation.
*Dilutions/concentrations used in a titration experiment to determine the optimal amount of primary antibody organelle markers to use. Those chosen are in bold and underlined. *Dilution of GOLGA2 was subsequently altered to 1:500*

3.4.3 HEK293T cells

Localisation studies were initially performed in HEK293T cells as they are highly transfectable. These cells posed many issues including their semi adherent nature and their tendency to clump which resulted in the loss of many cells during the transfection and immunofluorescence staining process. The images obtained from these cells were also not of the desired quality. Cell clumping made imaging a single cell difficult and the close proximity of the cells made identifying organelle specific staining impossible as illustrated in Figure 3.9.a. Low levels of fluorescence and widespread background fluorescence lead to pixelated images and low resolving power. To overcome these issues a number of different concentrations of cells on coated and uncoated cover slips were used, as well as careful washing during the transfection and immunofluorescence processes (Sections 2.3 and 2.4.2). Background fluorescence was minimised during imaging through the use of the imaging parameters by increasing the offset and decreasing the gain (Section 2.4.6). Despite these attempts the images acquired were of insufficient quality to properly decipher specifically where the protein was localising. These images did however show a staining pattern similar to that seen in U87-MG cells which were used for all subsequent localisation and functional studies.

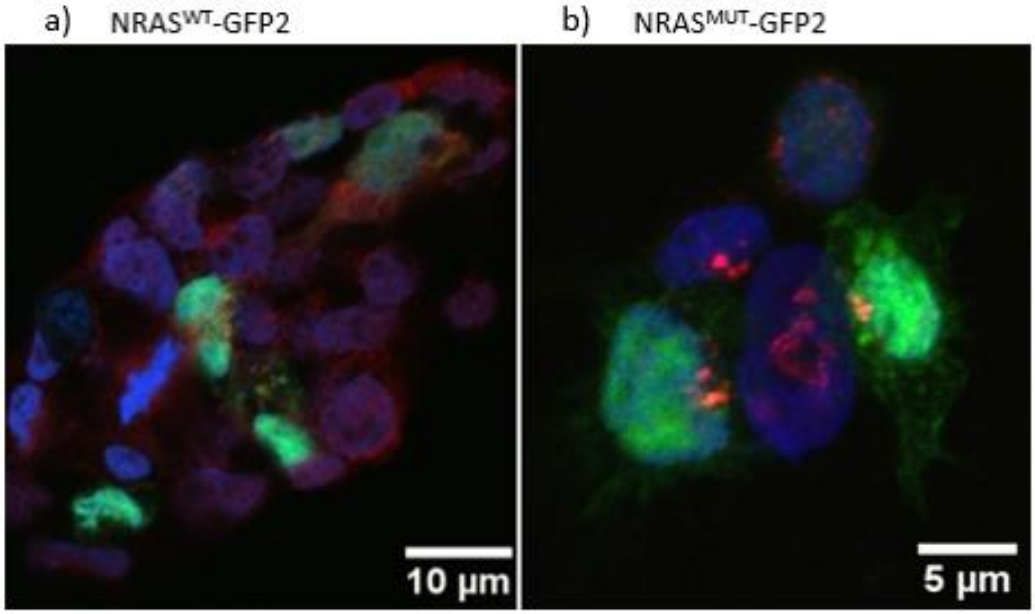


Figure 3.9: Confocal images of HEK293T cells.
These cells are transfected with GFP2 tagged constructs and stained with a) PDI (endoplasmic reticulum marker), and b) GM130/GOLGA2 (Golgi apparatus marker)

Given the quality issues described above when using HEK293T cells the alternative U87-MG cell type was selected for ongoing transfection studies. As well as the quality issues with HEK293T there is an additional reason for selecting an alternative cell type, while NRAS is known to be expressed in most tissues[48] and is likely to have an active role in cell survival growth and apoptosis in these, there are no documented patient symptoms that affect the kidney directly. To take account of this, U87-MG cells were also used. While this cell line has recently been discovered not to derive from the cell line originally described they are human glioblastoma cells[94] and thus provides a different cell type in which to study the effects of the mutation.

3.5 Localisation of NRAS^{WT} and NRAS^{MUT}

The localisation of NRAS^{WT} and NRAS^{MUT} was studied to determine if the mutation resulted in altered localisation of NRAS and thus be closer to explaining the role of this mutation in the patient's disease. The *NRAS^{WT}* and *NRAS^{MUT}* inserts were designed to be sub-cloned into two vectors to provide either the GFP2 or 3xFLAG tag as described

in section 2.1. The GFP2 tag allows the protein to be easily identified using fluorescent microscopy or flow cytometry, and thus allows the proteins location to be easily determined. However, the GFP2 protein is large and could affect the localisation of NRAS. As a control to check that this was not the case, the constructs were also tagged with the smaller 3xFLAG tag which was less likely to alter the protein localisation. Since the 3xFLAG tag does not fluoresce, an anti-flag antibody conjugated with a fluorescent tag was used to visualise the transfected protein. Cells were also labelled with organelle markers to determine whether the transfected proteins co-localise with specific organelles.

Markers for the plasma membrane (Glut1), for the Golgi apparatus (GM130/GOLGA2) and for the *trans*-Golgi network (TGN38) were selected as strong candidates for co-localisation with NRAS based on the known localisation of NRAS. The results of the bioinformatic studies suggested that NRAS could also be targeted the mitochondria, ER and vesicles of the secretory system therefore several additional markers; PDI, EEA1, MitoTracker® orange and LAMP1, which stain the endoplasmic reticulum, early endosomes, mitochondria and lysosomes, respectively, were selected to be used. The Glut1 marker did not display staining specific to the plasma membrane as expected, instead showing strong cytoplasmic staining (Figure 3.10). Therefore, an additional plasma membrane marker (PMCA) was selected, as seen in Figure 3.10*.b* this marker showed the expected membrane specific staining pattern. Due to the inaccuracy of the Glut1 staining these images have been excluded.

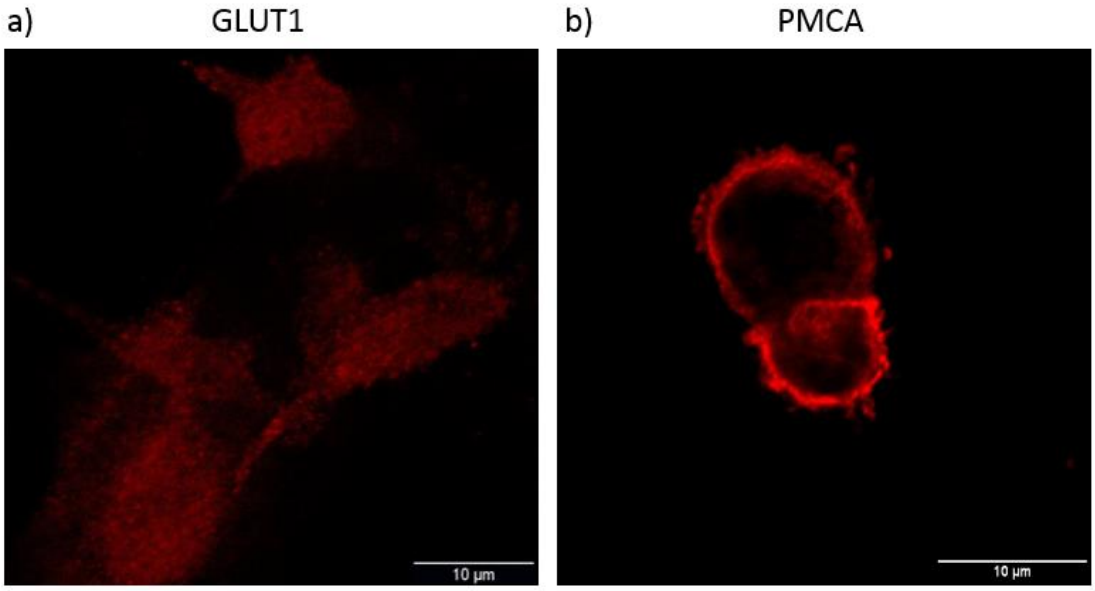


Figure 3.10: Comparison of the staining pattern of a) GLUT1 and b) PMCA.
Both of these antibodies are plasma membrane markers however Glut1 appears to have a cytoplasmic staining pattern (a) as opposed to the expected membrane specific staining as seen in b)

U87-MG Cells transfected with the pTagGFP2-N vector were used as a control for the localisation of the GFP2 tagged NRAS proteins. This was to confirm that the GFP2 tag was not altering the localisation of the NRAS proteins due to its large size. As presented in Figure 3.11 the GFP2 protein appears to localise throughout the cell, though there is an area of condensed localisation to the nucleus. Other controls included a; no primary antibody control for non-specific binding of the secondary antibodies, no secondary antibody control for auto fluorescence of the primary antibodies, no antibody control for cell auto-fluorescence and a no-DNA transfection control for GFP2 auto-fluorescence in cells (data not shown).

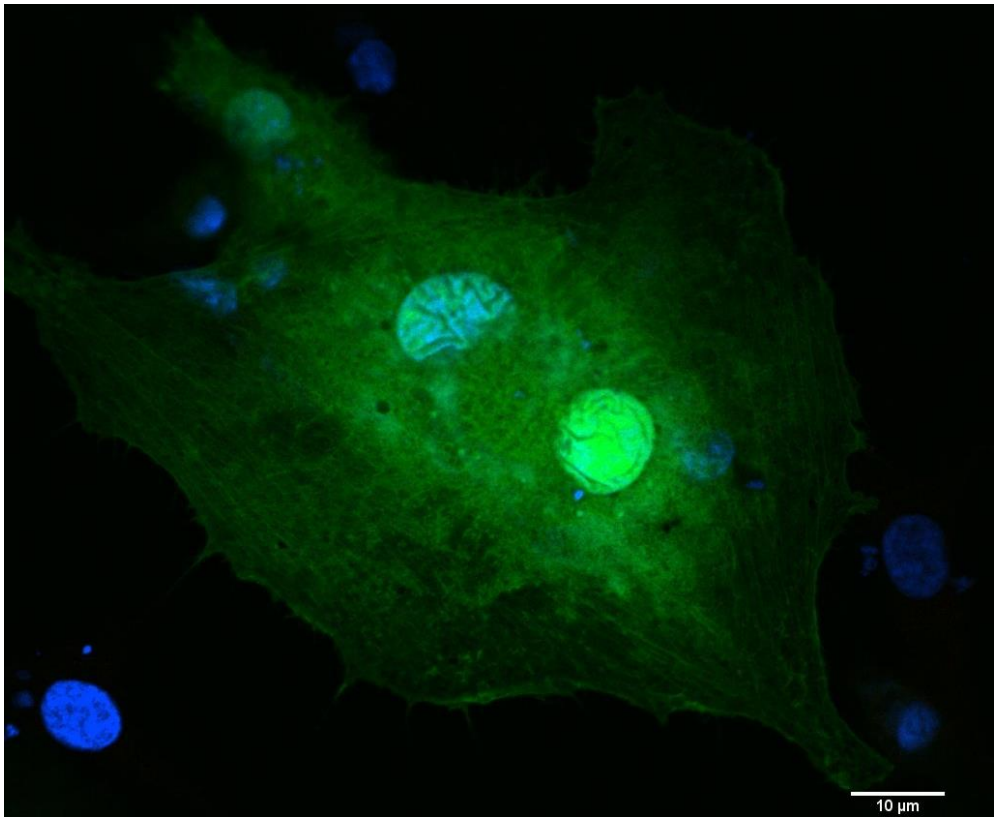


Figure 3.11: Confocal image of U87-MG cells transfected with pTagGFP2-N vector.
This was used as a control for the localisation of the GFP2 tagged NRAS proteins. The GFP2 protein appears to be localising throughout the cell.

Cells transfected with constructs and stained with organelle markers were visualised using confocal microscopy to take an image at a single point within the cell and thus determine if there was any co-localisation of the NRAS proteins with the markers used. In each of the results figures to follow, a series of four is presented: the nucleus stained with DAPI (blue), NRAS (green), the organelle of interest (red) and an overlay of all three images.

3.5.1 Golgi apparatus

NRAS has been observed to be constitutively active on the Golgi apparatus[51,53] and is known to travel bi-directionally between the plasma membrane and Golgi apparatus.[49] Cells transfected with either the GFP2 tagged (Figure 3.12.a) or the 3xFLAG tagged constructs (Figure 3.12.b) and stained with GM130/GOLGA2 show no evidence of localisation of $NRAS^{WT}$ to the Golgi apparatus. There does however appear to be co-localisation of $NRAS^{MUT}$-GFP2 with GOLGA2 as indicated by the arrow in Figure 3.12.a. This staining pattern is not replicated with the 3xFLAG tagged construct although the image quality is too poor to make any firm conclusion.

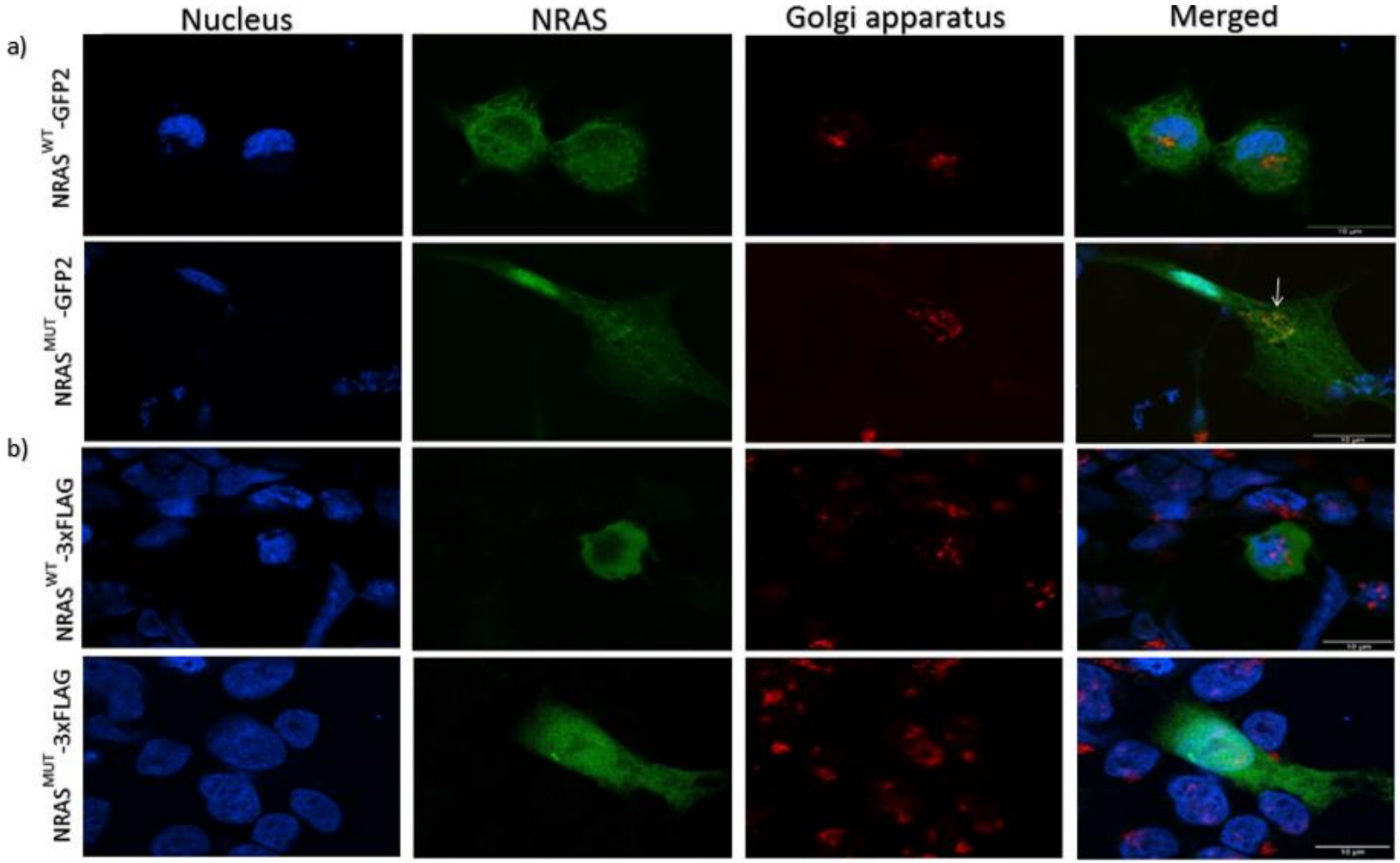


Figure 3.12: GM130/GOLGA2 (Golgi apparatus marker) stained U87-MG cells

Transfected with a) GFP2 and b).3xFLAG $NRAS^{WT}$ and $NRAS^{MUT}$ constructs. There appears to be co-localisation between $NRAS^{MUT}$-GFP and GOLGA2 indicated by the white arrow. There is no evidence of co-localisation in the remaining images. Scale bars are 10µm

3.5.2 Plasma membrane

The plasma membrane was another location that NRAS has been observed to localise to.[30] NRAS has been shown to be transiently active at the plasma membrane.[51,53] Results presented here (Figure 3.13.a) show that neither the wild type nor the mutant protein appear to be solely localising to the plasma membrane. There is no evidence that there is any co-localisation between PMCA and $NRAS^{WT}$, However there does appear to be partial co-localisation between PMCA and $NRAS^{MUT}$ (Figure 3.13.a)

3.5.3 Mitochondria

NRAS has been observed on both the inner and outer membranes of the mitochondria.[34,86] Therefore it was hypothesised that, if this localisation was disrupted in the mutant, it could lead to altered function of NRAS. No evidence was found for co-localisation of either the wild type or the mutant protein with the mitochondrial marker MitoTracker® (Figure 3.13.b)

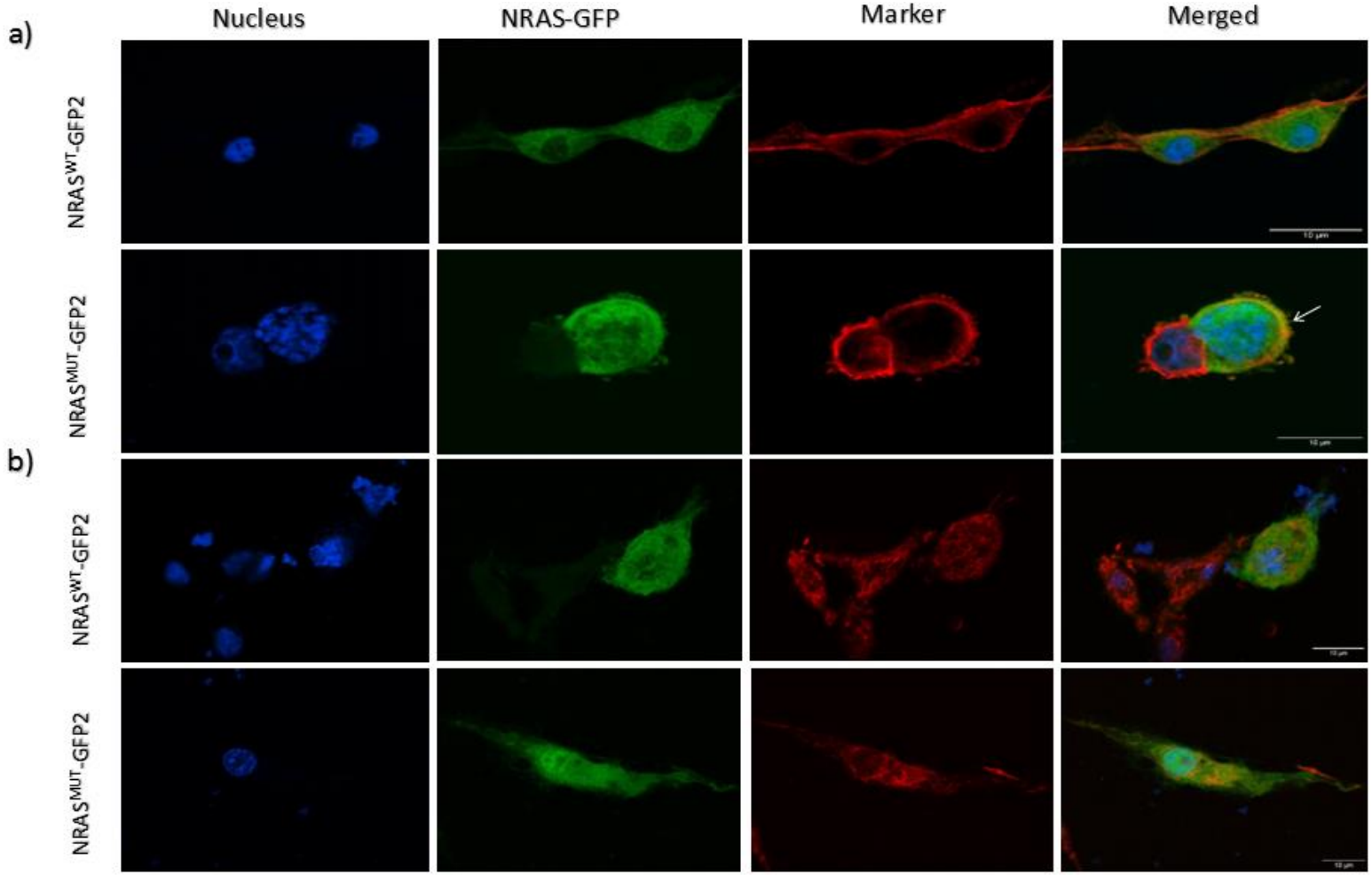


Figure 3.13: U87-MG Cells stained with a) PMCA (plasma membrane marker) and b) MitoTracker® Orange (mitochondria marker). *The arrow highlights an area of co-localisation between $NRAS^{MUT}$ and PMCA. There is no evidence of co-localisation between $NRAS^{WT}$ and PMCA nor either construct with MitoTracker®. Scale bars are 10µm*

3.5.4 Trans-Golgi network

NRAS was found to be one of a number of proteins that were transported from the *trans*-Golgi network to the cell surface.[50] To determine if the mutation disrupts this transport, cells stained with TGN38 were visualised. This antibody also showed non-specific binding and high background levels even after optimisation (Figure 3.14.a). While neither protein appears to co-localise with the TGN38, the marker itself does not show the expected localisation. Hence, firm conclusions cannot be drawn from the use of this marker.

3.5.5 Endoplasmic Reticulum

NRAS has been observed on the endoplasmic reticulum.[50,51] This organelle was of particular interest as the endoplasmic reticulum is involved in folding and transport of proteins. If the mutation caused the disruption of localisation signals it could be retained by the endoplasmic reticulum for degradation. Results presented here show no evidence of co-localisation between $NRAS^{WT}$ and the endoplasmic reticulum marker PDI (Figure 3.14.b). There does, however, appear to be partial co-localisation between $NRAS^{MUT}$ and PDI as evident by the orange staining seen in the merged image (Figure 3.14.b), suggesting that the mutant protein may be stuck in the endoplasmic reticulum.

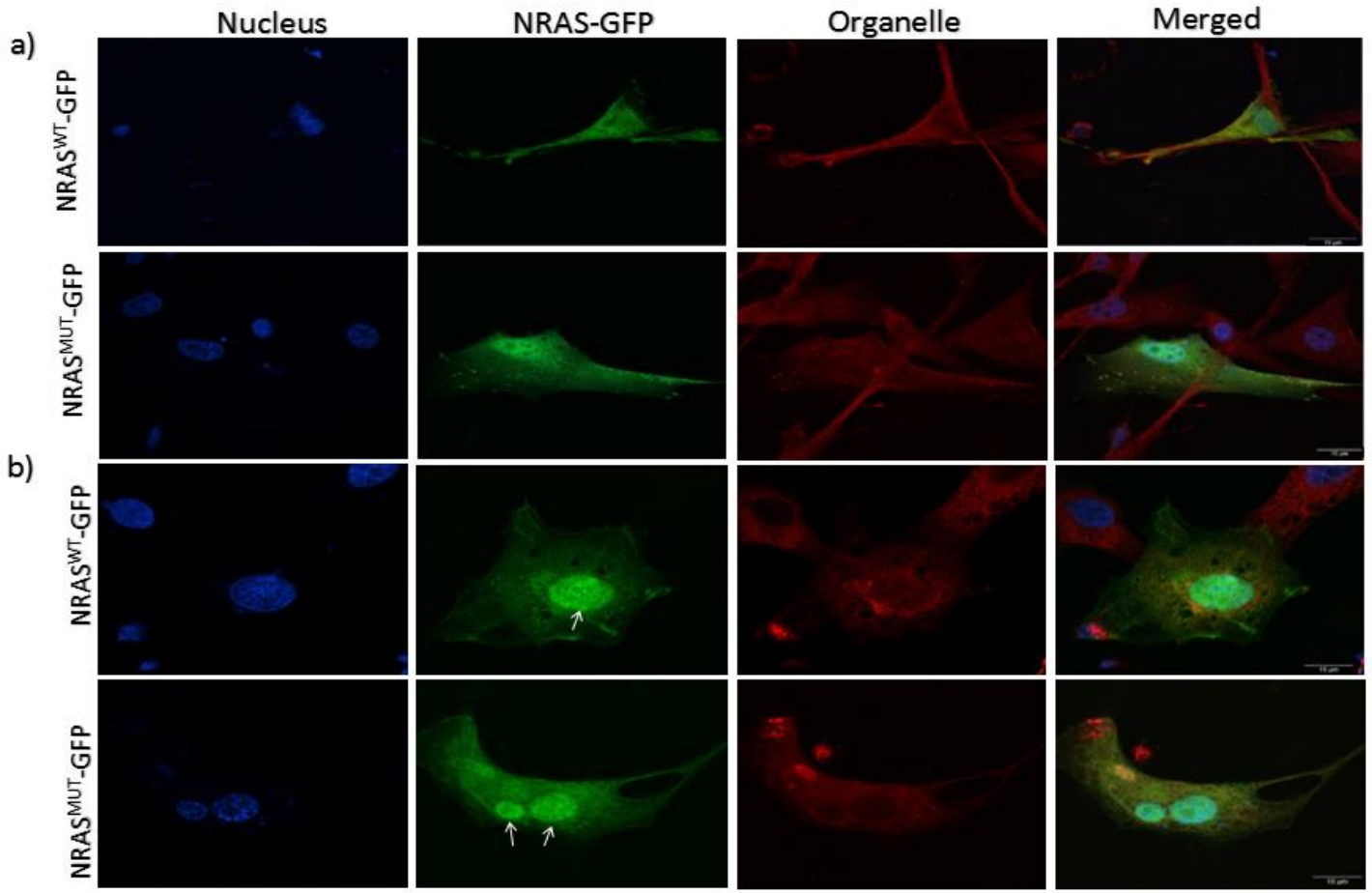


Figure 3.14: U87-MG Cells stained with a) TGN38 (trans-Golgi network marker) and b) PDI (endoplasmic reticulum marker).
The TGN38 antibody shows a staining pattern which is more cytoplasmic and less vesicular than expected. There is no evidence of co-localisation between either construct with TGN38. There appears to be partial co-localisation between NRAS^{MUT} and PDI however there is no ER localisation seen in wild type transfected cells. An example of strong nuclear localisation is indicated by arrows. Scale bars are 10µm

3.5.6 Lysosomes and Early endosomes

The bioinformatic tool LocSigDB predicted that there was a lysosome localisation signal within the NRAS sequence (Section 3.2.2.2). In addition to this the lysosomes and early endosomes are involved in the recycling and transportation of proteins from and to the Golgi apparatus, *trans*-Golgi network and plasma membrane.[95,96] Results obtained here (Figure 3.15) show no evidence of co-localisation of $NRAS^{WT}$ or $NRAS^{MUT}$ to either lysosomes (LAMP1: Figure 3.15.a) or to early endosomes (EEA1: Figure 3.15.b).

3.5.7 Nucleus

A number of tool including cNLS Mapper and PSORTII indicated that there was the potential for the NRAS proteins to localise to the nucleus (Section 3.2.2.1). All cells were stained with the nuclear stain DAPI which allowed for the localisation of NRAS to the nucleus to be observed. There was a large amount of variability in the localisation of the NRAS proteins to the nucleus as seen in Figures 3.12-3.15, however $NRAS^{MUT}$ was consistently observed to co-localise with DAPI. An example of this localisation is highlighted with arrows in Figure 3.14.b. Nuclear localisation was rarely seen in cells containing the wild type protein the only example presented here is shown in Figure 3.14.b.

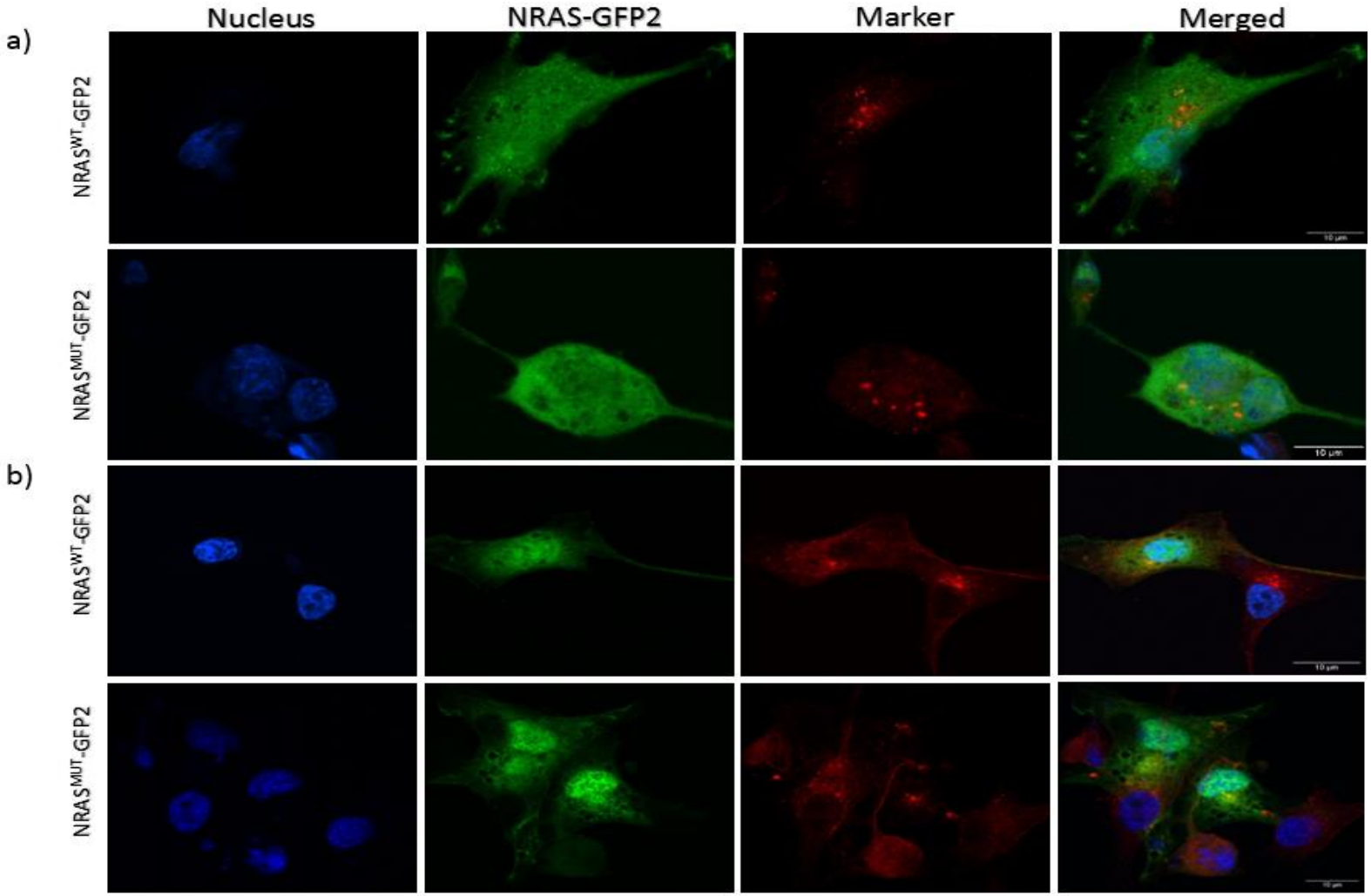


Figure 3.15: U87-MG cells stained with a) LAMP1 (lysosome marker) and b) EEA1 (early endosome marker).
Neither protein shows evidence of co-localisation with these markers. Scale bars are 10µm

3.6 Functional studies:

3.6.1 Apoptosis assay

An Andy Fluor 594 conjugated Annexin V product was used to detect and quantify the levels of apoptosis. Annexin V is one of a family of phospholipid binding proteins which have a high affinity for phosphatidylserine (PS).[97] PS, which is located on the inner side of the plasma membrane in healthy cells, is translocated to the extracellular portions of the membrane during apoptosis.[97] Therefore through the use of fluorescently labelled Annexin V paired with a cell viability stain, to discount those cells which are dead, apoptosing cells were detected and quantified using flow cytometry.

The gating strategy for the apoptosis assay is shown in Figure 3.16. This was achieved by selecting the population of interest from a no-DNA control sample (Section 2.6.1) as shown in Figure 3.16.a, this forms population 1(P1). From P1 the populations of transfected and untransfected cells were selected (Figure 3.16.b). Transfected cells from the test population, untransfected cells acted as an internal control. As shown in Figure 3.16.c two samples of untransfected cells one stained with the LIVE/DEAD stain alone the other with Annexin V alone were used to choose the appropriate gating. In addition to these the following positive controls were used: fixed/dead cells (data not shown) and cells in which apoptosis had been induced though incubation with sodium azide (data not shown). Figure 3.16.c shows that there are a small proportion of cells in an untransfected population that are dead or apoptosing as illustrated by the cell populations observed in Q2-3/Q4-3 and Q1-3 respectively.

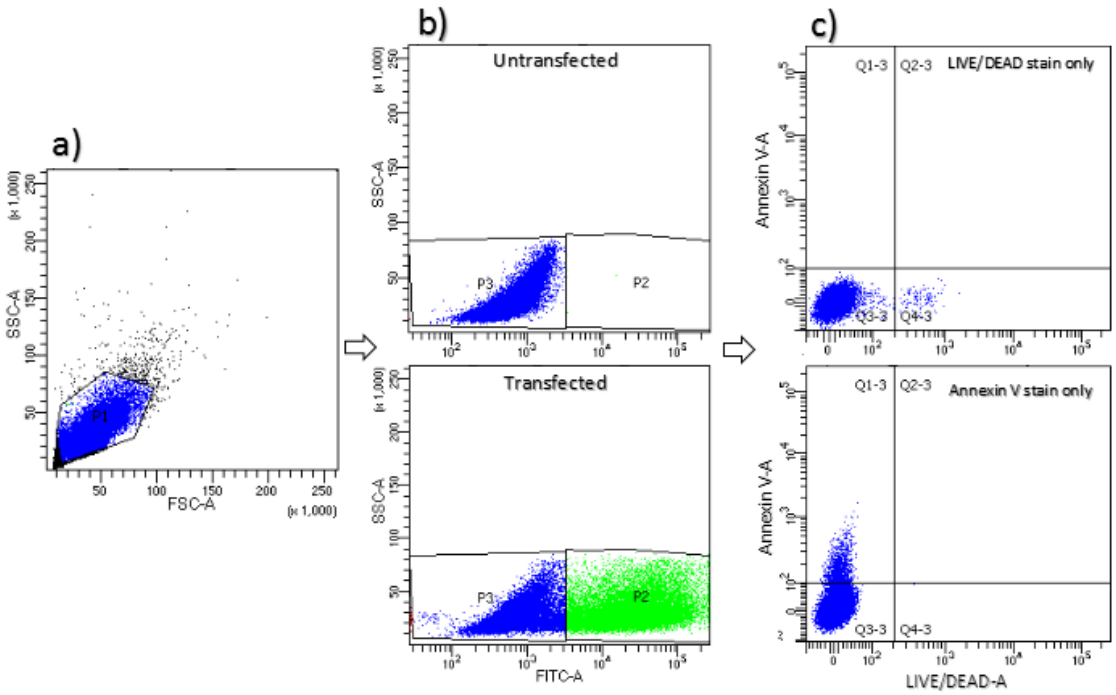


Figure 3.16: Gating strategy and controls for the Annexin V apoptosis assay.
a) Shows the selection of cells to form P1 using an untransfected population. b). shows the gating of transfected (green) and untransfected (blue) cells from population 1 using the untransfected population as the bench mark to select untransfected cells. c) Shows gating of P2 using a quadrant to select those cells that are apoptosing (Q1-3), live (Q3-3), and dead (Q2-3&4-3). The same quadrant was used for population 3 as well (data not shown).

The results presented in Table 3.3 show that after transfection with the $NRAS^{WT}$ or $NRAS^{MUT}$ GFP2 constructs, there is a marginal increase in the percentage of apoptosing cells between the transfected and untransfected populations. This increase in apoptosis was observed in both wild type and mutant samples. However the increase is greater in cells transfected with the $NRAS^{MUT}$-GFP2 construct, with an increase of 2.42 and 1.14 for replicate samples 1 and 2 respectively. In contrast the increase in wild type samples were a lower 0.38 and 0.95 for samples 1 and 2 respectively (Table 3.3). There was no evidence of an increase in the percentage of dead cells between the transfected populations in the wild type (1.65%) and mutant samples (1.69%).

Table 3.3: Results of an Annexin V apoptosis assay.
This table presents the proportion of live, dead and apoptosing cells in the transfected and untransfected populations of each sample identified using flow cytometry. The difference in the percentage of apoptosis seen in each population is also presented.

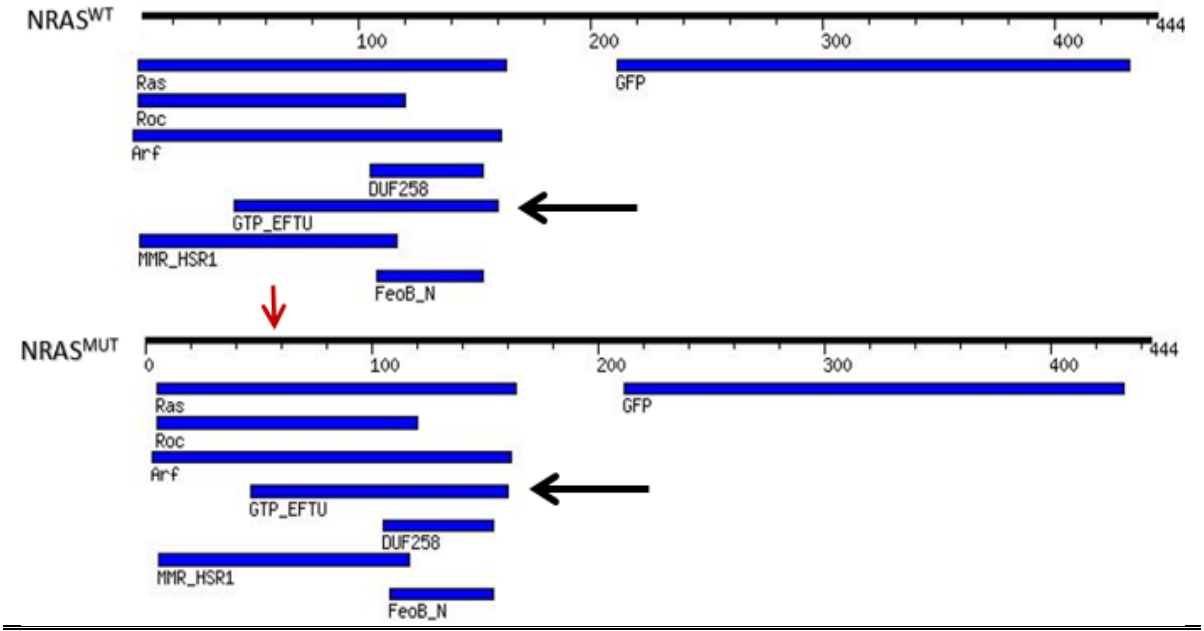
Sample		Live (%)	Dead (%)	Apoptosing (%)	n	Difference in % of apoptosis
$NRAS^{WT}$ Sample1	Transfected	96.83	0.51	2.67	22713	0.38
	Un-transfected	96.67	1.05	2.29	21957	
$NRAS^{WT}$ Sample 2	Transfected	94.63	1.14	4.23	22911	0.95
	Un-transfected	94.90	1.82	3.28	22119	
$NRAS^{MUT}$ Sample 1	Transfected	93.03	0.93	6.04	19315	2.42
	Un-transfected	95.39	0.99	3.62	27049	
$NRAS^{MUT}$ Sample 2	Transfected	94.95	0.76	4.29	19724	1.14
	Un-transfected	95.98	0.87	3.15	25991	

3.6.2 Bioinformatic prediction of the effect of the mutation.

Since little is known about the mutation and its potential effects on the function of NRAS, the sequences of the tagged $NRAS^{WT}$ and $NRAS^{MUT}$ proteins were analysed using the online tool MOTIF search (http://www.genome.jp/tools/motif/). This tool annotates query amino acid sequences with predicted binding domains and functional motifs using several available databases; the Pfam database[83] was used in this analysis. The result of this analysis is shown in Figure 3.17. The MOTIF tool predicted that the $NRAS^{WT}$-GFP and $NRAS^{MUT}$-GFP sequences contained both the RAS and GFP2 motifs, as expected. As seen in Figure 3.17 the tool also identified the similarities between the input sequence and the RAS of complex (ROC) domain, 50S ribosome-binding GTPase (MMR_HSR1) and ADP Ribosylation Factors (ARF). These are known to share sequence similarities as they are all in or associated with the RAS superfamily. This confirms that the tool is accurately identifying similar sequences

The tool further identified similarities to a GTP_EFTU motif, indicated by an arrow in Figure 3.17, in both the wild type and mutant sequences. GTP_EFTU is the GTP binding

domain of elongation factor-thermo unstable (EF-Tu), which is highly conserved across all guanine nucleotide-binding proteins (G proteins), including the RAS proteins. The mutation site is located within this motif as shown by a red arrow in Figure 3.17. The independent E-values, calculated by MOTIF, confirm that the substitution of the 58th amino acid reduces the confidence of this prediction 10-fold; from 5.8e^{-05} for NRAS^{WT} to 6.7e^{-04} for NRAS^{MUT}. The same motifs, excluding GFP, were also identified in the 3xFLAG tagged protein sequences (data not shown).



Name	Position	Description	E-value NRAS^{WT}	E-value NRAS^{MUT}
GFP	212…432	Green fluorescent protein	7.1e^{-76}	7.1e^{-76}
RAS	5…164	Ras family	8.3e^{-57}	2.6e^{-55}
ROC	5…120	Ras of complex, ROC, domain of DAPkinase	4.8e^{-20}	1.5e^{-19}
ARF	3…162	ADP-ribosylation factor family	7.4e^{-08}	1.1e^{-08}
GTP_EFTU	47…160	Elongation factor Tu GTP binding domain	5.8e^{-05}	0.00067
DUF258	105…154	Protein of unknown function	0.095	0.089
MMR_HSR1	6…117	50S ribosome-binding GTPase	0.00089	0.0026
FeoB_N	108…154	Ferrous iron transport protein B	0.061	0.059

Figure 3.17: Result from MOTIF search using Pfam database.
The GFP2 tagged protein sequences were annotated with known motifs and functional domains. Both the wild type and mutant show the same motifs, with the tool correctly identifying both the RAS and GFP2 sequences. Of particular interest is the GTP_EFTU motif which is indicated with a black arrow. The approximate location of the mutation site is indicated with a red arrow.

The same sequences were also analysed using the CD-Search tool[84] which searches the conserved domains database[84] and annotates the input sequence with known protein domains. From this analysis it was found that the mutation site, highlighted by a red arrow in Figure 3.18, is predicted to be situated within a number of the conserved domains annotated in the sequence. These include a GEF interaction sight, G3 box which is an effector interaction site and GTP/Mg^{2+} binding site (Figure 3.18). The mutation is also adjacent to the Switch II region which is known to be involved in the change in conformation in response to GTP binding which allows NRAS to interact with effector proteins and activate them.[20,45]

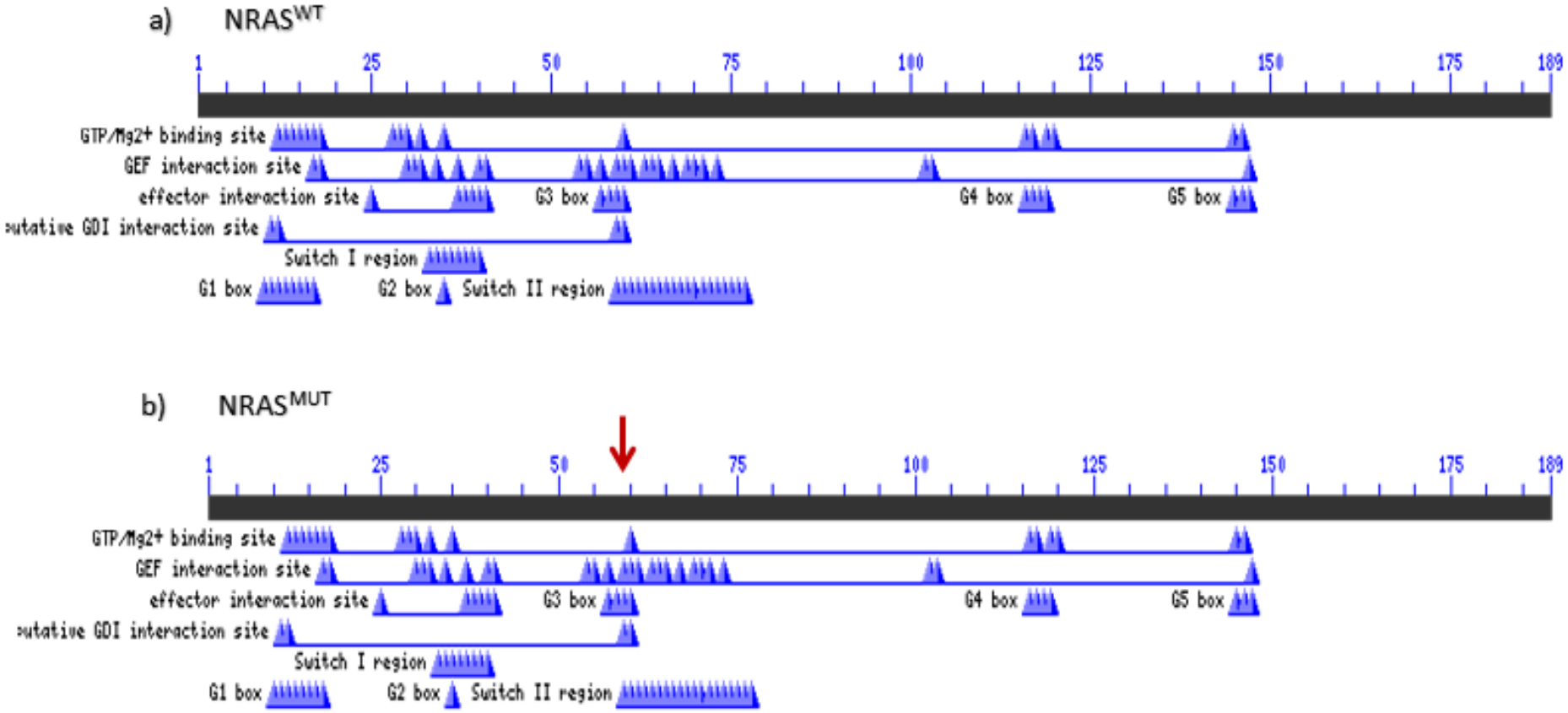


Figure 3.18: Annotated sequence for a) the NRAS^{WT} and b) the NRAS^{MUT} protein sequences using CD search
There are a number of domains within the NRAS sequence which could be affected by the T58I mutation (red arrow) including a GEF interaction sight, G3box and GTP/MG^{2+} binding site.

3.7 Summary

This study aimed to determine if the localisation and/or function of NRAS was effected by a novel c.173C>T (p.T58I) mutation identified in a patient with Noonan-like syndrome. A combination of bioinformatic analyses and literature searches predicted that the NRAS constructs could localise to a number of sub-cellular locations including the plasma membrane, endoplasmic reticulum, nucleus and secretory vesicles. These findings directed localisation studies, which were performed in U87-MG cells transfected with the GFP2 tagged $NRAS^{WT}$ and $NRAS^{MUT}$ constructs, to look for co-localisation with markers for the; plasma membrane, Golgi apparatus, endoplasmic reticulum, lysosomes, early endosomes and mitochondria. This study identified localisation of the mutant protein to the Golgi apparatus, plasma membrane and endoplasmic reticulum, while wild type protein was not observed to co-localise with any of the markers used.

An apoptosis assay was used to determine if the mutation resulted in an imbalance of the pro-survival/pro-apoptosis pathways as hypothesised to do. This assay identified an increase in the percentage of apoptosing cells in NAS^{MUT} transfected samples in comparison to both the wild type and an internal control of untransfected cells. Further functional analysis was performed using the bioinformatics tools MOTIF search and CD search which annotate protein sequences with known functional motifs and binding domains based on the Pfam and conserved domains database respectively. These tools identified that the mutation is located within/near a number of motifs including a GEF interaction sight, G3 effector interaction site, Switch II region, GTP/Mg2+ binding site and the GTP binding domain of EF-Tu, all of which are involved in the activation and/or inactivation of proteins.

4 DISCUSSION

As outlined in the introduction to this thesis, the patient in this study was diagnosed with Noonan-like syndrome based on a number of phenotypic traits including gross motor delay, craniofacial anomalies: macrocephaly, high forehead, lowset ears hypertelorism, strabismus and palpebral ptosis and musculoskeletal abnormalities: webbed neck, pectus excavatum and hyper-extensible joints. This patient was not found to harbour mutations in any of the genes usually associated with Noonan syndrome such as *PTPN11*, *KRAS* and *SOS1*. However a novel mutation in *NRAS* was identified. *NRAS* is one of the three isoforms of *RAS* which work as molecular switches for signalling pathways[31,32] thus playing a role in the regulation of cell growth, migration, survival, apoptosis and endocytosis.[33] NRAS achieves this through its intrinsic GTP*ase* activity by cycling between its inactive GDP-bound and active GTP-bound states.[34]

This potentially disease causing variant is a single nucleotide substitution (c.173C>T) in the 3rd exon of *NRAS* causing a substitution missense (p.Thr58Ile) mutation in NRAS. As this specific mutation has not been previously identified as a cause of Noonan syndrome this study aimed to validate its potential role in disrupting vital cell functions. The aim of this study was to identify bioinformatic and/or experimental evidence to suggest that the mutation results in disruption of localisation, and/or function, of NRAS. In doing so this study aimed to contribute to a lab-based pipeline which can be adapted for the validation of any rare disease variant discovered in Western Australian patients.

4.1 Localisation studies:

The RAS proteins have a fairly wide distribution within a cell having been observed to localise to the: Golgi apparatus,[50,51,85] plasma membrane,[30] endoplasmic reticulum[50,51] and the inner and outer membranes of the mitochondria.[34,86] This localisation has also been associated with their activation, a study by Chui *et al* (2002) showed that RAS stimulated with a growth factor is transiently active on the plasma membrane while it is constitutively active on the Golgi apparatus.[51] Given that this mutation was found to be an activating mutation (M. Tartaglia and G. Baynam, personal communication, 2016), it was hypothesised that the mutation would result in the localisation of $NRAS^{MUT}$ to the Golgi apparatus more so than the plasma membrane. While the results presented here suggest there was co-localisation between the mutant and the Golgi apparatus marker there was also co-localisation with the plasma membrane, suggesting that the activation of NRAS caused by the mutation is not the only factor affecting its localisation. While the localisation studies performed were qualitative, using intensity scatter plots, a quantitative analysis of localisation would be possible. An intensity scatter plot, as the name suggests, plots the intensity of each pixel of one colour against each pixel of a second[98], in this case the red of a marker and the green of NRAS. A quantitative analysis of localisation would allow for an increase, or decrease, in the co-localisation of $NRAS^{MUT}$ with these markers, in comparison to $NRAS^{WT}$, to be determined.

The activation of RAS on the Golgi apparatus is still not fully understood. Two current hypotheses suggest that the activation is the result of retrograde transport of RAS from the plasma membrane or that RAS relies on diffusible mediators which transfer messages from the plasma membrane to the Golgi apparatus which determine the

activation of RAS.[33] Current experimental data[33,85] indicates that the second hypothesis is likely. These data suggests that calcium is the mediator, causing an increase in calcium dependant exchange factors to the Golgi apparatus and GAPs to the plasma membrane, leading to activation of RAS at the Golgi apparatus and inactivation at the plasma membrane.[33,85] If this process is disrupted by the mutation resulting in active NRAS at both locations, or in incomplete activation at the plasma membrane this would explain the gain-in-function resulting from this mutation. To test this hypothesis the interaction of NRAS with effector molecules could be studied using a gel based pull-down assay. This would allow for the interaction of NRAS with the proteins known to be involved in the activation and inactivation, such as GEF and GAPs, to be studied (Section 4.2). If the mutation is causing NRAS to interact with these protein more or less than normal this would likely lead to a disruption of this activation/inactivation cycle.

Localisation of $NRAS^{MUT}$ to both the plasma membrane and Golgi apparatus indicate that the mutant protein is functioning normally and is thus able to be shuttled bidirectionally between the Golgi apparatus and plasma membrane, as wild type NRAS has previously been shown to do.[49] While this appears to be the natural subcellular localisation of NRAS, in the work carried out in this thesis the $NRAS^{WT}$ protein showed no evidence of co-localisation with either of the markers for these subcellular compartments as would have been expected.

$NRAS^{MUT}$ also showed strong nuclear localisation. While weak nuclear localisation was predicted when analysing the insert sequences using the online tools cNLS mapper and PSORTII, it was predicted for both the wild type and mutant proteins. There is no

evidence to suggest that the mutation would result in nuclear localisation. In addition if this localisation pattern were caused by the addition of the GFP2 tag there is no explanation as to why nuclear localisation was consistently seen in $NRAS^{MUT}$ transfected cells but not in $NRAS^{WT}$ transfected cells. Further evaluation of the 3xFLAG-tagged constructs could identify if this nuclear localisation is NRAS-GFP2 specific. To come to any conclusion further studies using smaller tags or completely untagged $NRAS^{MUT}$ and $NRAS^{WT}$ proteins, would be required. The latter would require the use of an antibody marker to the NRAS protein itself, followed by a fluorescent-tagged secondary antibody layer for visualisation.

In addition to nuclear localisation, $NRAS^{MUT}$ displayed partial co-localisation with PDI; an endoplasmic reticulum marker. This result was not seen in all images taken and thus further studies would be required to make any firm conclusions about endoplasmic reticulum localisation. If $NRAS^{MUT}$ is localising to the endoplasmic reticulum this may be because it is getting stuck during the post-translational process as the endoplasmic reticulum is one of the intermediate localisations prior too binding with the Golgi apparatus. There is no evidence to suggest that NRAS is retained in the endoplasmic reticulum, both the wild type and mutant sequences lack specific endoplasmic reticulum retention signals like KDEL[99] and neither the 3xFLAG nor GFP2 tag introduces one.

The lack of co-localisation of $NRAS^{WT}$ with any of the markers used, despite the fact that NRAS has been observed at many of these locations, suggests that the localisation or activation of the wild type protein is disrupted by the addition of the GFP2 tag. When looking at images taken of cells transfected with the pTagGFP2-N

vector alone, there are similarities between the staining pattern of GFP2 and $NRAS^{WT}$. This could be caused by the GFP2 tag altering the localisation of $NRAS^{WT}$ due to the size of the tag, however, there is no reason why only the wild type protein would be affected and not the mutant protein. The second possibility is that the $NRAS^{WT}$ insert was lost during the process of bulking up the plasmid in bacteria and its subsequent extraction, resulting in expression of GFP2 only. The third possibility is that $NRAS^{WT}$ is truly localising to the cytoplasm. These three possibilities are discussed further below.

The first point was addressed through the use of cells transfected with 3xFLAG tagged constructs. 3xFLAG is a smaller less obtrusive tag that is well known and frequently used for tagging of proteins. Both $NRAS^{WT}$-3xFLAG and $NRAS^{MUT}$-3xFLAG show a similar pattern of localisation as they do when tagged with the GFP2 tag. However, the 3xFLAG images were not of sufficient quality to make a definite conclusion. This suggests that the GFP2 tag is not altering the localisation of the protein. This suggestion is supported by the results of a study performed by Choy et al (1999) in which the localisation of endogenous RAS was used to confirm that their GFP-RAS construct was not localising unnaturally.[50] It is possible, however, that both the GFP2 and 3xFLAG tags lead to mislocalisation of NRAS. If this is the case overcoming this would require the use of cells which expressed only the wild type or mutant NRAS form, such as patient derived cell lines.

The second possibility was addressed through restriction digest of the plasmid DNA which confirmed that both the $NRAS^{WT}$ and $NRAS^{MUT}$ inserts were present in the plasmid DNA extracted from transformed bacteria and that they were of the expected size. This was further supported by a western blot of protein extracted from

transfected cells which showed that the NRAS-GFP2 protein was expressed in both the wild type and mutant samples, that they were of the expected size and that there was significant difference in size to that of GFP2 alone.

Thirdly the cytoplasmic staining pattern observed could be the result of $NRAS^{WT}$, and $NRAS^{MUT}$, being unable to bind strongly to its intended target, this is supported by a study which showed that the anchorage of RAS to the membrane is relatively weak.[52] Therefore it is possible that the addition of tags further reduces the efficiency of NRAS to bind to the membrane. Alternatively $NRAS^{WT}$ may naturally be localising to the cytoplasm, as the transport of depalmitoylated RAS from the plasma membrane to the Golgi apparatus has been shown to involve a cytosolic intermediate.[100]

4.2 Functional studies:

Preliminary analysis of the novel p.T58I substitution identified it to be a gain-of-function mutation that results in the enhanced expression of the ERK and AKT proteins (M. Tartaglia and G. Baynam, personal communication, 2016). AKT has a major role in regulating cell survival[87] and an increase in active AKT has been shown to protect a number of cell lines, including COS-7[88] and MDCK[89] epithelial cells, from apoptosis induced in a number of ways. Therefore, it was hypothesised that an increase in AKT caused by the mutation would result in lower levels of apoptosis. The results of this study are directly conflicting with this hypothesis as there was an increase in apoptosis observed in cells transfected with $NRAS^{MUT}$. This may be the result of a pro-apoptotic response from RAS to an increased activation of the RAS–RAF–MAPKK–MAPK pathway. This has been identified as one of the triggers causing RAS; which normally has a pro-survival function, to promote apoptotis.[101,102] While

there is no evidence as yet to suggest that the mutation affects the activation of the whole RAF–MAPKK–MAPK pathway, the mutation is known to cause the increased expression of ERK, also known as MAPK.

Due to the small sample size used in the apoptosis assay the statistical significance of the increase in apoptosis was unable to be determined. Only a single experiment with duplicates of each sample was performed, therefore definitive conclusions cannot be made from this data. This study was not repeated due to time constraints. Further replication of this assay needs to be performed in cells which do not contain the endogenous copy, for example in cells in which the gene is knocked-down, or using patient derived cell lines. This is because it is ambiguous as to whether the effect seen is due to an over expression of NRAS, and it is likely that any effects of $NRAS^{MUT}$ are masked by the endogenous copy.

Due to the limited time to generate experimental evidence for change of function, additional bioinformatic analyses were undertaken to predict putative effects of the mutation on protein function. The bioinformatics tool MOTIF showed that the mutation was found within a region which is similar to the conserved GTP_EFTU domain. GTP_EFTU is the GTP binding domain of the EF-Tu which is a G-protein that works in a similar way to the RAS proteins by cycling between its inactive GDP bound and active GTP bound states.[103] The binding of GTP to EF-Tu occurs in this domain, which is highly conserved across all G-proteins.[103] There are four main motifs within this binding domain, the second of which is a DxxG sequence. In RAS this sequence is DTAG and spans the 57^{th}-60^{th} amino acids[103] indicating that a T58I mutation could

disrupt this motif. The disruption of this motif could result in the disruption of the GTP*ase* activity of NRAS.

The second tool used supported these findings, this tool, CD-Search, identified that the mutation sits within/near a number of conserved domains. These include a GEF interaction site, a G3 box; an effector interaction site, and a GTP/Mg^{2+} binding site. All three of these domains are involved in the activation and inactivation of NRAS. The importance of the 58^{th} amino acid in these regions is unknown however it is possible that this is the cause of the gain-of-function that results from the mutation.

The mutation is also adjacent to the Switch II region which is known to be involved in the change in conformation in response to GTP binding which allows NRAS to interact with effector proteins and activate them.[20,45] Mutations identified within or near the switch II region have previously been shown to interfere with the GTPase function of RAS.[17] Therefore, if the efficiency of Switch II is affected by this mutation, it would likely lead to NRAS being stuck in its inactive or active state. This in turn would lead to more or less interaction with effector proteins, leading to the disruption of downstream cellular activity. Since this mutation is known to be a gain-of-function mutation it is likely that if this Switch II region is affected it would result in the protein being stuck in the active conformation.

4.3 Experimental validation pipeline:

Although this study failed to provide definitive data to validate the role of this novel NRAS mutation in causing Noonan-like syndrome, it was successful in providing information that can contribute to the development of a pipeline for the validation of

other novel rare disease variants. In particular this study highlights the areas of the pipeline which need to be altered or extended to obtain more conclusive results.

Optimisation of the experimental pipeline was a large part of this study. The optimisations and quality controls performed in this study as described in sections 2.1.1, 2.1.2, 2.1.4, 2.3.1 and 2.4.2.1 are necessary for any variants studied with this pipeline. Antibody marker optimisation, in particular, allowed for the inefficiency of the Glut1 antibody to be identified. The nonspecific staining of Glut1 seen in this study may be related to the use of it in association with RAS, as RAS has been associated with an increased transcription of Glut1 in order for increased glucose uptake.[104] Therefore it is possible that, in transfected cells which have an over expression of NRAS, there is also an overabundance of Glut1. In addition to this it was further discovered that U87-MG cells have low levels of membrane bound Glut1,[105] making Glut1 the wrong antibody to use in this particular cell type.

The pipeline used in this study involved designing gene inserts to be sub-cloned into both the GFP2 and 3xFLAG tagged expression vectors. Due to the size of the GFP2 tag the results gained with GFP2 tagged constructs needs to be validated if any anomalies are seen. This could be overcome with the use of patient derived cell lines in which NRAS could be directly probed.

In order to perform the functional analyses needed to determine the effects of a mutation on protein function the pipeline needs in include the creation of stably

transfected cells. This was one of the main limitations of this study as the transfected cells used were transiently transfected, thus the plasmid was episomal and lost during replication making longer-term assays impossible. Using stably transfected cells would allow for long-term analysis such as a proliferation or protein interaction assays to be performed. In addition to stably transfected cells, to successfully study the function of the mutant protein, the endogenous copy needs to be under expressed or absent from the cells. Using patient derived cell line would allow this to be done in a more efficient fashion and would make the designing of construct unnecessary, thus removing the ambiguity revolving around the addition of tags to NRAS.

Two examples of simplified validation pipelines that could result in less ambiguous results are shown in Figure 4.1. Figure 4.1.a shows a pipeline which utilises the same construct design, bioinformatic analyses and transfection process as this study, with the addition of stable transfection cells to allow for further functional studies to be performed. Figure 4.1.b shows an ideal pipeline which uses patient derived cell lines for functional and localisation studies. This would remove the need for the GFP2 and 3xFLAG tags.

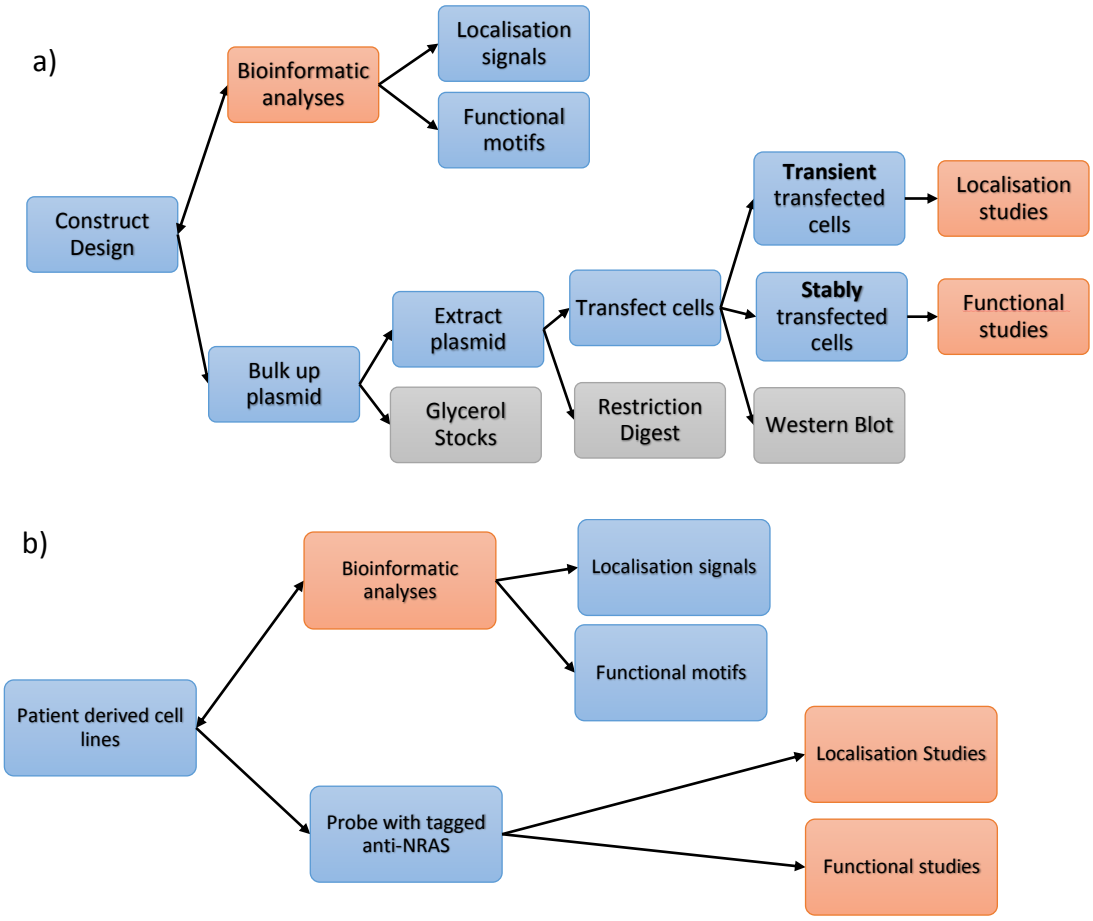


Figure 4.1: Two simplified experimental pipelines for the validation of rare disease variants.
a) Shows a pipeline which utilises the same construct design as used in this study with the addition of stably transfected cells to allow for further functional studies to be performed. b) Shows an ideal pipeline which uses patient derived cell line for functional and localisation studies.

4.4 Conclusion and further studies

In conclusion the results of this study are insufficient to validate the role of the c.173C>T p.T58I substitution in the rare disease Noonan-like syndrome. The preliminary information from this study indicates a potential difference in localisation and function between the mutant and wild type proteins. The mutant protein was seen to co-localise with the Golgi apparatus as expected if the mutant protein was constitutively active. However the mutant protein was also observed to co-localise with the markers for the plasma membrane, nucleus and endoplasmic reticulum. This

could indicate that the mutation affects more than just the activation of the protein. The wild type protein however was not seen to co-localise with any of the markers used leading to questions as to the validity of these results.

The functional assays performed, while limited to studies in transiently rather than stably transfected cells, indicate an increase in apoptosis in cells containing the mutant protein. This is the opposite of expected. These results however are preliminary and further replicates need to be performed to determine the significance of this difference before any conclusions are made. Bioinformatic analysis of the insert sequences revealed that the mutation is likely to affect the binding of GTP and/or the change in conformation associated with this binding. This is in line with the preliminary finding on the same mutation which showed that the mutation results in a gain in function.

To fully understand the effect of this mutation and thus validate its role in Noonan, further study is required. In particular further functional analyses in stably transfected cells looking at the: proliferation and differentiation of cells containing the mutant protein, activation of NRAS and interaction of NRAS with other proteins such as the GEF and GAPs. These functions are of particular interest as ERK and AKT, which are known to have enhanced expression caused by this mutation, are known to have a role in the regulation of apoptosis, cell proliferation and differentiation.[87,101,102,106] In addition bioinformatic analysis in this study identified that the mutation was found within/near a number of motifs involved in GTP binding and the interaction GTP bound NRAS with effector molecules.

Ideally functional and localisation studies would also be repeated in patient derived cell line. For the localisation studied this would remove the ambiguity of the GFP2 and 3xFLAG tags and their effects on the localisation of the NRAS. This would also enable the results of functional assays to be less ambiguous as the endogenous wild type NRAS expressed in the cell lines used in this study would be under expressed.

An alternate way to achieve this is to using CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated protein) system. This system is derived from the adaptive immune system used by the many bacteria and archaea to protect against the invasion of nucleic acids, for example during viral infection.[107-109] When transcribed the CRISPR sequences act as guide RNA which identify foreign DNA and target it for inactivation by nuclease-mediated cleavage.[110] One of the more commonly used systems is CRISPR/Cas9 which was adapted from *Streptococcus pyogenes* and optimised for genome editing in vertebrate cells.[110-113] Using this system would allow for two cell lines to be produced that are identical other than the NRAS mutation of interest. This would also, in theory, allow for the mutation in patient derived cell line to be repaired and return to normal function proving the effects of the mutation.

The information gained from these additional studies would allow for the mutation to be properly validated and in combination with the information gained in this study provided the information needed to develop a more efficient pipeline to perform these studies on other rare disease variants.

5 REFERENCES

1. Ayme, S., Rodwell, C. & eds. 2011 Report on the State of the Art of Rare Disease Activities in Europe of the European Union Committee of Experts on Rare Diseases - Part I: Overview of Rare Disease Activities in Europe and Key Developments in 2010. (2011).
2. Remuzzi, G. & Garattini, S. Rare diseases: what's next? *The Lancet* **371**, 1978-1979 (2008).
3. Department of Health Western Australia. WA Rare Diseases Strategic Framework 2015-2018. (ed. Office of Population Health Genomics, P.H.D.) (Department of Health Western Australia, Perth, 2015).
4. The Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223-228 (2015).
5. Baynam, G. *et al.* The rare and undiagnosed diseases diagnostic service – application of massively parallel sequencing in a state-wide clinical service. *Orphanet Journal of Rare Diseases* **11**, 1-7 (2016).
6. Mosteller, R.D., W, P. & Broek, D. Analysis of interaction between Ras and CDC25 guanine nucleotide exchange factor using yeast GAL4 two-hybrid system. *Methods in Enzymology* **255**, 135-148 (1995).
7. Ng, S.B., Nickerson, D.A., Bamshad, M.J. & Shendure, J. Massively parallel sequencing and rare disease. *Human molecular genetics* **19**, R119-24 (2010).
8. Jamuar, S.S. & Tan, E.-C. Clinical application of next-generation sequencing for Mendelian diseases. *Human Genomics* **9**, 10 (2015).
9. Dello Russo, C. *et al.* Next generation sequencing in the identification of a rare genetic disease from preconceptional couple screening to preimplantation genetic diagnosis. *Journal of prenatal medicine* **8**, 17-24 (2014).
10. Sato-Otsubo, A., Sanada, M. & Ogawa, S. Single-nucleotide polymorphism array karyotyping in clinical practice: where, when, and how? *Seminars in oncology* **39**, 13-25 (2012).
11. Altshuler, D., Daly, M.J. & Lander, E.S. Genetic mapping in human disease. *Science* **322**, 881-8 (2008).
12. Guryev, V. & Cuppen, E. Next-generation sequencing approaches in genetic rodent model systems to study functional effects of human genetic variation. *FEBS Lett* **583**, 1668-73 (2009).
13. Sawyer, S.L. *et al.* Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clinical genetics* **89**, 275-84 (2016).
14. Tabor, H.K., Berkman, B.E., Hull, S.C. & Bamshad, M.J. Genomics really gets personal: how exome and whole genome sequencing challenge the ethical framework of human genetics research. *American journal of medical genetics. Part A* **155A**, 2916-24 (2011).
15. Beaulieu, C.L. *et al.* FORGE Canada Consortium: outcomes of a 2-year national rare-disease gene-discovery project. *American journal of human genetics* **94**, 809-17 (2014).
16. Firth, H.V., Wright, C.F. & for the, D.D.D.S. The Deciphering Developmental Disorders (DDD) study. *Developmental Medicine & Child Neurology* **53**, 702-703 (2011).
17. Rauen, K.A. The RASopathies. *Annual Review Genomics Human Genetics* **14**, 355-69 (2013).
18. Cizmarova, M. *et al.* New Mutations Associated with Rasopathies in a Central European Population and Genotype-Phenotype Correlations. *Annals of human genetics* **80**, 50-62 (2016).
19. Allanson, J. & Roberts, A. Noonan Syndrome. in *GeneReviews®* Vol. 1993-2016 (eds Pagon, R. & Ardinger, H.) (University of Washington Seattle (WA), 2001).
20. Cirstea, I.C. *et al.* A restricted spectrum of NRAS mutations causes Noonan syndrome. *Nature Genetics* **42**, 27-9 (2010).

21. Van Der Burgt, I. *et al.* Clinical and molecular studies in a large Dutch family with Noonan syndrome. *American Journal of Medical Genetics* **53**, 187-191 (1994).
22. Wingbermuhle, E., Van der Burgt, I., Egger, J. & Verhoeven, W. Noonan Syndrome. in *Cognitive and Behavioral Abnormalities of Pediatric Diseases* (eds. Nass, R. & Yitzchak, F.) (Oxford Press, 2010).
23. National Human Genome Research Institute. Elements of Morphology. (2014).
24. Allanson, J.E. *et al.* The Face of Noonan Syndrome: Does Phenotype Predict Genotype. *American journal of medical genetics. Part A* **152A**, 1960-1966 (2010).
25. Allanson, J.E. Noonan syndrome. *Journal of medical genetics* **24**, 9-13 (1987).
26. Abdel-Salam, E. & Temtamy, S.A. Familial Turner phenotype. *The Journal of pediatrics* **74**, 67-72 (1969).
27. Maximilian, C., Ioan, D.M. & Fryns, J.P. A syndrome of mental retardation, short stature, craniofacial anomalies with palpebral ptosis and pulmonary stenosis in three siblings with normal parents. An example of autosomal recessive inheritance of the Noonan phenotype? *Genetic counseling* **3**, 115-8 (1992).
28. van Der Burgt, I. & Brunner, H. Genetic heterogeneity in Noonan syndrome: evidence for an autosomal recessive form. *American journal of medical genetics* **94**, 46-51 (2000).
29. OMIM, O.M.I.i.M. McKusick-Nathans Institute of Genetic Medicine. (Johns Hopkins University (http://omim.org/), Baltimore, MD, Accessed: 2016).
30. Henis, Y.I., Hancock, J.F. & Prior, I.A. Ras acylation, compartmentalization and signaling nanoclusters (Review). *Molecular Membrane Biology* **26**, 80-92 (2009).
31. Rodriguez-Viciana, P. & McCormick, F. RalGDS comes of age. *Cancer Cell* **7**, 205-6 (2005).
32. Rodriguez-Viciana, P., Sabatier, C. & McCormick, F. Signaling specificity by Ras family GTPases is determined by the full spectrum of effectors they regulate. *Molecular and cellular biology* **24**, 4943-54 (2004).
33. Mironov, A. & Pavelka, M. (eds.). *The Golgi Apparatus: State of the Art 110 Years After Camillo Golgi's Discovery*, 714 (Springer Wien NewYork, Austria, 2010).
34. Wolfman, J.C., M., P.S., Liao, J. & Wolfman, A. Structural and functional consequences of c-N-Ras constitutively associated with intact mitochondria. *Biochimica et Biophysica Acta* **1763**, 1108-1124 (2006).
35. Castellano, E. & Santos, E. Functional Specificity of Ras Isoforms: So Similar but So Different. *Genes & Cancer* **2**, 216-231 (2011).
36. Nikiforova, M.N. *et al.* RAS point mutations and PAX8-PPAR gamma rearrangement in thyroid tumors: evidence for distinct molecular pathways in thyroid follicular carcinoma. *The Journal of clinical endocrinology and metabolism* **88**, 2318-26 (2003).
37. Vasko, V. *et al.* Specific pattern of RAS oncogene mutations in follicular thyroid tumors. *The Journal of clinical endocrinology and metabolism* **88**, 2745-52 (2003).
38. Bezieau, S. *et al.* High incidence of N and K-Ras activating mutations in multiple myeloma and primary plasma cell leukemia at diagnosis. *Human mutation* **18**, 212-24 (2001).
39. Niemela, J.E. *et al.* Somatic KRAS mutations associated with a human nonmalignant syndrome of autoimmunity and abnormal leukocyte homeostasis. *Blood* **117**, 2883-6 (2011).
40. Oliveira, J.B. *et al.* NRAS mutation causes a human autoimmune lymphoproliferative syndrome. *Proc Natl Acad Sci U S A* **104**, 8953-8 (2007).
41. Kerr, B. *et al.* Genotype-phenotype correlation in Costello syndrome: HRAS mutation analysis in 43 cases. *Journal of medical genetics* **43**, 401-5 (2006).
42. Zampino, G. *et al.* Diversity, parental germline origin, and phenotypic spectrum of de novo HRAS missense changes in Costello syndrome. *Human mutation* **28**, 265-72 (2007).

43. Dessars, B. *et al.* Genotypic and gene expression studies in congenital melanocytic nevi: insight into initial steps of melanotumorigenesis. *The Journal of investigative dermatology* **129**, 139-47 (2009).
44. Hall, A. & Brown, R. Human N-ras: cDNA cloning and gene structure. *Nucleic Acids Res* **13**, 5255-68 (1985).
45. Schubbert, S. *et al.* Germline KRAS mutations cause Noonan syndrome. *Nature Genetics* **38**, 331-336 (2006).
46. Schaber, M.D. *et al.* Ras interaction with the GTPase-activating protein (GAP). *Proteins: Structure, Function and Genetics* **6**, 306-315 (1989).
47. Karnoub, A.E. & Weinberg, R.A. Ras oncogenes: split personalities. *Nature reviews. Molecular cell biology* **9**, 517-531 (2008).
48. The Human Protein ATLAS. NRAS. Vol. 2016 (accessed May 2016).
49. Quatela, S.E. & Philips, M.R. Ras Signaling on the Golgi. *Current Opinion in Cell Biology* **18**, 162-167 (2006).
50. Choy, E. *et al.* Endomembrane trafficking of ras: the CAAX motif targets proteins to the ER and Golgi. *Cell Press* **98**, 69-80 (1999).
51. Chiu, V.K. *et al.* Ras signalling on the endoplasmic reticulum and the Golgi. *Nature Cell Biology* **4**, 343-350 (2002).
52. Bivona, T.G. *et al.* PKC regulates a farnesyl-electrostatic switch on K-Ras that promotes its association with Bcl-XL on mitochondria and induces apoptosis. *Molecular cell* **21**, 481-93 (2006).
53. Prior, I.A., Lewis, P.D. & Mattos, C. A comprehensive survey of Ras mutations in cancer. *Cancer Research* **72**, 2457-2467 (2012).
54. Kinsler, V.A. *et al.* Multiple congenital melanocytic nevi and neurocutaneous melanosis are caused by postzygotic mutations in codon 61 of NRAS. *The Journal of investigative dermatology* **133**, 2229-36 (2013).
55. Oliveira, J.B. *et al.* NRAS mutation causes a human autoimmune lymphoproliferative syndrome. *Proceedings of the National Academy of Sciences of the U.S.A.* **104**, 8953-8 (2007).
56. Matsuda, K. *et al.* Spontaneous improvement of hematologic abnormalities in patients having juvenile myelomonocytic leukemia with specific RAS mutations. *Blood* **109**, 5477-80 (2007).
57. Zenker, M. *et al.* Expansion of the genotypic and phenotypic spectrum in patients with KRAS germline mutations. *Journal of medical genetics* **44**, 131-5 (2007).
58. Gripp, K.W. *et al.* Costello syndrome associated with novel germline HRAS mutations: An attenuated phenotype? *American Journal of Medical Genetics Part A* **146A**, 683-690 (2008).
59. De Filippi, P. *et al.* Germ-line mutation of the NRAS gene may be responsible for the development of juvenile myelomonocytic leukaemia. *British Journal of Haematology* **147**, 706-9 (2009).
60. Runtuwene, V. *et al.* Noonan syndrome gain-of-function mutations in NRAS cause zebrafish gastrulation defects. *Disease Models and Mechanisms* **4**, 393-399 (2011).
61. Denayer, E. *et al.* NRAS Mutations in Noonan Syndrome. *Molecular Syndromology* **3**, 34-38 (2012).
62. Kraoua, L. *et al.* Constitutional NRAS mutations are rare among patients with Noonan syndrome or juvenile myelomonocytic leukemia. *American Journal of Medical Genetics Part A* **158a**, 2407-11 (2012).
63. Ekvall, S. *et al.* Mutation in NRAS in familial Noonan syndrome--case report and review of the literature. *BMC Medical Genetics* **16**(2015).
64. Marshall, C.J., Hall, A. & Weiss, R.A. A transforming gene present in human sarcoma cell lines. *Nature* **299**, 171-3 (1982).
65. Shimizu, K., Goldfarb, M., Perucho, M. & Wigler, M. Isolation and preliminary characterization of the transforming gene of a human neuroblastoma cell line.

Proceedings of the National Academy of Sciences of the United States of America **80**, 383-387 (1983).
66. Niihori, T. *et al.* Germline KRAS and BRAF mutations in cardio-facio-cutaneous syndrome. *Nature genetics* **38**, 294-6 (2006).
67. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
68. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006 (2002).
69. Vincze, T., Posfai, J. & Roberts, R.J. NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic acids research* **31**, 3688-91 (2003).
70. Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic acids research* **40**, W597-603 (2012).
71. Magnan, C.N. & Baldi, P. SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. *Bioinformatics* **30**, 2592-7 (2014).
72. Asarnow, D. *et al.* Validation and international regulatory experience for a mycoplasma touchdown PCR assay. *Biologicals* **38**, 224-31 (2010).
73. Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10171-6 (2009).
74. Shen, H.B. & Chou, K.C. A top-down approach to enhance the power of predicting human protein subcellular localization: Hum-mPLoc 2.0. *Analytical biochemistry* **394**, 269-74 (2009).
75. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic acids research* **27**, 49-54 (1999).
76. King, B.R. & Guda, C. ngLOC: an n-gram-based Bayesian method for estimating the subcellular proteomes of eukaryotes. *Genome biology* **8**, R68 (2007).
77. Negi, S., Pandey, S., Srinivasan, S.M., Mohammed, A. & Guda, C. LocSigDB: a database of protein localization signals. *Database (Oxford)* **2015**(2015).
78. Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic acids research* **35**, W585-7 (2007).
79. Bairoch, A. *et al.* The Universal Protein Resource (UniProt). *Nucleic acids research* **33**, D154-9 (2005).
80. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**, 25-9 (2000).
81. Pierleoni, A., Martelli, P.L., Fariselli, P. & Casadio, R. BaCelLo: a balanced subcellular localization predictor. *Bioinformatics* **22**, e408-16 (2006).
82. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671-5 (2012).
83. Finn, R.D. *et al.* Pfam: the protein families database. *Nucleic acids research* **42**, D222-30 (2014).
84. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic acids research* **43**, D222-6 (2015).
85. Bivona, T.G. & Philips, M.R. Ras pathway signaling on endomembranes. *Current Opinion in Cell Biology* **15**, 136-142 (2003).
86. Rebollo, A., D., P.-S. & Martinez, A.C. Bcl-2 differentially targets K-, N-, and H-Ras to mitochondria in IL-2 supplemented or deprived cells: implications in prevention of apoptosis. *Oncogene* **18**, 4930-4939 (1999).
87. Downward, J. Mechanisms and consequences of activation of protein kinase B/Akt. *Current opinion in cell biology* **10**, 262-7 (1998).

88. Kulik, G., Klippel, A. & Weber, M.J. Antiapoptotic signalling by the insulin-like growth factor I receptor, phosphatidylinositol 3-kinase, and Akt. *Molecular and cellular biology* **17**, 1595-606 (1997).
89. Khwaja, A., Rodriguez-Viciana, P., Wennstrom, S., Warne, P.H. & Downward, J. Matrix adhesion and Ras transformation both activate a phosphoinositide 3-OH kinase and protein kinase B/Akt cellular survival pathway. *The EMBO journal* **16**, 2783-93 (1997).
90. Dingwall, C., Robbins, J., Dilworth, S.M., Roberts, B. & Richardson, W.D. The nucleoplasmin nuclear location sequence is larger and more complex than that of SV-40 large T antigen. *The Journal of cell biology* **107**, 841-9 (1988).
91. Kalderon, D., Roberts, B.L., Richardson, W.D. & Smith, A.E. A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499-509 (1984).
92. Lange, A. *et al.* Classical nuclear localization signals: definition, function, and interaction with importin alpha. *The Journal of biological chemistry* **282**, 5101-5 (2007).
93. Kosugi, S. *et al.* Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *The Journal of biological chemistry* **284**, 478-85 (2009).
94. Allen, M., Bjerke, M., Edlund, H., Nelander, S. & Westermark, B. Origin of the U87MG glioma cell line: Good news and bad news. *Science translational medicine* **8**, 354re3 (2016).
95. Brown, W.J., Goodhouse, J. & Farquhar, M.G. Mannose-6-phosphate receptors for lysosomal enzymes cycle between the Golgi complex and endosomes. *The Journal of cell biology* **103**, 1235-47 (1986).
96. Schmid, S.L., Fuchs, R., Male, P. & Mellman, I. Two distinct subpopulations of endosomes involved in membrane recycling and transport to lysosomes. *Cell* **52**, 73-83 (1988).
97. van Engeland, M., Nieland, L.J., Ramaekers, F.C., Schutte, B. & Reutelingsperger, C.P. Annexin V-affinity assay: a review on an apoptosis detection system based on phosphatidylserine exposure. *Cytometry* **31**, 1-9 (1998).
98. Dunn, K.W., Kamocka, M.M. & McDonald, J.H. A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology - Cell Physiology* **300**, C723-C742 (2011).
99. Stornaiuolo, M. *et al.* KDEL and KKXX retrieval signals appended to the same reporter protein determine different trafficking between endoplasmic reticulum, intermediate compartment, and Golgi complex. *Mol Biol Cell* **14**, 889-902 (2003).
100. Rocks, O. *et al.* An acylation cycle regulates localization and activity of palmitoylated Ras isoforms. *Science* **307**, 1746-52 (2005).
101. Kauffmann-Zeh, A. *et al.* Suppression of c-Myc-induced apoptosis by Ras signalling through PI(3)K and PKB. *Nature* **385**, 544-8 (1997).
102. Arber, N. Janus faces of ras: anti or pro-apoptotic? *Apoptosis* **4**, 383-8 (1999).
103. Kjeldgaard, M., Nyborg, J. & Clark, B.F. The GTP binding motif: variations on a theme. *FASEB Journal* **10**, 1347-68 (1996).
104. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. *Nature reviews. Cancer* **11**, 761-74 (2011).
105. Jin, Q., Agrawal, L., Vanhorn-Ali, Z. & Alkhatib, G. GLUT-1-independent infection of the glioblastoma/astroglioma U87 cells by the human T cell leukemia virus type 1. *Virology* **353**, 99-110 (2006).
106. McCubrey, J.A. *et al.* Roles of the Raf/MEK/ERK pathway in cell growth, malignant transformation and drug resistance. *Biochimica et biophysica acta* **1773**, 1263-84 (2007).
107. Koonin, E.V. & Makarova, K.S. CRISPR-Cas: an adaptive immunity system in prokaryotes. *F1000 biology reports* **1**, 95 (2009).
108. Makarova, K.S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology* **9**, 467-77 (2011).

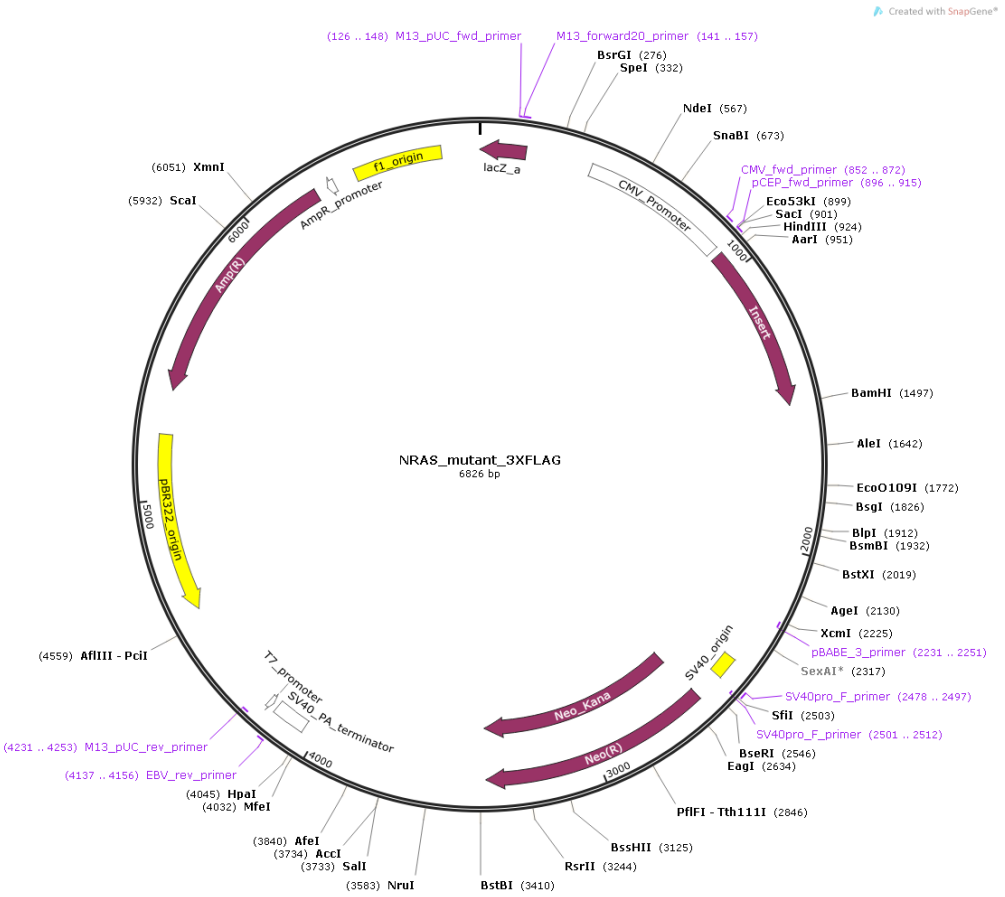
109. Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331-8 (2012).
110. Hruscha, A. *et al.* Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development* **140**, 4982-7 (2013).
111. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology* **31**, 230-2 (2013).
112. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-6 (2013).
113. Shen, B. *et al.* Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell research* **23**, 720-3 (2013).

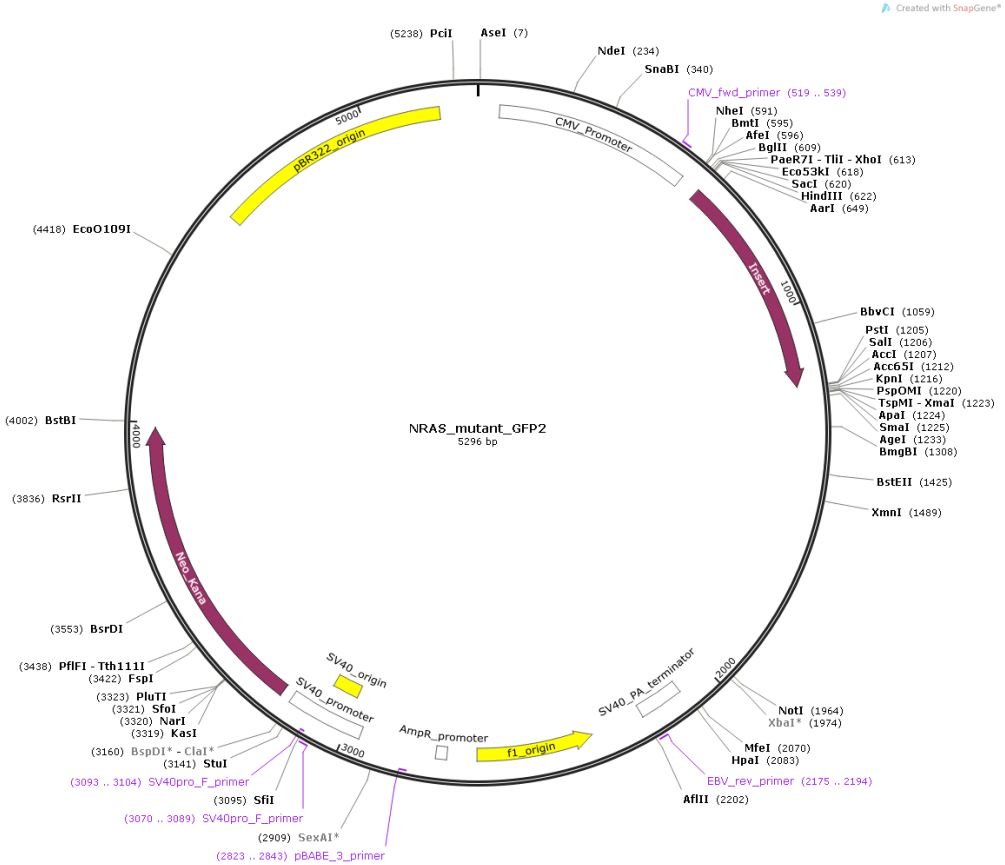
6 APPENDICES

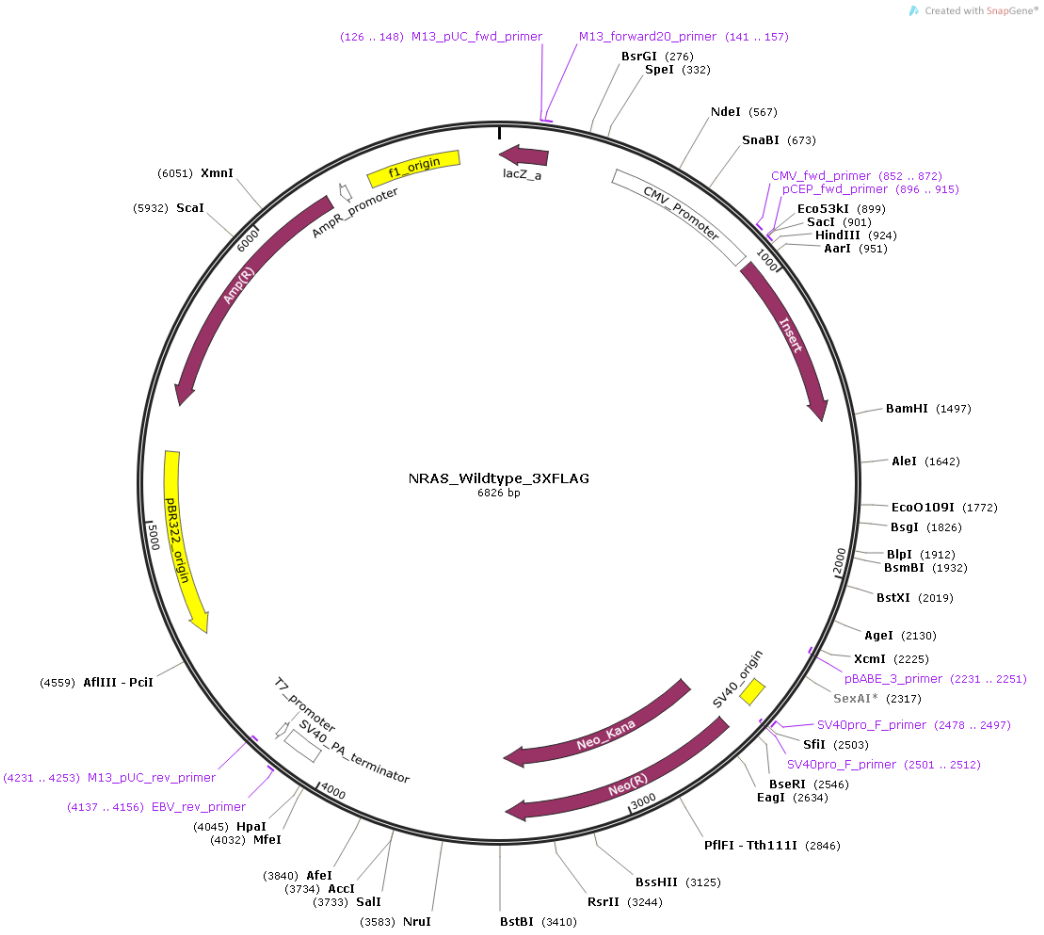
Appendix I: *Full gene names used in thesis*

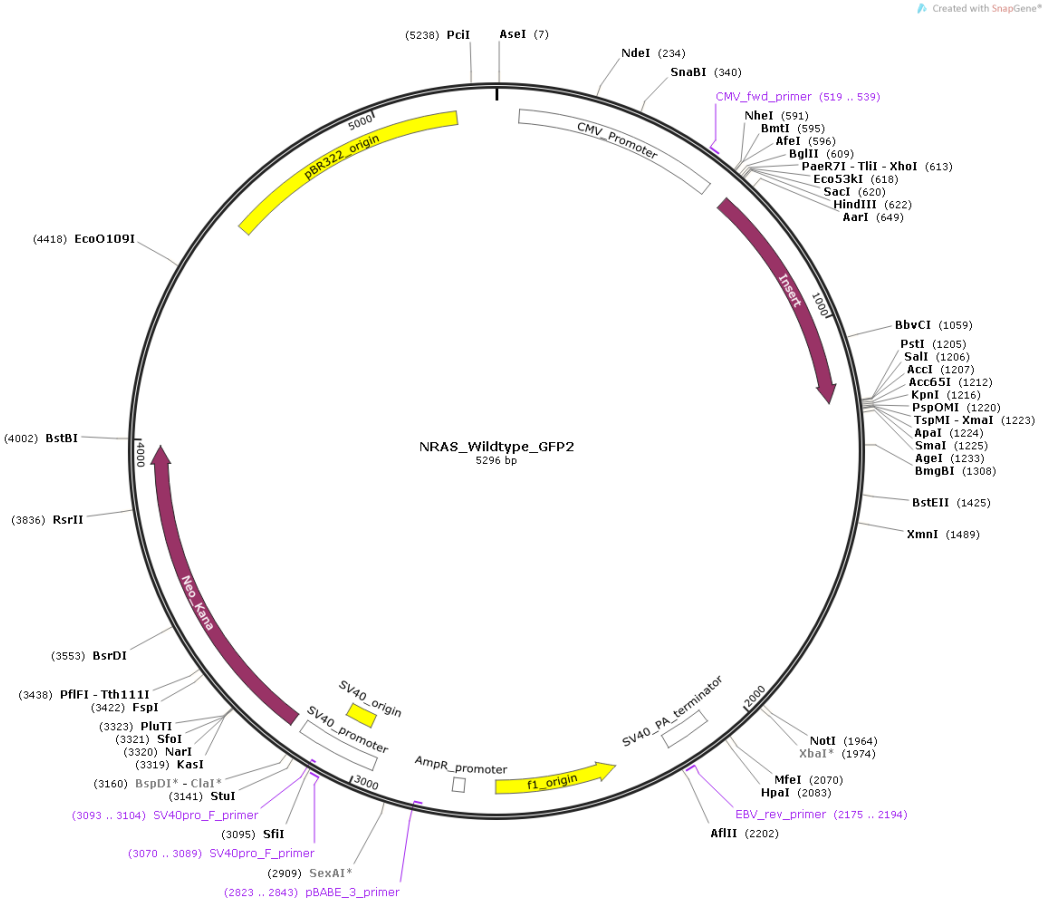
Official gene symbol	Official gene name
AKT	AKT serine/threonine kinase 1
BRAF	B-Raf proto-oncogene, serine/threonine kinase
CBL	Cbl proto-oncogene
ERK	Extracellular regulated MAP kinase
GRB2	Growth factor receptor bound protein 2
HRAS	V-HA-RAS Harvey rat sarcoma viral oncogene homolog
KRAS	V-KI-RAS2 Kirsten rat sarcoma viral oncogene homolog
LZTR1	Leucine-zipper-like transcription regulator 1
MEK	MAP kinase-ERK kinase
NF1	Neurofibromin 1
NORE1/ RASSF5	RAS association domain family member 5
NRAS	neuroblastoma RAS viral oncogene homolog
PDK1	Phosphoinositide-dependent kinase 1
PI3K	Phosphatidylinositol 3-kinase
PTPN11/ SHP2	Protein tyrosine phosphatase, non-receptor type 11
RAF1/ CRAF	Raf-1 proto-oncogene, serine/threonine kinase
RAL	RAS Like protein
RalGDS	Ral guanine nucleotide dissociation stimulator
RASSF1	RAS association domain family member 1
RIT1	RAS like without CAAX 1
SHC/NF2	Neurofibromin 2
SHOC2	SHOC2, leucine rich repeat scaffold protein
SOS1	SOS RAS/Rac guanine nucleotide exchange factor 1
SOS2	SOS RAS/Rho guanine nucleotide exchange factor 2

Appendix II: *Full plasmid maps provided by Invitrogen GeneArt™, images generated using SnapGene® Viewer ver 3.1.2*






Created with SnapGene®
(126 .. 148) M13_pUC_fwd_primer
M13_forward20_primer (141 .. 157)
BsrGI (276)
SpeI (332)
NdeI (567)
SnaBI (673)
CMV_fwd_primer (852 .. 872)
pCEP_fwd_primer (896 .. 915)
Eco53kI (899)
SacI (901)
HindIII (924)
AarI (951)
(6051) XmnI
(5932) ScaI
f1_origin
lacZ_a
AmpR_promoter
CMV_Promoter
6000
1000
Amp(R)
Insert
BamHI (1497)
AleI (1642)
EcoO109I (1772)
BsgI (1826)
BlpI (1912)
BsmBI (1932)
BstXI (2019)
AgeI (2130)
XcmI (2225)
pBABE_3_primer (2231 .. 2251)
SexAI* (2317)
SV40pro_F_primer (2478 .. 2497)
SfiI (2503)
SV40pro_F_primer (2501 .. 2512)
BseRI (2546)
EagI (2634)
PflFI - Tth111I (2846)
BssHII (3125)
RsrII (3244)
BstBI (3410)
NruI (3583)
SalI (3733)
AccI (3734)
AfeI (3840)
MfeI (4032)
HpaI (4045)
EBV_rev_primer (4137 .. 4156)
M13_pUC_rev_primer (4231 .. 4253)
AflIII - PciI (4559)
NRAS_Wildtype_3XFLAG
6826 bp
pBR322_origin
5000
T7_promoter
SV40_PA_terminator
4000
Neo_Kana
SV40_origin
Neo(R)
3000
2000


Created with SnapGene®
(5238) PciI
AseI (7)
NdeI (234)
SnaBI (340)
CMV_fwd_primer (519 .. 539)
NheI (591)
BmtI (595)
AfeI (596)
BglII (609)
PaeR7I - TliI - XhoI (613)
Eco53kI (618)
SacI (620)
HindIII (622)
AarI (649)
5000
pBR322_origin
CMV_Promoter
Insert
1000
(4418) EcoO109I
BbvCI (1059)
PstI (1205)
SalI (1206)
AccI (1207)
Acc65I (1212)
KpnI (1216)
PspOMI (1220)
TspMI - XmaI (1223)
ApaI (1224)
SmaI (1225)
AgeI (1233)
BmgBI (1308)
BstEII (1425)
XmnI (1489)
(4002) BstBI
(3836) RsrII
4000
Neo_Kana
(3553) BsrDI
(3438) PflFI - Tth111I
(3422) FspI
(3323) PluTI
(3321) SfoI
(3320) NarI
(3319) KasI
(3160) BspDI* - ClaI*
(3141) StuI
(3093 .. 3104) SV40pro_F_primer
(3095) SfiI
(3070 .. 3089) SV40pro_F_primer
(2909) SexAI*
(2823 .. 2843) pBABE_3_primer
SV40_origin
SV40_promoter
AmpR_promoter
f1_origin
3000
SV40_PA_terminator
2000
NotI (1964)
XbaI* (1974)
MfeI (2070)
HpaI (2083)
EBV_rev_primer (2175 .. 2194)
AflII (2202)
NRAS_Wildtype_GFP2
5296 bp

Appendix III: *Nucleotide sequences for the NRASWT and NRASMUT inserts with and without the GFP2 and 3xFLAg tags. These sequences were used to get the predicted protein sequences using ExPASy.*

NRAS^{WT} insert

```
5'-AAGCTTATGACTGAGTACAAACTGGTGGTGGTTGGAGCAGGTGGTGTTGGGAAAAGCGCACTGACAA
TCCAGCTAATCCAGAACCACTTTGTAGATGAATATGATCCCACCATAGAGGATTCTTACAGAAAACAAGT
GGTTATAGATGGTGAAACCTGTTTGTTGGACATACTGGATACAGCTGGACAAGAAGAGTACAGTGCCATG
AGAGACCAATACATGAGGACAGGCGAAGGCTTCCTCTGTGTATTTGCCATCAATAATAGCAAGTCATTTG
CGGATATTAACCTCTACAGGGAGCAGATTAAGCGAGTAAAAGACTCGGATGATGTACCTATGGTGCTAGT
GGGAAACAAGTGTGATTTGCCAACAAGGACAGTTGATACAAAACAAGCCCACGAACTGGCCAAGAGTTAC
GGGATTCCATTCATTGAAACCTCAGCCAAGACCAGACAGGGTGTTGAAGATGCTTTTTACACACTGGTAA
GAGAAATACGCCAGTACCGAATGAAAAAACTCAACAGCAGTGATGATGGGACTCAGGGTTGTATGGGATT
GCCATGTGTGGTGATGGGATCCCTGCAG-3'
```

NRAS^{WT} insert tagged with GFP2

```
5'-AAGCTTATGACTGAGTACAAACTGGTGGTGGTTGGAGCAGGTGGTGTTGGGAAAAGCGCACTGACAA
TCCAGCTAATCCAGAACCACTTTGTAGATGAATATGATCCCACCATAGAGGATTCTTACAGAAAACAAGT
GGTTATAGATGGTGAAACCTGTTTGTTGGACATACTGGATACAGCTGGACAAGAAGAGTACAGTGCCATG
AGAGACCAATACATGAGGACAGGCGAAGGCTTCCTCTGTGTATTTGCCATCAATAATAGCAAGTCATTTG
CGGATATTAACCTCTACAGGGAGCAGATTAAGCGAGTAAAAGACTCGGATGATGTACCTATGGTGCTAGT
GGGAAACAAGTGTGATTTGCCAACAAGGACAGTTGATACAAAACAAGCCCACGAACTGGCCAAGAGTTAC
GGGATTCCATTCATTGAAACCTCAGCCAAGACCAGACAGGGTGTTGAAGATGCTTTTTACACACTGGTAA
GAGAAATACGCCAGTACCGAATGAAAAAACTCAACAGCAGTGATGATGGGACTCAGGGTTGTATGGGATT
GCCATGTGTGGTGATGGGATCCCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATG
AGCGGGGGCGAGGAGCTGTTCGCCGGCATCGTGCCCGTGCTGATCGAGCTGGACGGCGACGTGCACGGCC
ACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGACGCCGACTACGGCAAGCTGGAGATCAAGTTCATCTG
CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTCTGCTACGGCATCCAGTGCTTC
GCCCGCTACCCCGAGCACATGAAGATGAACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACATCCAGG
AGCGCACCATCCAGTTCCAGGACGACGGCAAGTACAAGACCCGCGGCGAGGTGAAGTTCGAGGGCGACAC
CCTGGTGAACCGCATCGAGCTGAAGGGCAAGGACTTCAAGGAGGACGGCAACATCCTGGGCCACAAGCTG
GAGTACAGCTTCAACAGCCACAACGTGTACATCCGCCCCGACAAGGCCAACAACGGCCTGGAGGCTAACT
TCAAGACCCGCCACAACATCGAGGGCGGCGGCGTGCAGCTGGCCGACCACTACCAGACCAACGTGCCCCT
GGGCGACGGCCCCGTGCTGATCCCCATCAACCACTACCTGAGCACTCAGACCAAGATCAGCAAGGACCGC
AACGAGGCCCGCGACCACATGGTGCTCCTGGAGTCCTTCAGCGCCTGCTGCCACACCCACGGCATGGACG
AGCTGTACAGGTAA-3'
```

NRAS^{WT} insert tagged with 3xFLAG

```
5'-AAGCTTATGACTGAGTACAAACTGGTGGTGGTTGGAGCAGGTGGTGTTGGGAAAAGCGCACTGACAA
TCCAGCTAATCCAGAACCACTTTGTAGATGAATATGATCCCACCATAGAGGATTCTTACAGAAAACAAGT
GGTTATAGATGGTGAAACCTGTTTGTTGGACATACTGGATACAGCTGGACAAGAAGAGTACAGTGCCATG
AGAGACCAATACATGAGGACAGGCGAAGGCTTCCTCTGTGTATTTGCCATCAATAATAGCAAGTCATTTG
CGGATATTAACCTCTACAGGGAGCAGATTAAGCGAGTAAAAGACTCGGATGATGTACCTATGGTGCTAGT
GGGAAACAAGTGTGATTTGCCAACAAGGACAGTTGATACAAAACAAGCCCACGAACTGGCCAAGAGTTAC
GGGATTCCATTCATTGAAACCTCAGCCAAGACCAGACAGGGTGTTGAAGATGCTTTTTACACACTGGTAA
GAGAAATACGCCAGTACCGAATGAAAAAACTCAACAGCAGTGATGATGGGACTCAGGGTTGTATGGGATT
GCCATGTGTGGTGATGGGATCCCGGGCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATC
GACTACAAGGATGACGATGACAAGTAA-3'
```

ATCG	*Hind*III restriction enzyme cleavage site
ATCG	*Bam*HI restriction enzyme cleavage site
ATCG	*Pst*I restriction enzyme cleavage site
ATCG	Remainder of Multiple cloning region
ATCG	GFP2 or 3xFLAG tag
C/T	Mutation site

NRAS^{MUT} insert

5'-AAGCTTATGACTGAGTACAAACTGGTGGTGGTTGGAGCAGGTGGTGTTGGGAAAAGCGCACTGACAA
TCCAGCTAATCCAGAACCACTTTGTAGATGAATATGATCCCACCATAGAGGATTCTTACAGAAAACAAGT
GGTTATAGATGGTGAAACCTGTTTGTTGGACATACTGGATATAGCTGGACAAGAAGAGTACAGTGCCATG
AGAGACCAATACATGAGGACAGGCGAAGGCTTCCTCTGTGTATTTGCCATCAATAATAGCAAGTCATTTG
CGGATATTAACCTCTACAGGGAGCAGATTAAGCGAGTAAAAGACTCGGATGATGTACCTATGGTGCTAGT
GGGAAACAAGTGTGATTTGCCAACAAGGACAGTTGATACAAAACAAGCCCACGAACTGGCCAAGAGTTAC
GGGATTCCATTCATTGAAACCTCAGCCAAGACCAGACAGGGTGTTGAAGATGCTTTTTACACACTGGTAA
GAGAAATACGCCAGTACCGAATGAAAAAACTCAACAGCAGTGATGATGGGACTCAGGGTTGTATGGGATT
GCCATGTGTGGTGATGGGATCCCTGCAG-3'

NRAS^{MUT} insert tagged with GFP2

5'-AAGCTTATGACTGAGTACAAACTGGTGGTGGTTGGAGCAGGTGGTGTTGGGAAAAGCGCACTGACAA
TCCAGCTAATCCAGAACCACTTTGTAGATGAATATGATCCCACCATAGAGGATTCTTACAGAAAACAAGT
GGTTATAGATGGTGAAACCTGTTTGTTGGACATACTGGATATAGCTGGACAAGAAGAGTACAGTGCCATG
AGAGACCAATACATGAGGACAGGCGAAGGCTTCCTCTGTGTATTTGCCATCAATAATAGCAAGTCATTTG
CGGATATTAACCTCTACAGGGAGCAGATTAAGCGAGTAAAAGACTCGGATGATGTACCTATGGTGCTAGT
GGGAAACAAGTGTGATTTGCCAACAAGGACAGTTGATACAAAACAAGCCCACGAACTGGCCAAGAGTTAC
GGGATTCCATTCATTGAAACCTCAGCCAAGACCAGACAGGGTGTTGAAGATGCTTTTTACACACTGGTAA
GAGAAATACGCCAGTACCGAATGAAAAAACTCAACAGCAGTGATGATGGGACTCAGGGTTGTATGGGATT
GCCATGTGTGGTGATGGGATCCCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCCACCATG
AGCGGGGGCGAGGAGCTGTTCGCCGGCATCGTGCCCGTGCTGATCGAGCTGGACGGCGACGTGCACGGCC
ACAAGTTCAGCGTGCGCGGCGAGGGCGAGGGCGACGCCGACTACGGCAAGCTGGAGATCAAGTTCATCTG
CACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTCTGCTACGGCATCCAGTGCTTC
GCCCGCTACCCCGAGCACATGAAGATGAACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACATCCAGG
AGCGCACCATCCAGTTCCAGGACGACGGCAAGTACAAGACCCGCGGCGAGGTGAAGTTCGAGGGCGACAC
CCTGGTGAACCGCATCGAGCTGAAGGGCAAGGACTTCAAGGAGGACGGCAACATCCTGGGCCACAAGCTG
GAGTACAGCTTCAACAGCCACAACGTGTACATCCGCCCCGACAAGGCCAACAACGGCCTGGAGGCTAACT
TCAAGACCCGCCACAACATCGAGGGCGGCGGCGTGCAGCTGGCCGACCACTACCAGACCAACGTGCCCCT
GGGCGACGGCCCCGTGCTGATCCCCATCAACCACTACCTGAGCACTCAGACCAAGATCAGCAAGGACCGC
AACGAGGCCCGCGACCACATGGTGCTCCTGGAGTCCTTCAGCGCCTGCTGCCACACCCACGGCATGGACG
AGCTGTACAGGTAA-3'

NRAS^{MUT} insert tagged with 3xFLAG

5'-AAGCTTATGACTGAGTACAAACTGGTGGTGGTTGGAGCAGGTGGTGTTGGGAAAAGCGCACTGACAA
TCCAGCTAATCCAGAACCACTTTGTAGATGAATATGATCCCACCATAGAGGATTCTTACAGAAAACAAGT
GGTTATAGATGGTGAAACCTGTTTGTTGGACATACTGGATATAGCTGGACAAGAAGAGTACAGTGCCATG
AGAGACCAATACATGAGGACAGGCGAAGGCTTCCTCTGTGTATTTGCCATCAATAATAGCAAGTCATTTG
CGGATATTAACCTCTACAGGGAGCAGATTAAGCGAGTAAAAGACTCGGATGATGTACCTATGGTGCTAGT
GGGAAACAAGTGTGATTTGCCAACAAGGACAGTTGATACAAAACAAGCCCACGAACTGGCCAAGAGTTAC
GGGATTCCATTCATTGAAACCTCAGCCAAGACCAGACAGGGTGTTGAAGATGCTTTTTACACACTGGTAA
GAGAAATACGCCAGTACCGAATGAAAAAACTCAACAGCAGTGATGATGGGACTCAGGGTTGTATGGGATT
GCCATGTGTGGTGATGGGATCCCGGGCTGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATC
GACTACAAGGATGACGATGACAAGTAA-3'

ATCG	*Hind*III restriction enzyme cleavage site
ATCG	*Bam*HI restriction enzyme cleavage site
ATCG	*Pst*I restriction enzyme cleavage site
ATCG	Remainder of Multiple cloning region
ATCG	GFP2 or 3xFLAG tag
C/T	Mutation site

Appendix IV: *Predicted amino acid sequences of tagged and untagged NRAS^{WT} and NRAS^{MUT} proteins, as predicted by ExPASy.*

NRAS_Wildtype

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAG
QEEYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDL
PTRTVDTKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQG
CMGLPCVVM

NRAS_Mutant

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDIAG
QEEYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDL
PTRTVDTKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQG
CMGLPCVVM

NRAS_Wildtype_GFP2

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDT
AGQEEYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKC
DLPTRTVDTKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGT
QGCMGLPCVVMGSLQSTVPRARDPPVATMSGGEELFAGIVPVLIELDGDVHGHKFSVRGE
GEGDADYGKLEIKFICTTGKLPVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGY
IQERTIQFQDDGKYKTRGEVKFEGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYI
RPDKANNGLEANFKTRHNIEGGGVQLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRN
EARDHMVLLESFSACCHTHGMDELYR

NRAS_Wildtype_3xFLAG

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDT
AGQEEYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKC
DLPTRTVDTKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGT
QGCMGLPCVVMGSRADYKDHDGDYKDHDIDYKDDDDK

NRAS_Mutant_GFP2

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDI
AGQEEYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKC
DLPTRTVDTKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGT
QGCMGLPCVVMGSLQSTVPRARDPPVATMSGGEELFAGIVPVLIELDGDVHGHKFSVRGE
GEGDADYGKLEIKFICTTGKLPVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGY
IQERTIQFQDDGKYKTRGEVKFEGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYI
RPDKANNGLEANFKTRHNIEGGGVQLADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRN
EARDHMVLLESFSACCHTHGMDELYR

NRAS_Mutant_3xFLAG

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDI
AGQEEYSAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKC
DLPTRTVDTKQAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGT
QGCMGLPCVVMGSRADYKDHDGDYKDHDIDYKDDDDK

Appendix V: *Full results from LocSigDB*

NRAS^{WT}

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEY
SAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVDTK
QAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQGCMGLPCVVM

Signal	Coordinate(s)	Localization
Yx{2}[VILFWCM]	3-7, 63-67, 156-160	Lysosome
Dx{1}E	46-49	Endoplasmic reticulum

NRAS^{WT}-GFP

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEY
SAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVDTK
QAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQGCMGLPCVVMGSL
QSTVPRARDPPVATMSGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKFICTT
GKLPVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKYKTRGEVKF
EGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFKTRHNIEGGGVQL
ADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDHMVLLESFSACCHTHGMDELYR

Signal	Coordinate(s)	Localization
[DE]x{3}L[LI]	395-401, 421-427	Lysosome|melanosome
Yx{2}[VILFWCM]	3-7, 63-67, 156-160, 244-248	Lysosome
Dx{1}E	46-49	Endoplasmic reticulum
[HK]x{1}K	282-285, 310-313, 331-334	Endoplasmic reticulum

NRAS^{WT}-3xFLAG

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEY
SAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVDTK
QAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQGCMGLPCVVMGSR
ADYKDHDGDYKDHDIDYKDDDDK

Signal	Coordinate(s)	Localization
Yx{2}[VILFWCM]	3-7, 63-67, 156-160	Lysosome
Dx{1}E	46-49	Endoplasmic reticulum

NRAS^{MUT}

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEY
SAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVDTK
QAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQGCMGLPCVVM

Signal	Coordinate(s)	Localization
Yx{2}[VILFWCM]	3-7, 63-67, 156-160	Lysosome
Dx{1}E	46-49	Endoplasmic reticulum

NRAS^{MUT}-GFP2

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEY
SAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVDTK
QAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQGCMGLPCVVMGSL
QSTVPRARDPPVATMSGGEELFAGIVPVLIELDGDVHGHKFSVRGEGEGDADYGKLEIKFICTT
GKLPVPWPTLVTTLCYGIQCFARYPEHMKMNDFFKSAMPEGYIQERTIQFQDDGKYKTRGEVKF
EGDTLVNRIELKGKDFKEDGNILGHKLEYSFNSHNVYIRPDKANNGLEANFKTRHNIEGGGVQL
ADHYQTNVPLGDGPVLIPINHYLSTQTKISKDRNEARDHMVLLESFSACCHTHGMDELYR

Signal	Coordinate(s)	Localization
[DE]x{3}L[LI]	395-401, 421-427	Lysosome|melanosome
Yx{2}[VILFWCM]	3-7, 63-67, 156-160, 244-248	Lysosome
Dx{1}E	46-49	Endoplasmic reticulum
[HK]x{1}K	282-285, 310-313, 331-334	Endoplasmic reticulum

NRAS^{MUT}-3xFLAG

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEY
SAMRDQYMRTGEGFLCVFAINNSKSFADINLYREQIKRVKDSDDVPMVLVGNKCDLPTRTVDTK
QAHELAKSYGIPFIETSAKTRQGVEDAFYTLVREIRQYRMKKLNSSDDGTQGCMGLPCVVMGSR
ADYKDHDGDYKDHDIDYKDDDDK

Signal	Coordinate(s)	Localization
Yx{2}[VILFWCM]	3-7, 63-67, 156-160	Lysosome
Dx{1}E	46-49	Endoplasmic reticulum