# Post-transcriptional determinants of RAS protein abundance.

Hannah Benisty

DOCTORAL THESIS UPF / YEAR 2019

THESIS SUPERVISOR

Dr. Luis Serrano Pubul

Systems Biology Unit, Centre for Genomic Regulation

# Acknowledgements

The four years as a PhD student at the Centre for Genomic Regulation have been supported by many people and I would like to thank them for their contributions.

In particular, thanks go to Luis, who accepting to supervise me as her PhD student gifted me with a very enriching and rewarding experience. Luis has provided me keen insight for my research and has encouraged me to develop my hypotheses. I find amazing his dedication to science, how he can meet often, discuss and remember all details of the projects despite his fully packed agenda. I have been inspired by his curiosity and his outside-the-box thinking, which will for sure influence my future scientific work. This curiosity is reflected in his group, where during these years I have collected a wide variety of scientific knowledge on *Mycoplasma pneumoniae*, human eye diseases, genetic engineering of bacteria, microexon regulation, whole-cell modeling, and many other topics.

Many thanks to Christina Kiel, who did not hesitate to support me when I decided to start my PhD and for her supervision during the first years of my thesis and her feedback in this dissertation.

I also extend my thanks to the members of my Thesis Advisory Committee Mara Dierssen, Bill Keyes, Susana de la Luna and Eduard Batlle for the valuable feedback on my research.

I feel very lucky to have developed as a scientist in such a friendly laboratory atmosphere. First, I have to thank the “codon usage” people Marc Weber and Xavier Hernandez-Alias, the co-authors of the manuscript presented in this thesis for contributing to my project, for their effort and patience when explaining me the bioinformatics. It has been a very nice collaboration, always enjoying the discussions in a good mood. Marc, fa dos anys et vaig demanar de participar en el meu projecte i la teva implicació ha sigut clau. Gràcies especialment per ajudar-me amb molta didàctica a entrar en el món del codon usage. Xavi! Quina sort he tingut que l'any passat comencessis el PhD amb nosaltres! Gràcies per la teva ajuda en aquest projecte i per tantes discussions sobre els tRNAs. Many thanks to my deskmates, Violeta Beltrán Sastre and Sarah Di Bartolo, for your help on a daily basis with great

scientific advice in my projects, for helping me with my many hesitations, and for always having encouraging words for me. Sarah, thanks for our the meetings to discuss about the project presented in Chapter 3 of this thesis. Claire Lastrucci, merci mon amie pour ton soutien, ta générosité et surtout de m'avoir aidé à gérer cette dernière année de thèse et prendre soin de moi, avec toutes tes petites attentions, toujours prête à m'écouter attentivement et me donner des précieux conseils que je garde pour toujours, pour nos petites pauses et pour me forcer à rentrer à la maison après des longues journées de manips. To Ludovica Ciampi, from bringing to the lab a lot of joy, sweetness and good humor and for making the group stick even more together. Raul Burgos gracias por tu buen humor y por animarme siempre con mis presentaciones y en los partidos de volley. Rocco Mazzolini, por poner ritmo al laboratorio con tus canciones silbadas, por tu zen attitude y el acroyoga. To Dan Shaw for your kindness, humor, English corrections, for cheering me during our volley matches, and for being my swimming companion in these rough waters. Eva Garcia-Ramallo por cuidar de todos nosotros y por siempre estar dispuesta a ayudar. A Carlos Pinero por tu ayuda con la biología molecular, por tus consejos para enfrentarme a la escritura de esta tesis, por inculcarme tu frikismo de organización de alícuotas que me ha sido increíblemente útil, y por traerme un poco de Madrid al laboratorio. Ariadna Montero, por ser un ejemplo de energía y eficacia, por tus mimos y por la generosidad con tus entrenos que me han ayudado a pasar este último año de doctorado con las endorfinas necesarias para acabar la tesis con fuerza. A Maria Lluch por tu entusiasmo y sobretodo por transmitirme seguridad con tu sonrisa constante durante todas mis presentaciones. To the dry lab, for their patience with the experimentalist and for always being ready to help with any bioinformatic question. Especialmente gracias a Javier Delgado, por tunearme el ordenador, por ayudarme con scripts, por las clases particulares de FoldX, por tu sarcasmo y tu sentido de la provocación que han animado tanto nuestras comidas en la cantina y tardes en el stadium bar. A Leandro Radusky por las aguas, por motivarme a empezar a escribir mi tesis, por incitarnos a filosofar, por los asados y por siempre transmitirnos tu serenidad. To Damiano Cianferoni, for bringing a bit of Italian sassiness to the lab, for helping with FoldX and for motivating me to stay until late playing volley this summer which helped to clear my mind. Anash Garrab por tu amabilidad y simpatía diarias. Samuel por implicarte tanto para que lo pasemos bien, por ser el mejor capitán de volley y llevar a Pelotas Chungas a la victoria, por

habernos hecho el video más divertido de “Dance your lab”, y por estar siempre dispuesto a ayudar.
I also want to thank past members. Thanks to Kiana Toufighi, Carolina Gallo, Besray Unal, Veronica Llorens for their support and advice when I decided to start my PhD. A Sira Martínez, por siempre motivarme y aconsejarme desde la experiencia. To Jae-Seong Yang, for your feedback on my work, you are probably the most brilliant mind and the best dancer I know! Thanks to Tony Ferrar por ser tan acogedor cuando llegué al laboratorio, por nuestras comidas en la playa y nuestras pausas de gossips, por transmitirme tu espíritu objetivo y a la vez crítico y por preocuparte siempre por mi. A Marie Trussart, por haberme sacado a celebrar a Ultramarinos el día que decidí hacer este doctorado, por contagiarme tus ganas de disfrutar, por estar tan presente aunque ahora estés en la otra punta del mundo, por siempre intentar darme confianza y seguridad con mis proyectos y presentaciones, por estar disponible siempre para ayudarme, tu amistad es un tesoro.

Aside from the lab mates, the CRG is a nice environment to work. Muchas gracias Reyes Perza, por tu buen humor, por ser tan resolutiva y por ayudarnos tanto. Thanks to our neighbors from the HTS facility, especially to Carlo Carolis and Katie Broadbent for their valuable support in my project presented in Chapter 3 and always ready to help. To the PhD community, especially Reza Soda, Sebastian Ulrich, Alessandro Dasti, Andrea Tassinari, Júlia Domingo for keeping always optimistic and sharing our thesis fears in the corridor. Thanks to the members of the Tissue engineering facility Laura Batlle, Martin Gigerey and Marta Vila for their support with organoids experiments. Especialment a tu Laura, per invitar-me a ser instructora al curs d'intestinal i liver organoids, va ser una experiència molt enriquidora i per sempre interessar-te pels meus avanços.

A Samira Jaeger por tus rigurosas correcciones de la introducción de esta tesis y por ayudarme a hacer el planning cuando estaba desorientada con la escritura, por animarme a hacer todas esas clases de baile en el centro cívico y por convencerme de ir a la clase de impro teatral que me ha ayudado a perder la vergüenza de hacer presentaciones en público. A les bioquímiques, Georgina Balcells, Mireia Rossell, Gemma Bullich, Laura Padilla i Astrid Bull gràcies per retrobar-nos amb soparets compartint consells de doctores a futures doctores. A mis raíces de Andorra, Sandra Moreira, Ester Carrera, Marc Gomez, Matias Rey e Ismael Haddad que, aunque hayamos estado

dispersos por los cuatro vientos todos habéis desempeñado un papel importante a lo largo de este doctorado. Sandruki, gracias por tus mensajes de ánimo, por tu apoyo durante mi tesis, y por tu comprensión cuando el tiempo me ha faltado.

Martin, it has been amazing to have you by my side on this journey! Thank you for being a steadfast support during my PhD, your scientific guidance has been invaluable. Thanks for listening one hundred times to my rehearsals, for sharing your contagious passion for science, for manipulating the whole lab to go to the stairs of Poble Sec late in the night to play guitar, for the after-work evenings in our adored Plaça Sortidor brainstorming about our projects, and most importantly for your love and making me laugh so much.

Y para terminar, gracias a mi familia. Gracias por vuestro cariño, amor y apoyo incondicional. Por visitarme en todos mis laboratorios y siempre estar interesados por mis experimentos y proyectos. Por inculcarme los valores de tolerancia y constancia. Por animarme y ayudarme en todos mis pasos desde el día que me llevasteis todos a Toulouse hasta hoy que acabo de escribir mi tesis.

# Abstract

The RAS oncogenes –KRAS, NRAS and HRAS– are mutated in one third of human cancers where they exhibit different mutation patterns. A potential factor contributing to this mutation bias is the variation of RAS expression levels. Here, I investigate some of the determinants of RAS protein abundance. First, I examine whether codon bias among RAS genes and within other cancer gene families plays a role in cell context-specific expression. I further describe a tRNA expression program that favors oncogene translation in proliferating cells. Second, I investigate why oncogenic RAS mutants exhibit a higher protein abundance than the RAS wild type. In this context, I study the underlying mechanisms leading to this variation and more specifically how protein-protein interactions between RAS and its downstream binding partners change the protein turnover of RAS and therefore, its protein abundance. Overall, this thesis provides insight into the possible relevance of RAS protein synthesis and protein degradation as determinants of RAS mutation patterns in human cancers.

# Resum

Els oncogens KRAS, NRAS i HRAS estan mutats en un terç dels càncers en humans on hi exhibeixen patrons de mutació diferents. Un possible factor que contribueix a aquest biaix de mutació és la variació dels nivells d'expressió de RAS. En aquesta tesi investigo els elements determinants de l'abundància de la proteïna RAS. Primer, examino si el biaix de codó entre els gens RAS i entre gens d'altres famílies implicades en càncer contribueix a les diferències d'expressió, en funció del context cel·lular. Així mateix, descric un programa d'expressió de tRNA que facilita la traducció d'oncogens en cèl·lules proliferatives. En segon lloc, investigo per què mutants oncogènics de RAS tenen una abundància de proteïna més elevada que la RAS salvatge. Així mateix, estudio els mecanismes subjacents responsables d'aquesta variació i més concretament el paper de les interaccions de RAS amb altres proteïnes en la regulació de la seva abundància. Així doncs, aquesta tesi estudia la possible rellevància dels mecanismes de síntesi i degradació de la proteïna RAS en els patrons de mutació en càncer.

# Resumen

Los oncogenes KRAS, NRAS y HRAS, están mutados en un tercio de cánceres humanos, exhibiendo en ellos patrones de mutación diferentes. Un factor potencial que contribuye a este sesgo mutacional es la variación de los niveles de expresión de RAS. En esta tesis investigo los determinantes de la abundancia de proteína de RAS. En primer lugar, examino si el sesgo de uso de codones entre los genes RAS y entre genes de otras familias implicadas en cáncer contribuye a las diferencias de expresión en función del contexto celular. Con este objetivo, describo un programa de expresión de tRNA que facilita la traducción de oncogenes en células proliferativas. En segundo lugar, investigo por qué mutantes oncogénicos de RAS presentan una abundancia de proteína más elevada que RAS de tipo salvaje. Así mismo, estudio los mecanismos subyacentes responsables de esta variación y más concretamente el rol de las interacciones de RAS con otras proteínas en la regulación de la abundancia de proteína de RAS. En suma, esta tesis estudia la posible relevancia de los mecanismos de síntesis y degradación de proteína de RAS en los patrones de mutación en cáncer.

# Preface

In the early 1980s, RAS genes were identified as oncogenes. The point mutation responsible for the transforming properties of RAS was found later in 1982 in a human cancer cell line. This was reported by the Weinberg, Barbacid and Wigler groups. Subsequently, the cancer field started buzzing with the discovery of RAS oncogenes, which marked a shift in human cancer research.

"For a brief moment in 1982, there was the illusion that cancer was as simple as it possibly could be – a normal cell differed from its neoplastic counterpart by one base out of three billion!" (Weinberg, 2014)

Almost 40 years and around 20,000 publications later, RAS continues to be of great interest. Early efforts aiming to develop anti-RAS drugs failed and no effective inhibitor of RAS activity has reached the clinics yet, which has prompted its perception as an "undruggable" cancer target. Although past failures dampened the enthusiasm in the RAS research field, lately, renewed efforts have focused on reaching a deeper understanding of RAS biology and biochemistry, e.g. signal transduction, effector engagement, feedback loops, three-dimensional structures of RAS bound to its effectors as well as RAS gene- and substitution-specific features. Therefore, a tremendous progress has been made in understanding its role in the complexity of cancer. Remarkably, this has resulted in an overwhelming amount of information on RAS biology.
Nevertheless, one of the most intriguing clinical observations remains unsolved: Why are KRAS mutations much more frequent in human cancer compared to NRAS or HRAS mutations? This observation is surprising because the three forms are nearly identical, each of them is capable of cell transformation in different model systems, and, all forms are widely expressed across tissues. In addition, the underlying reasons for the different frequencies of each type of substitution of activating mutations are yet to be fully elucidated.

Multiple factors may be partly responsible for generating different RAS mutation patterns in cancer. My aim during this thesis is to contribute to a more comprehensive understanding of these intriguing observations. Specifically, I focus on two determinants of RAS protein levels. In the first

chapter of this thesis I provide the necessary background to these studies. In the second chapter, I investigate how codon usage influences RAS protein levels and how it may account for the prevalence of KRAS mutants in tumors, in comparison to HRAS and NRAS. In the third chapter, I investigate how RAS protein-protein interactions with its different downstream binding partners affect RAS degradation and how it may contribute to mutation-specific protein levels. Finally, in the last chapter I present a general overview and discussion of the results.

# Contents

**Acknowledgements** ..... **iii**
**Abstract** ..... **vii**
**Resum** ..... **ix**
**Resumen** ..... **xi**
**Preface** ..... **xiii**
**Contents** ..... **xv**
**List of figures** ..... **xix**
**List of tables** ..... **xxiii**
Chapter 1 ..... 1
1 Introduction ..... 1
1.1 Cancer genes ..... 1
1.1.1 Clonal evolution ..... 1
1.1.2 Oncogenes and tumor suppressors ..... 2
1.2 The oncogene RAS ..... 2
1.2.1 RAS a small G protein ..... 3
1.2.2 RAS signaling pathways ..... 4
1.2.2.1 Study of RAS signaling pathways ..... 5
1.2.3 RAS structure ..... 6
1.2.4 RAS sequence ..... 8
1.2.5 RAS post-translational modifications ..... 9
1.2.6 RAS degradation ..... 10
1.2.7 Main biological differences between RAS family members 10
1.2.7.1 Processing and compartmentalization of RAS ..... 11
1.2.7.2 RAS expression ..... 13
1.2.7.3 RAS during development ..... 13
1.2.1 RAS in cancer ..... 14
1.2.1.1 RAS mutations ..... 14
1.2.1.2 Differences between RAS genes ..... 16
1.2.1.3 RAS expression in cancer ..... 18
1.3 Protein synthesis and degradation ..... 19
1.3.1 mRNA translation ..... 21
1.3.1.1 The degeneracy of the code ..... 21
1.3.1.2 Transfer RNA ..... 22
1.3.1.3 tRNA abundance and codon usage ..... 24
1.3.1.4 Quantification of tRNA abundance ..... 24

1.3.1.5 Variability in tRNA abundance ..... 25
1.3.1.6 tRNA biology and RAS in cancer ..... 25
1.3.1.7 Translation and mRNA stability ..... 26
1.3.2 Protein degradation ..... 28
1.4 Structure of this thesis ..... 28
1.4.1 Chapter 2 ..... 29
1.4.2 Chapter 3 ..... 29
Chapter 2 ..... 31
2 Proliferation codon usage facilitates oncogene translation ..... 31
2.1 Summary ..... 31
2.2 Introduction ..... 31
2.3 Results ..... 33
2.3.1 Codon usage of cancer genes ..... 33
2.3.2 Codon usage-specific changes of KRAS protein abundance under different cell states ..... 40
2.3.3 Specific differences in tRNA levels explain differences in KRAS abundance between cell lines ..... 43
2.4 Discussion ..... 53
2.5 Appendix ..... 57
2.6 Methods ..... 60
2.6.1 Data Sources ..... 60
2.6.2 Computational analysis ..... 61
2.6.3 Sample preparation and experimental procedures ..... 63
Chapter 3 ..... 67
3 Implications of point mutations on KRAS protein abundance ..... 67
3.1 Summary ..... 67
3.2 Introduction ..... 67
3.3 Results ..... 69
3.3.1 $KRAS_{WT}$ and oncogenic KRAS have different protein abundance. ..... 69
3.3.2 Investigating possible mechanisms leading to different protein abundance. ..... 72
3.3.2.1 Changes in mRNA levels ..... 73
3.3.2.2 KRAS-mediated stimulation of protein synthesis ..... 76
3.3.2.3 Effect of G12D mutation on KRAS stability ..... 77
3.3.2.4 Effector binding protects from degradation ..... 79
3.3.3 Designed RAS mutants ..... 82
3.3.4 Experimental validation of designed RAS mutants ..... 85

3.3.4.1 Validation of interactions ..... 85
3.3.4.2 Mutant-specific protein abundances ..... 90
3.3.4.3 Mutant-specific protein degradation ..... 92
3.3.5 KRAS-RAF1 interaction increases KRAS protein abundance. ..... 93
3.4 Discussion ..... 95
3.5 Appendix ..... 98
3.5.1 Expression of KRAS mutants in mouse $APC^{-/-}$ intestinal organoids. 98
3.5.2 Microscale thermophoresis experiments. ..... 100
3.5.3 Comparison between EF1 and CMV promoters. ..... 102
3.6 Methods ..... 103
3.6.1 Structural analysis ..... 103
3.6.2 Sample preparation and experimental procedures ..... 104
Chapter 4 ..... 113
4 Discussion ..... 113
4.1 Quantification of oncogene abundances ..... 114
4.2 Codon usage and tRNA ..... 115
4.2.1 Effect of tRNA pool on protein levels ..... 117
4.2.2 Effect of tRNA pool in mRNA stability ..... 119
4.3 RAS genes and levels ..... 119
4.4 Interaction with effectors ..... 120
4.5 KRAS turnover ..... 122
4.6 Concluding remarks ..... 122
**Contributions ..... 125**
**References ..... 127**

# List of figures

Figure 1-1 The RAS pathway. ....4
Figure 1-2 Crystal structure of $HRAS_{WT}$ in complex with GppNHp. ....7
Figure 1-3 Structures of RAS-effector complexes ....7
Figure 1-4 Overview of RAS protein sequence identity....9
Figure 1-5 RAS processing and subcellular localizations. ....12
Figure 1-6 Summary of post-translational modifications of RAS proteins .13
Figure 1-7 Overview of aberrant RAS signaling in cancer. ....15
Figure 1-8 RAS mutations in all cancers. ....17
Figure 1-9 RAS mutations by gene and by substitution. ....18
Figure 1-10 Four fundamental processes involved in gene expression ....20
Figure 1-11 Transfer RNA....23
Figure 1-12 Effects of different tRNA pools and codon choice discussed in this thesis ....27
Figure 2-1 Selected cancer gene families. ....34
Figure 2-2 Average codon and amino acid identity of the eight cancer gene families. ....35
Figure 2-3 PCA projection of the human codon usage....38
Figure 2-4 Covariance mutation count and PC1 ....39
Figure 2-5 Experimental design. ....40
Figure 2-6 $KRAS_{WT}$ and $KRAS_{HRAS}$ protein expression during starvation and non-starvation ....41
Figure 2-7 Comparison of expression between the productive and non-productive (without RBS and ATG) expression cassette. ....42
Figure 2-8 Transcript abundance with and without translation....43
Figure 2-9 $KRAS_{WT}$ and $KRAS_{HRAS}$ protein expression in BJ/hTERT, HEK293 and HeLa. ....44
Figure 2-10 HEK293, HeLa and BJ/hTERT cell count over 96 hours. ....45
Figure 2-11 Control with inverted tags in the expression construct. ....45
Figure 2-12 Comparison of transcript abundance with and without translation. ....46
Figure 2-13 Volcano plot showing relative tRNA differential expression in log2 fold change between HEK293 and HeLa....48
Figure 2-14 tRNA levels associated to $KRAS_{WT}$ and $KRAS_{HRAS}$ codon usage ....49
Figure 2-15 Immunoblot data (from Lampson et al., 2013) ....50

Figure 2-16 Relative codon usage between the most frequently and the less frequently mutated gene..... 52
Figure 2-17 Proliferation- versus differentiation-related codons. ..... 53
Figure 2-18 Based on the three-stage carcinogenesis model for RAS-induced tumors proposed (Sarkisian et al., 2007)..... 55
Figure 2-19 Comparison with HCT116. ..... 58
Figure 2-20 Cancer gene pairs covariance codon usage and mutation counts. ..... 59
Figure 2-21 Correlation of average codon frequencies of proliferation and differentiation gene set with the codon frequencies of RAS genes. ..... 60
Figure 3-1 Example of $KRAS_{WT}$ and $KRAS_{G12D}$ protein abundance in HeLa cells..... 70
Figure 3-2 $KRAS_{WT}$ and $KRAS_{G12D}$ protein abundance in mouse intestinal organoids..... 70
Figure 3-3 KRAS expression in $APC^{-/-}$ $KRAS_{WT}$ and $APC^{-/-}$;$KRAS_{G12D}$ mouse intestinal organoids..... 71
Figure 3-4 Comparison between FLAG-KRAS WT, G12D and G12C in HeLa cells. ..... 72
Figure 3-5 Protein and mRNA levels of FLAG-$KRAS_{WT}$ and FLAG-$KRAS_{G12D}$ in HEK293 cells. ..... 74
Figure 3-6 Protein and mRNA levels of TET when $KRAS_{WT}$ and $KRAS_{G12D}$ are expressed in HEK293 cells..... 75
Figure 3-7 Total protein quantification of HEK293 cell lysates with FLAG-$KRAS_{WT}$ and FLAG-$KRAS_{G12D}$ expression..... 77
Figure 3-8 Protein abundance comparison between effector binding mutant D38A, WT and G12D..... 80
Figure 3-9 Quantification of FLAG-KRAS WT and mutants at four different time points after cycloheximide addition..... 81
Figure 3-10 RAS-effector structures..... 82
Figure 3-11 Water-mediated interaction of RAS residues. ..... 83
Figure 3-12 Designed RAS mutants binding different downstream effectors. ..... 84
Figure 3-13 Overview of the NanoBRET assay principle. ..... 86
Figure 3-14 RAS interactions with effectors measured with different methods..... 88
Figure 3-15 NanoBRET results with $KRAS_{mutants}$-effectors interactions.... 89
Figure 3-16 Expression of FLAG-$KRAS_{mutants}$ in HEK293 cells..... 91

Figure 3-17 Degradation experiment with Y40C-G12D and Y40F-G12D mutants. .....92
Figure 3-18 Effectors co-expression with FLAG-KRAS$_{WT}$.....93
Figure 3-19 Co-expression of FLAG-KRAS with kinase dead RAF1. .....94
Figure 3-20 Expression of FLAG-KRAS$_{mutants}$ in APC$^{-/-}$ intestinal orgnanoids. .....99
Figure 3-21 Intestinal organoids sphere formation assay. .....100
Figure 3-22 Microscale Thermophoresis measurements between HRAS-GppNHp and four different effectors. .....101
Figure 3-23 Change from CMV to EF1 promoter.....102
Figure 3-24 Inducible TetOn3G expression system. .....106

# List of tables

Table 1-1 RAS effectors and the main cellular processes they control. .....5
Table 1-2 Oncogenic potential of RAS mutations at Glycine 12.....16
Table 1-3 The degenerate genetic code. .....22
Table 2-1 Mutation and protein similarity data from the eight cancer gene families. .....36
Table 2-2 Hydro-tRNA sequencing expression data (rpm). .....47
Table 2-3 RT-qPCR primers.....64
Table 3-1 Stability change by G12D mutations. .....78
Table 3-2 Comparison of predicted and measured interactions. .....90
Table 3-3 Main components for media preparation for intestinal organoid culture. .....105
Table 3-4 Primers for RT-qPCR .....111

# Chapter 1

# 1 Introduction

## 1.1 Cancer genes

The 1980s and the 1990s were a milestone regarding the discovery of cancer genes and thus, the genetic basis of cancer development. The identification and characterization of cancer genes uncovered not only their role in cancerous cells but also shed a light on their function in normal cells. Furthermore, the finding that cancer progression is associated with the accumulation of genetic and epigenetic alterations and that the identity of mutant cancer-causing genes varied dramatically between cancer types, highlighted the complexity and diversity of this disease. Importantly, molecular cancer biology research adopted a model of clonal evolution to better understand the accumulation of mutations in specific genes.

### 1.1.1 Clonal evolution

The concept of clonal evolution of tumors emerged from the view of cancer as a multistep process with the accumulation of genetic changes and selection acting on them. Peter Nowell described in 1976 that during tumorigenesis cells acquire a series of genetic changes which in some cases provide a growth advantage that result in the transformation of normal cells into cancer cells (Nowell, 1976). Nowell proposed that the first alterations result in cell proliferation, allowing for a selective growth advantage. Successive rounds of clonal selection would cause cell populations to have more severe phenotypes. Subsequently, different types of genetic alterations were identified. For instance, simple changes in the DNA coding sequence, such as single-base mutations or more drastic deletions or rearrangements of the DNA, may either abolish or significantly enhance a protein function or expression. Still, the majority of alterations are considered to be evolutionary neutral, not affecting the cancer cell growth phenotype. Hence, building upon

the model of cancer as an evolutionary system, tools from population genetics have been employed to leverage the increasing availability of sequenced cancer genomes to distinguish between driver mutations (those conferring a growth phenotype to the cell) versus passenger mutations (those not undergoing selection).

### 1.1.2 Oncogenes and tumor suppressors

Having discovered the presence of cancer genes, attention turned immediately to study their identity, the function of their protein products and the mechanisms by which genetic alterations modify their function or expression. Accordingly, cancer driver genes can be distinguished into oncogenes and tumor suppressor genes depending on the effect that mutations have on the activity of the cancer gene. Activating mutations, duplications or other genetic events, increase the activity and the expression of certain genes in cancer, so-called oncogenes and therefore correspond to gain-of-function mutations. In healthy cells, these genes promote and maintain normal cell growth. On the other hand, alterations that impair the activity of genes that control cell growth and respond to DNA damage, called tumor suppressor genes, correspond to loss-of-function mutations (Weinberg, 2006). Therefore, both oncogenes and tumor suppressors play a crucial role through opposite mechanisms which can cause uncontrolled cell proliferation.
Cancer formation is a multistep process in which the acquisition of phenotypic traits (e.g. sustaining proliferative signaling or resisting cell death; Hanahan, Weinberg, and Francisco 2000) is mirrored by the consecutive genetic or epigenetic activation of oncogenes and inactivation of tumor suppressors. For example, Eric Fearon and Bert Vogelstein published a highly influential paper regarding the multistep model of tumorigenesis, where they described the consistent and sequential acquisition of alterations in colorectal cancer, starting with the loss of APC followed by activating mutations in the oncogene RAS and then the loss of the tumor suppressor gene TP53 (Fearon and Volgelstein, 1990).

In the following part of the Introduction I will focus on the biology of RAS, one of the most studied oncogenes and the main focus of this thesis.

## 1.2 The oncogene RAS

The identification of mutationally activated RAS genes in human cancer cell lines in 1982 started an endeavor to study its structure, biochemistry, and

biology. With the availability of a large number of sequenced cancer genomes, it turned out to be one of the most frequently mutated oncogenes (~30% of all cancers). Therefore, efforts to characterize RAS continue to this day with the ultimate goal to develop anti-RAS drugs for cancer treatment. Yet, this is a very challenging goal because RAS lacks hydrophobic pockets on the surface allowing for small molecules to bind with high affinity. Moreover, RAS is a member of a large family with similar GTP/GDP binding domains, which complicates the development of highly specific drugs. As a result, RAS won the reputation of being an 'undruggable' protein (Ledford, 2015).

Years of RAS research, have established two broad scientific areas:

1. Discovery of RAS as the first oncogene led to the subsequent discovery of many other genes mutated in cancer and therefore played a significant role in discerning the origin of cancer at the molecular level.
2. Understanding the biochemical mechanisms by which RAS facilitates signal transduction both in a healthy context and in cancer. This revealed that aberrant RAS activation contributes to several of the 'Hallmarks of Cancer' (Hanahan and Weinberg, 2011).

In the following paragraphs, I will summarize the main biological characteristics of RAS.

### 1.2.1 RAS a small G protein

RAS proteins are the best characterized members of a large superfamily of small GTPases that regulate key cellular processes. RAS GTPases switch between the GDP-bound inactive and GTP-bound active states with the help of guanine nucleotide exchange factors (GEFs), which promote activation, and GTPase-activating proteins (GAPs), which inactivate RAS by catalyzing GTP hydrolysis (Wennerberg, 2005). Once activated, RAS-GTP changes its conformation, and binds and activates downstream effectors with distinct functions. RAS proteins act as signal transducers, converting upstream extracellular signals to downstream intracellular effects. This binary switch between an active and an inactive state allows for a tight regulation of signal transduction.

### 1.2.2 RAS signaling pathways

RAS resides on the inner leaflet of the plasma membrane. Activation of cell surface receptors stimulates signals that lead to the activation of RAS by GEFs. RAS serves to amplify and diversify the incoming information, initiating a wide variety of downstream signaling cascades through the recruitment and binding of its effectors at the membrane, leading to changes in many different cellular phenotypes. RAS signaling is not limited to the plasma membrane but also occurs from endomembranes (Chiu et al., 2002). Effectors usually contain a RAS binding domain (RBD) (Emerson et al., 1995). The three downstream pathways RAF/MAPK, PI3K/AKT, and RALGEF are the best described canonical pathways that control cell proliferation and cell survival (Figure 1-1). Therefore, activating cancer mutations in RAS lead to increased cell proliferation.

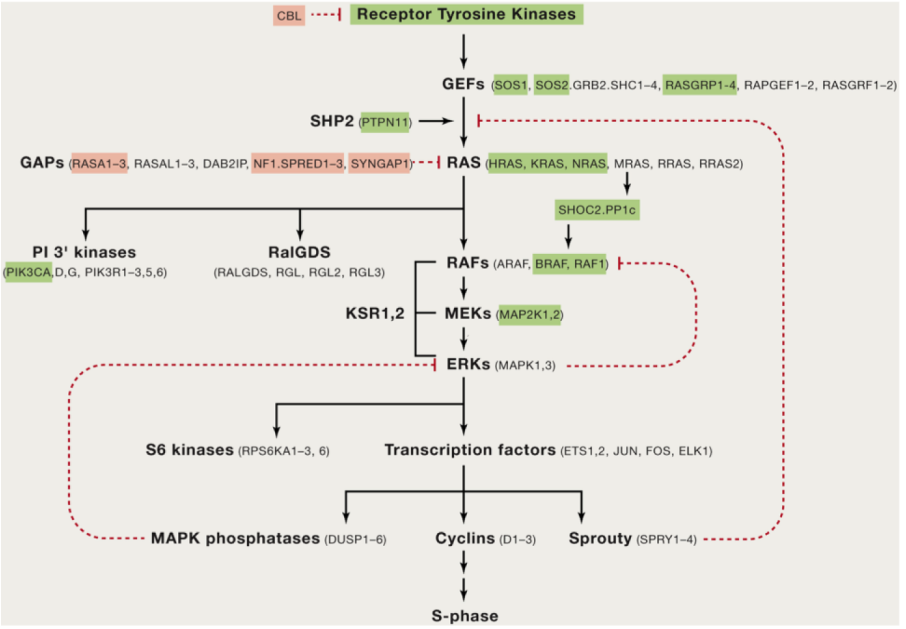



**Figure 1-1 The RAS pathway.**

*Simplified representation of the RAS pathway with the three main canonical downstream effectors (PI3Ks, RALGDS, and RAFs). The RAF/MAPK signaling cascade is displayed and cancer genes are highlighted as follows: genes in red are frequently deleted while genes in green are frequently activated by mutations in human cancers Figure extracted from (Simanshu et al., 2017).*

#### 1.2.2.1 Study of RAS signaling pathways

RAS signal transduction involves physical interaction with a wide spectrum of downstream effectors, e.g. RAF1, RALGDS, PI3Kα, PLCε, and RASSF5, each of them controlling different cellular processes (Table 1-1). Although effectors show a very limited sequence similarity, they all bind to RAS through the common RBD domain to the effector domain of RAS (see Section 1.2.3). Previously, our group has been able to predict RAS interactors using a computational approach based on structural homology modeling using different RAS-effector structures as templates and pull-down experiments (Kiel et al. 2005).

| Protein | Substrate | Cellular process | Reference |
|---|---|---|---|
| **RAF1** | MEK1 and MEK2 Ser/Tyr kinases | Gene expression and cell proliferation | (Desideri et al., 2015) |
| **p110α,β,δ,γ** | PtdIns(4,5)P2 | Cell growth, cell survival, cytoskeleton reorganization and metabolism | (Castellano and Downward, 2011) |
| **PLCε** | PtdIns(4,5)P2 | Increase of calcium levels | (Bunney and Katan, 2006) |
| **RALGDS** | RalA/RalB small GTPases | Vesicular trafficking and migration | (Neel et al., 2011) |
| **RASSF5** | MST1 Ser/Thr kinase | Apoptosis, cell cycle arrest | (Feig and Buchsbaum, 2002) |

**Table 1-1 RAS effectors and the main cellular processes they control.**

Accordingly, RAS has been described as a protein of central importance acting as a signaling hub by binding different effectors. These binding partners might not interact simultaneously with RAS due to the steric hindrance at the RBD. Therefore, a key question is how RAS and its effectors act together to transduce signals within the cell and how cell context-specific variability influences RAS signaling. Previously, our group found that the protein abundance of RAS and its effectors contributes to context-dependent signaling through competition for binding (Kiel et al. 2013).

Effector domain mutants, i.e. mutations at the interaction interface of RAS preventing the binding with its partners, have been used as a tool to study RAS signaling in different cellular contexts. For instance, using a combination of random mutagenesis in combination with yeast-two-hybrid protein-protein interaction screens, showed that residues T35, E37, D38, and Y40 are key residues for the binding of RAS to different effectors (Khosravi-far et al., 1996; White et al., 1995). Mutations at these positions have been used for studying the different signaling pathways controlled by RAS activation. Interestingly, different mutations affect distinct sets of effectors. For example, the E37G mutation hinders the interaction with PI3K and RAF1 but does not affect the interaction with RALGDS and RASSF5. D38 has important contacts with the RBD of the different effectors. The D38A mutation, for instance, results mainly in the loss of RALGDS, RASSF5, and PI3K binding but retains some RAF1 interaction. Finally, the Y40C mutant has been used to selectively activate PI3K and at the same time preventing the interaction with RAF1, RALGDS, RASSF5 (Rodriguez-Viciana et al., 1997).

### 1.2.3 RAS structure

The first crystal structure of RAS was solved in 1988 (de Vos et al., 1988). Currently, more than 300 entries in the Protein Data Bank (PDB) provide structures of RAS in nucleotide-free, GTP-, GDP-, various effector-, GEF- and GAP-bound states. The structures of RAS proteins contain 6 β-sheets flanked by 5 α-helices interconnected by 10 loops as illustrated in Figure 1-2 (Santos and Nebreda, 1989; Wittinghofer and Vetter, 2011). The G domain of the proteins (residues 1-166) is generally divided into two halves. The effector lobe (residues 1-86) through which the contacts between RAS and the nucleotide and between RAS and the effectors are established. This part is composed by the P-loop (phosphate-binding loop, residues 10-17), switch I (residues 32-38) and switch II (residues 59-67) and is followed by the allosteric lobe (residues 87-166) which is involved in establishing the contacts with the membrane together with the hypervariable region (residues 167-188/189). The activation of RAS induces a GTP-specific conformation through the movement of switch I and II which allows the interaction with the effectors with a common interacting interface (Figure 1-3).

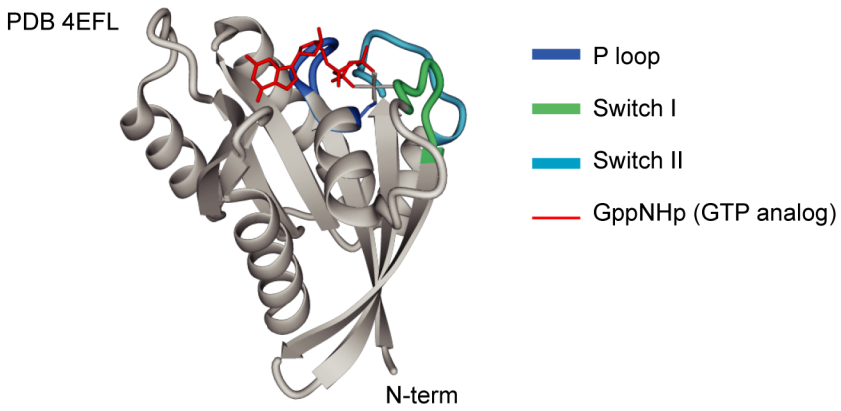



**Figure 1-2 Crystal structure of HRAS$_{WT}$ in complex with GppNHp.**

*Structure available in the PDB (ID: 4EFL). The regions of the P-loop switch I and switch II change in conformation during GDP–GTP cycling and correspond to the interaction interface with the effectors. The figure has been created with Yasara (Krieger and Vriend, 2014).*

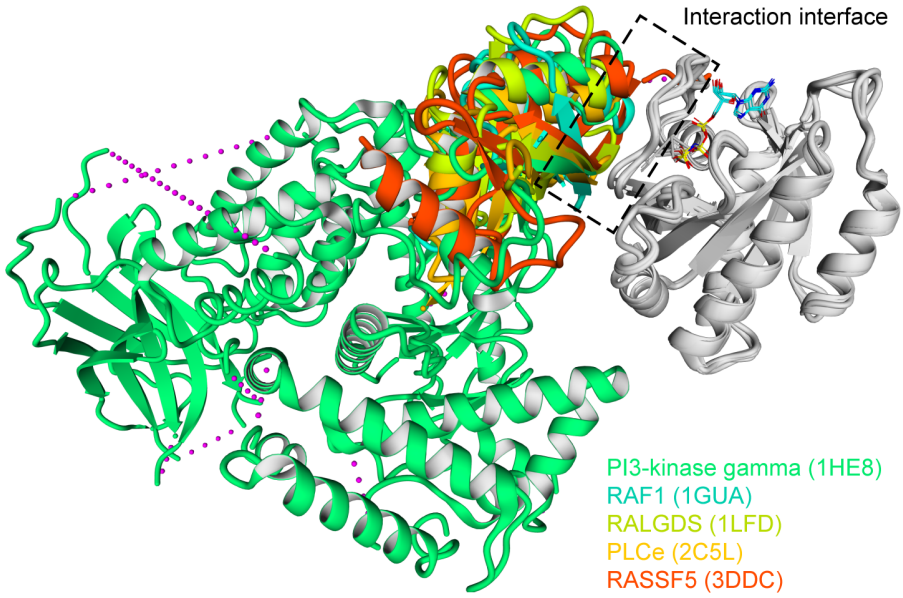



**Figure 1-3 Structures of RAS-effector complexes**

*The five 3D structures of RAS-effector complexes studied in this thesis. An overlaid structure in ribbon representation illustrates that the five effectors interact with RAS through the same interface. The effectors are represented in different colors and RAS in grey. The figure has been created with Yasara (Krieger and Vriend, 2014).*

### 1.2.4 RAS sequence

The human RAS family contains 36 members (Vigil et al., 2010). Here, we focus on a subgroup of three highly related proteins within this family, encoded by three different genes: KRAS, NRAS, and HRAS. HRAS is 189 amino acids long while NRAS and KRAS are 188 amino acids in length. The KRAS transcript can be spliced into two isoforms (KRAS4A and KRAS4B). Even though KRAS4A is expressed together with KRAS4B in some tissues (Tsai et al., 2015), KRAS4B is the main splice variant expressed in most tissues, which contributes to the fact that KRAS cancer research has traditionally focused on KRAS4B. The three RAS proteins show an overall sequence identity of 85% (Figure 1-4), with the effector lobe being completely conserved, composed of 86 identical amino acids. More specifically, the switch regions, which interact with downstream effectors, are identical. Thus, HRAS, KRAS and NRAS can activate the canonical signaling pathways RAF/MAPK, PI3K/AKT, and RALGEF. On the other hand, the regions with the highest sequence divergence are located at the C-terminal hypervariable region (see Figure 1-4), which enables membrane targeting and differential localization of the RAS proteins. The C-terminal includes the CAAX box, which is the target of post-translational modifications and allows the correct localization of RAS at the plasma- and endomembranes (see Section 1.2.5).

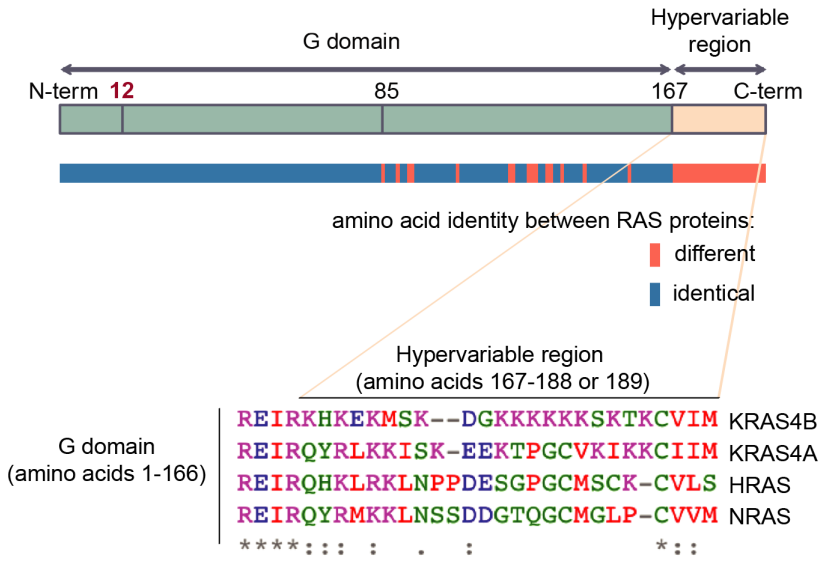



**Figure 1-4 Overview of RAS protein sequence identity.**

*RAS proteins are highly conserved throughout the amino-terminal domain (1-167), which contains the GTP-GDP binding and interaction sites for effector proteins but they differ in the carboxy-terminal domain. The alignment of the hypervariable carboxyl terminus of RAS proteins is highlighted, including both splice variants of KRAS (generated with ClustalW).*

In contrast to the high similarity on the amino acid level, there is a larger variation of the underlying DNA sequence: the codon composition differs highly between the family members (<20% codon identity) (Lampson et al., 2013). This intriguing property of the RAS family will be further discussed in Chapter 2.

### 1.2.5 RAS post-translational modifications

Post-translational modifications of RAS proteins contribute to the regulation of their cellular trafficking and modulate their protein activity. The most important modifications are the farnesylation of the CAAX motif and palmitoylation. These modifications have been shown to allow RAS to associate with membranes from various cellular compartments, which is required for RAS signaling and thus is essential for RAS protein function. Therefore, these modifications have been largely studied as they present targets for the development of inhibitors that limit dysregulated RAS activity

in cancer. Interestingly, these modifications are different for each RAS protein (see Section 1.2.7.1).
More recently studied modifications include phosphorylation, nitrosylation, and ubiquitylation. For example, cysteine 118 can be nitrosylated when exposed to nitric oxide which might enhance RAS activation. Therefore, it has been speculated that RAS activity can be controlled by redox reactions. Additionally, HRAS (Jura et al., 2006) and KRAS (Sasaki et al., 2011) were shown to be ubiquitylated, which modulates their signaling potential. However, the conditions and physiological significance of RAS ubiquitylation remain to be elucidated. RAS can also be modified by enzymes coming from pathogenic bacteria. For instance, the enzyme ExoS from *Pseudomonas aeruginosa,* ADP-ribosylates Arginine 41 and 128 on RAS, which attenuates GTP loading and inhibits the interaction with effectors (Ganesan et al., 1999). Thus, RAS proteins are substrates for a wide variety of post-translational modifications.

### 1.2.6 RAS degradation

The mode of RAS degradation remains controversial. KRAS has been reported to undergo lysosomal degradation as it has been found to signal from but also to be degraded in late endosomes and in the lysosome (Lu et al., 2009). More recently, it has been reported that RAS is polyubiquitinated through the E3 ubiquitin ligase enzyme (β-TrCP) leading to proteasomal degradation of RAS (Jeong et al., 2012; Shukla et al., 2014) (see Section 1.3). The different reported modes of RAS degradation may be due to multiple inducers of RAS degradation, but more detailed mechanisms remain to be studied.

### 1.2.7 Main biological differences between RAS family members

RAS genes are ubiquitously expressed across tissues and share a high sequence similarity as mentioned in Section 1.2.4 (Figure 1-4), therefore, it remains difficult to discern which are the specific roles of each RAS protein and which are redundant functions in physiological or pathological processes. Functional specificity for the different RAS proteins is supported by the distinct processing and cellular localization, their role during embryogenesis and their patters of expression.

#### 1.2.7.1 Processing and compartmentalization of RAS

The possibility of functional specificity for each distinct RAS protein is supported by the observation of the distinct amino acid sequences at the C-terminal hypervariable region in human but also in other species (Santos and Nebreda, 1989). As mentioned above the RAS hypervariable region promotes membrane binding, suggesting that this might contribute to differential localization of the RAS proteins. Indeed, the different RAS proteins show a varying degree of modifications in this C-terminal region, which has been shown to contribute to the specific RAS localization at the plasma membrane and endomembranes (Cox et al., 2015; Herrero et al., 2016). All three RAS are farnesylated to induce binding to the endoplasmic reticulum. Then, RAS undergoes the removal of the last three amino acids of the CAAX box. HRAS and NRAS need a second modification to bind to the membrane. This is the palmitoylation of one (NRAS) or two (HRAS) cysteines. Both HRAS and NRAS undergo a dynamic cycle consisting of palmitoylation at the Golgi and depalmitoylation at the membrane. For KRAS there is no need for a second modification because it is composed of lysines that confer a positive charge facilitating the electrostatic interaction with the negatively charged lipids of the membrane (Figure 1-5) (Hancock et al., 1990). Recently, it has been shown that KRAS goes as well through a dynamic process, with its delivery to recycling endosomes and its transport back to the plasma membrane (Schmick et al., 2014).

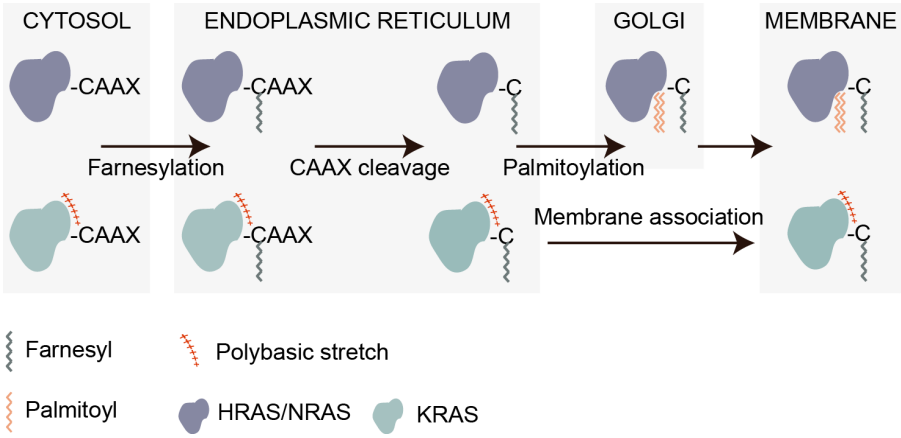



**Figure 1-5 RAS processing and subcellular localizations.**

*RAS processing changes the CAAX motif in the C-terminal hypervariable region, where RAS proteins diverge in sequence. The farnesylation of the CAAX cysteine allows the interaction of RAS with the ER membrane and is followed by the proteolysis of the final AAX amino acids. The palmitoylation of HRAS (two cysteines) and NRAS (one cysteine) at the Golgi allows for stable binding to the membrane. KRAS polybasic stretch promotes electrostatic interaction with the membrane.*

Despite the intensive study of RAS proteins little is known about the RAS protein-specific differences for other types of modifications. Some studied modifications include, for example, the acetylation at Lysine 104 of KRAS. This modification can perturb the GEF activity and thus reduce KRAS activation (Yang et al., 2012). Also, the phosphorylation of KRAS4B at Serine 181 can neutralize the positive charge of the polybasic stretch and therefore inhibit the association to the membrane (Alvarez-Moya et al., 2010). In the case of HRAS, a prolyl isomerase acts on Proline 179 and promotes HRAS depalmitoylation (Ahearn et al., 2011). Finally, Figure 1-6 summarizes the known modifications of the different RAS proteins and the specific residues affected.

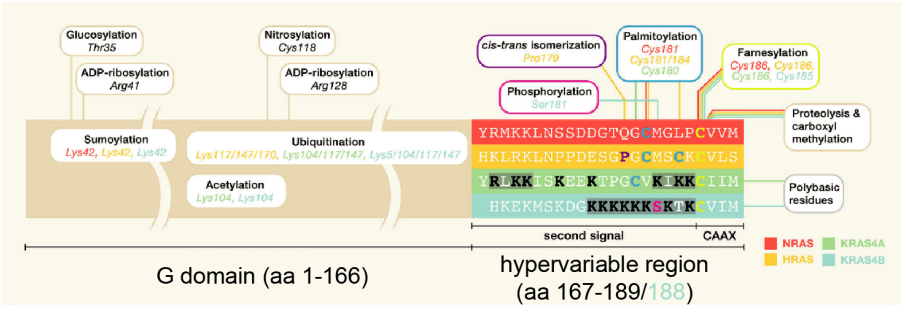



**Figure 1-6 Summary of post-translational modifications of RAS proteins**

*The figure represents the post-translational modifications of RAS proteins. Color-coding distinguishes each of the proteins and the corresponding residues modified. Polybasic lysines of KRAS4A and KRAS4B are indicated by the black color. Figure extracted from (Ahearn et al., 2018)*

### 1.2.7.2 RAS expression

Differences in RAS expression levels might also contribute to their functional specificity. Although KRAS, HRAS, and NRAS are ubiquitously expressed in human and mouse (Fiorucci and Hall, 1988), their relative abundance varies in different cell types and developmental stages. For example, KRAS mRNA levels are in general the lowest of the three in several human tissues (Consortium, 2013). On the other hand, recently it has been shown that during the different developmental stages KRAS4B is the most abundant transcript followed by NRAS, KRAS4A and finally HRAS (Newlaczyl et al., 2017).
RAS expression differences have been also identified at the level of translation. Unlike HRAS, the KRAS coding DNA sequence has a high frequency of non-optimal codons, resulting in low efficiency of KRAS protein translation in comparison to HRAS (Lampson et al., 2013). Therefore, codon usage might underlie functional differences between RAS members.

### 1.2.7.3 RAS during development

More intriguing evidence of the different functional roles of the RAS genes comes from work with knockout mice which showed that HRAS and NRAS knockouts developed normally with no detrimental impact on long term survival. KRAS knockout mice, in contrast, are not viable, indicating that only KRAS activity is necessary during embryogenesis. (Esteban et al., 2001; Koera et al., 1997; Umanoff et al., 1995). Interestingly, the knock-in of HRAS in the locus of KRAS resulted in embryo survival without impairment during development (Potenza et al., 2005). This result indicates that KRAS-specific

expression patterns, rather than unique functions, are critical in embryogenesis.

Finally, additional evidence for KRAS-, HRAS- or NRAS-specific roles came from the observation of different mutation frequencies of RAS in cancer. This aspect will be described in the next section.

### 1.2.1 RAS in cancer

Two mechanisms can lead to an advantageous level of RAS signaling within a cancer cell, contributing to the acquisition of a cellular growth advantage:
- Modulated activity of the RAS protein
- Modulated abundance of the RAS protein

In the following paragraphs, I will focus on RAS activating mutations and RAS expression levels in cancer. Particular emphasis is placed on the differences between KRAS, NRAS, and HRAS, as this comparison is linked to the first project presented in this thesis.

#### 1.2.1.1 RAS mutations

The tight regulation of RAS signaling can be evaded in 30% of all human cancers by activating mutations in RAS (Prior et al., 2012), which lock RAS in the active, GTP-bound state as illustrated in Figure 1-7. Most of the observed point mutations alter amino acids G12, G13 or Q61 (Hobbs et al., 2016), leading to a loss of the RAS intrinsic and GAP-stimulated GTPase activity. The GTP-bound active state of RAS is well established to be oncogenic. The mutation of the glycine in positions 12 or 13 (G12/13) introduces a side chain that prevents the GAP protein from accelerating GTP hydrolysis (Ahmadian et al., 1997). The glutamine 61 (Q61) catalyzes the GTP hydrolysis reaction (Krengel et al., 1990). Thus, mutations in these residues impair GTP hydrolysis and result in a constitutively active RAS protein.

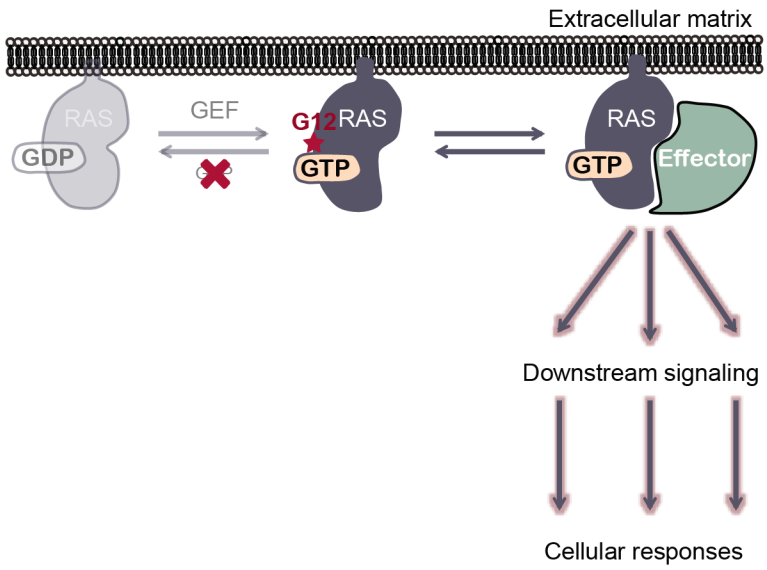



**Figure 1-7 Overview of aberrant RAS signaling in cancer.**

*GEFs are recruited to promote the exchange of GDP for GTP on RAS, generating an active GTP-bound RAS. GAPs accelerate the GTPase activity of RAS, converting RAS back to the inactive state. Oncogenic mutations in RAS proteins render the protein constitutively activating multiple downstream signaling pathways.*

Interestingly, there are studies showing that both the positions and the type of substitutions at each of the three missense mutation hotspots have distinct biochemical and oncogenic properties (Burd et al., 2014; Hunter et al., 2015; Ihle et al., 2012) (Table 1-2). The RAS mutants G12V and G12D, for example, have been described as having a low binding affinity for the downstream effector RAF whereas other mutants exhibit a higher affinity (G12C, G13D, and Q61L) (Hunter et al., 2015).

**Table 1** Transformation properties of *ras* mutants

| Position 12 codon | Activated genes | Focus formation | Agar growth |
|---|---|---|---|
| Gly | c-Ha-*ras*1 (refs 16–19), c-Ki-*ras*2 (ref. 41), N-*ras* (ref. 15) | – | – |
| Ile | | +++ | + |
| Val | T24/EJ Ha-*ras*1 (refs 16–19), SW480 Ki-*ras*2 (ref. 21) | +++ | + |
| Leu | | +++ | + |
| Ser | v-Ki-*ras* (ref. 12) | ++ | + |
| Thr | | +++ | + |
| Ala | | ++ | + |
| Met | | ++ | + |
| Cys | CaLu Ki-*ras*2 (refs 20, 21) PR371 Ki-*ras*2 (ref. 23) | ++ | + |
| Tyr | | ++ | + |
| Phe | | ++ | + |
| Trp | | ++ | + |
| Pro | | – | – |
| Lys | v-BALB-*ras* (ref. 17) | + | + |
| Arg | v-Ha-*ras* (ref. 11), v-Ra-*ras* (ref. 26), A1698, A2182 Ki-*ras*2 (ref. 24) | +++ | + |
| His | | ++ | + |
| Asp | 134-51 Ha-*ras*1 (ref. 22) | ++ | + |
| Glu | NMU-Ha-*ras* (ref. 25) | ++ | + |
| Asn | | ++ | + |
| Gln | | + | + |

Focus formation was determined by the ability of *ras*-cotransfected Rat-1 cells to form foci (see Fig. 3 legend). Agar growth was assessed by plating individual *ras*-cotransfected subclones, or cotransfected mass cultures, into soft agar[51] at two densites ($4 \times 10^3$ and $2 \times 10^4$ per 60 mm dish) and scoring colonies of >30 cells after 2–3 weeks.

**Table 1-2 Oncogenic potential of RAS mutations at Glycine 12**

*Focus formation assay to determine the oncogenic potential of RAS mutants. Table extracted from (Seeburg et al., 1984)*

#### 1.2.1.2 Differences between RAS genes

Historically, the majority of biochemical and structural studies of RAS have focused on HRAS as it was the first gene that was isolated. However, HRAS is the least frequently mutated RAS gene in human cancers, whereas KRAS is the predominantly mutated one, followed by NRAS (Figure 1-8).

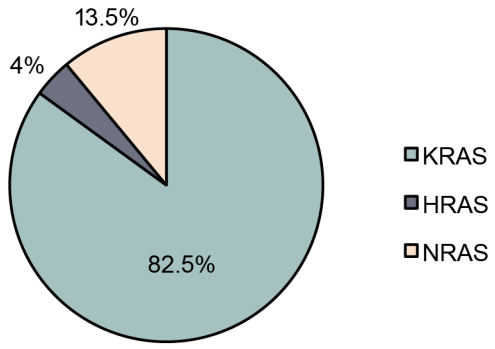



**Figure 1-8 RAS mutations in all cancers.**

*Percentages based on coding mutations reported in COSMIC v90.*

Indeed, the mutation frequency of each RAS gene varies widely in different cancer types. In pancreatic ductal adenocarcinoma, there is a near 100% frequency of KRAS mutations and 85% in colorectal adenocarcinoma. This contrasts with melanoma where NRAS comprises 94% of RAS mutations compared to HRAS and KRAS. Also, KRAS and NRAS mutations are found at equivalent frequencies in multiple myeloma. Conversely, HRAS mutations are prevalent in head and neck squamous cell carcinomas (86%) in comparison to NRAS and KRAS (Hobbs et al., 2016).

RAS genes can be also distinguished by their striking differences in the mutation frequency at each of the three hotspots (G12, G13, and Q61) as well as in the type of substitution. For instance, G12 mutations are prevalent in KRAS whereas Q61 mutations are rare in this protein. In contrast, Q61 is predominantly mutated in NRAS. Also, the frequencies of the substitutions are not uniform, G12D is predominant in KRAS in comparison to G12C and G12V and G12V mutations are more common in HRAS than G12D (Figure 1-9). The G13D mutant appears mainly for KRAS, while observed rarely in HRAS. Due to the high similarity of RAS proteins, the source of the mutation bias remains unclear and might hint at different roles of the RAS genes within the cell.

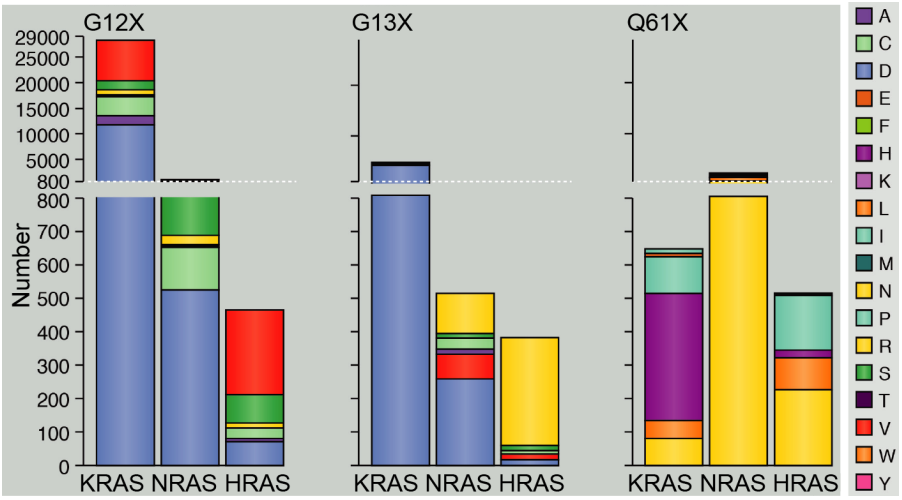



**Figure 1-9 RAS mutations by gene and by substitution.**

*Figure extracted from (Hobbs et al., 2016), COSMIC v75.*

### 1.2.1.3 RAS expression in cancer

Any event that affects RAS protein levels could influence the response of the cell. Accordingly, the level of oncogenic RAS expression is a critical determinant of tumorigenicity (Elenbaas et al., 2001; Li et al., 2018). However, it remains difficult to elucidate which is the ideal level of RAS that gives a growth advantage to the cancer cell. For instance, a high level of oncogenic RAS can lead to proliferation but can also inhibit tumorigenesis by inducing growth arrest (Serrano et al., 1997). On the other hand, very low levels would not lead to cell proliferation. This has been recently described as the "sweet spot" of oncogenic RAS signaling (Li et al., 2018).

Why the frequency of RAS mutations varies widely across cancer types remains largely unresolved. One classical assumption is that the mutated genes tend to cause disease in the tissues in which they are highly expressed (Lage et al., 2008). This requires that we understand not only the physiological expression patterns of RAS, but also how RAS expression levels support cancer initiation and progression (in a RAS gene-specific manner).

As mentioned above, RAS genes have different mutation frequencies in cancer. Therefore, we could expect a correlation between mutation frequencies and expression levels. A study using human fetal tissues and cell lines showed that the three RAS transcripts are ubiquitously expressed, however they did not find a correlation between RAS gene-specific mRNA

levels and RAS gene mutation frequencies (Fiorucci and Hall, 1988). Also, the preferential mutation of KRAS in colorectal carcinoma cannot be explained by differences on the transcript abundance because both NRAS and KRAS are expressed in mouse colon cells (Haigis et al., 2008) and human colon (GTEx Human Transcriptome Atlas), but KRAS is more frequently mutated.

Regarding protein abundances there are fewer studies due to technical limitations caused by the high similarity between the RAS proteins, which complicates their differentiation by mass spectrometry or the use of specific antibodies for immunoblotting. Only recently, an exhaustive study of specific RAS antibodies was published (Waters et al., 2017) that would facilitate quantification with immunoblotting assays. Previously, Omerovic et al., 2008 performed KRAS, HRAS and NRAS protein quantification in a panel of cancer cell lines, using a pan-RAS antibody (recognize all RAS proteins) following specific knockdown of each of the three genes independently. In this study, KRAS and NRAS are more abundant in comparison to HRAS. Later, a quantitative proteomics experiment with colorectal cancer cell lines concluded that KRAS is more abundant than HRAS and NRAS (Mageean et al., 2015). On the other hand, as mentioned previously, it has been described that the coding sequence of KRAS is composed by non-optimal codons (codons decoded by low abundant tRNAs), which leads to a lower protein abundance of KRAS in comparison to HRAS when exogenously expressed (Lampson et al., 2013). These results suggest that KRAS could be the least abundant protein and prompted the hypothesis that, since mutant and high levels of RAS causes cell cycle arrest in normal cells, KRAS is preferentially mutated in cancer because its lower levels are tolerated. Consequently, the observations regarding KRAS, NRAS, HRAS protein abundance are contradictory with the codon usage.

Work within this thesis examines how protein translation may control RAS protein levels. Therefore, the following section will describe aspects of translation.

## 1.3 Protein synthesis and degradation

During the last 30 years, the central dogma of biology (Crick, 1970) has been studied and complemented by detailed biological mechanisms. In this thesis, I will present the results concerning the translation of mRNA into protein and protein degradation. To better explain how this fit into the general picture, I

will first briefly summarize the interplay of the molecular species DNA, RNA, and protein (Figure 1-10).
Gene expression is a highly complex process, involving four main different steps (Figure 1-10) controlled by gene-regulatory events. Briefly, transcripts are the product of the transcription of DNA coding sequences by the RNA polymerase II. Transcription requires an extensive control, specifically transcription factors are known to regulate the transcription of genes in different cellular conditions. For example, transcript levels depend on gene transcription, pre-mRNA processing and mRNA decay.

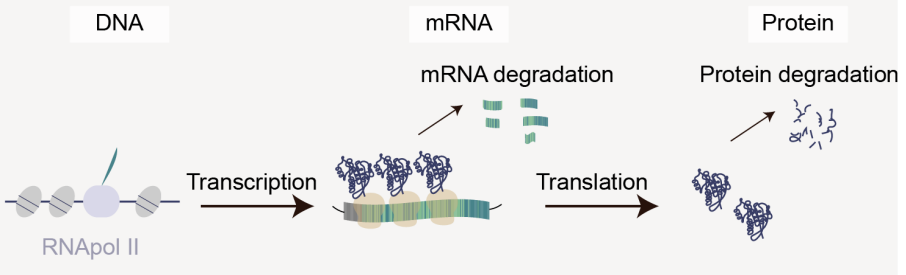



**Figure 1-10 Four fundamental processes involved in gene expression**

*mRNA and protein levels are controlled by both synthesis and degradation rates.*

More specifically, mRNA half-life is controlled by mRNA-binding proteins and non-coding RNAs, such as microRNAs, and can be affected by the translational status of the mRNA (see Section 1.3.1.7). This results in high variations of transcript levels, leading to different mRNA profiles under different biological conditions. Multiple processes contribute to establish the expression level of a protein. These include translation rates that are influenced by the mRNA sequence, e.g. through the codon composition. The translation rate can also be modulated by the binding of proteins to regulatory elements on the transcript, or through the relative availability of transcript and the local supply of charged ribosomes. Finally, protein half-life may influence protein abundances independently of transcript concentrations. Systematic studies have quantified at genomic and proteomic scales the concentration of transcripts and the corresponding proteins as well as mRNA and protein turnover (Lahtvee et al., 2017; Martin-Perez and Villén, 2017; Schwanhäusser et al., 2011). These studies have highlighted the difficulties of quantifying the contribution of each of these regulatory mechanisms to shape gene expression, as a consequence of their closely intertwined relationship.

### 1.3.1 mRNA translation

Translation of mRNA into protein is a central step in gene expression which requires the translation of an mRNA coding sequence into a distinct code composed of amino acids. The degeneracy of the genetic code, which enables a protein to be encoded by many alternative mRNA sequences, has an important role in regulating protein expression (Novoa and Ribas de Pouplana, 2012).

#### 1.3.1.1 The degeneracy of the code

There are 20 different amino acids and each one is encoded by a codon composed of a triplet of nucleotides. There are 64 possible codons, 61 of which correspond to amino acids (excluding stop codons), thus, three times more than the number of amino acids. This redundancy in the genetic code, where most amino acids can be encoded by more than one codon, is also called codon degeneracy. Codons encoding for the same amino acid are described as synonymous codons. During translation, codons of the transcript are paired up with their matching amino acid. This process is mediated by an adapter molecule acting as an interface between the codon and the amino acid, known as transfer RNA (tRNA).

| | U | | | C | | | A | | | G | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| U | UUU | - | Phe | UCU | AGA | Ser | UAU | AUA | Tyr | UGU | ACA | Cys | U |
| | UUC | GAA | | UCC | GGA | | UAC | GUA | | UGC | GCA | | C |
| | UUA | UAA | Leu | UCA | UGA | | UAA | - | Stop | UGA | - | Stop | A |
| | UUG | CAA | | UCG | CGA | | UAG | - | | UGG | CCA | Trp | G |
| C | CUU | AAG | Leu | CCU | AGG | Pro | CAU | - | His | CGU | ACG | Arg | U |
| | CUC | - | | CCC | GGG | | CAC | GUG | | CGC | - | | C |
| | CUA | UAG | | CCA | UGG | | CAA | UUG | Gln | CGA | UCG | | A |
| | CUG | CAG | | CCG | CGG | | CAG | CUG | | CGG | CCG | | G |
| A | AUU | AAU | Ile | ACU | AGU | Thr | AAU | AUU | Asn | AGU | ACT | Ser | U |
| | AUC | GAU | | ACC | - | | AAC | GUU | | AGC | GCU | | C |
| | AUA | UAU | | ACA | UGU | | AAA | UUU | Lys | AGA | UCU | Arg | A |
| | AUG | CAU | Met | ACG | CGU | | AAG | CUU | | AGG | CCU | | G |
| G | GUU | AAC | Val | GCU | AGC | Ala | GAU | AUC | Asp | GGU | - | Gly | U |
| | GUC | - | | GCC | GGC | | GAC | GUC | | GGC | GCC | | C |
| | GUA | UAC | | GCA | UGC | | GAA | UUC | Glu | GGA | UCC | | A |
| | GUG | CAC | | GCG | CGC | | GAG | CUC | | GGG | CCC | | G |

**Table 1-3 The degenerate genetic code.**

*The 1st, 2nd, and 3rd nucleotides of codons are indicated in the left, upper, and right sides of the table, respectively. Each codon is shown in the green column and its potential anticodon in the second column. Codons that lack the corresponding tRNA gene in humans are indicated with a dash. Anticodons highlighted in orange correspond to predicted tRNA genes (Chan and Lowe, 2016).*

### 1.3.1.2 Transfer RNA

Codon recognition is performed by tRNAs, i.e. RNA molecules that can interact with the codons through base pairing. There are multiple copies of tRNA genes (approximately 500) in the human genome to decode 61 codons (Chan and Lowe, 2016). Transcription of tRNA is driven by RNA polymerase III. The pre-tRNA are processed to acquire modifications and mature 5' and 3' ends. The canonical length of tRNAs is 76 nucleotides, with a secondary structure that resembles the shape of a cloverleaf. The interaction with the codon in the mRNA occurs through a triplet of nucleotides termed the anticodon, located at positions 34, 35 and 36, which are in a distant position

from the amino acid attachment site (Figure 1-11). tRNAs carrying the same anticodon belong to same isoacceptor family.

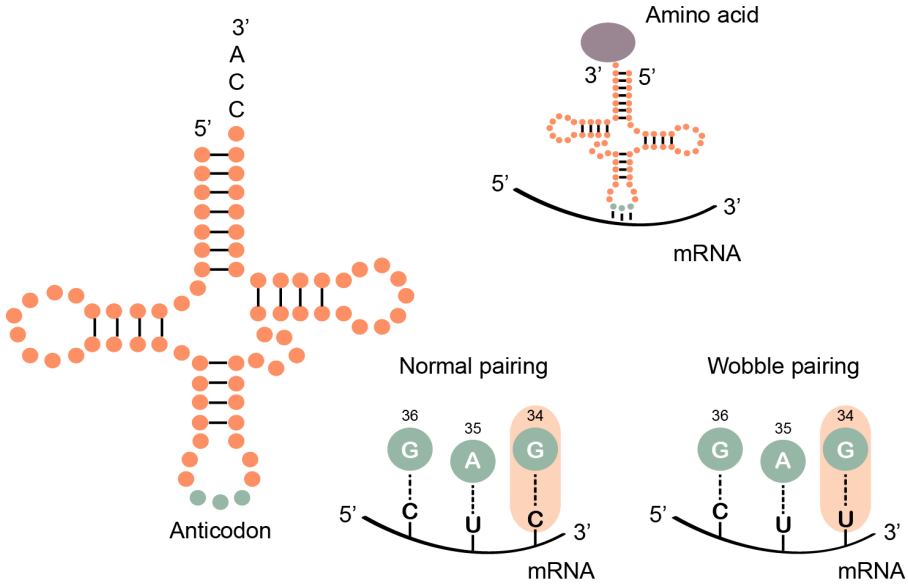



**Figure 1-11 Transfer RNA**

*Overall structure of a tRNA composed of 76 nucleotides, carrying the anticodon in the middle of the anticodon loop and interaction with mRNA. The 5' of the codon pairs with the 3' of the anticodon. tRNAs are the bridge between nucleic acids and amino acids. Example corresponding to identical Leucine tRNAs and their anticodon-codon interaction with Watson-Crick pairing (normal pairing) and non-Watson-Crick pairing (wobble pairing). The nucleotide in position 34 pairs with the third mRNA codon base during ribosome decoding.*

However, not all of the 61 possible codons have the corresponding anticodon tRNAs (Table 1-3). Thus, some codon-anticodon interactions occur through a non-Watson-Crick interaction at position 34, called wobble base pairing (Agris et al., 2007). All tRNA molecules terminate with three nucleotides, CCA, at the 3' terminal site of amino acid attachment. Aminoacyl-tRNA synthetases are the enzymes responsible for linking the cognate amino acid to the tRNA. The nuclear-encoded human tRNAs are among the most extensively post-transcriptionally modified type of RNA, with an average of 11 to 13 modifications per tRNA. Interestingly, a regulatory role has been associated to tRNA modifications as different human diseases have been linked to their absence (Torres et al., 2014).

#### 1.3.1.3 tRNA abundance and codon usage

As mentioned above, tRNA genes are present in multiples copies in the genome. Typically, transcripts composed of codons that correspond to abundant tRNAs would be more highly expressed. Codon usage has been reported to correlate with translation efficiency and protein expression levels in organism such as *Escherichia coli* or *Saccharomyces cerevisiae* (Ikemura, 1985). A popular measure of translation efficiency is the tRNA adaptation index which is based on the tRNA gene copy number and the strength of codon-anticodon interactions (dos Reis et al., 2004). Furthermore, an effective protein synthesis in ensured when the anticodon demand by the cellular transcripts is balanced by the tRNA supply in the cell (Gingold et al., 2012). Since the gene expression levels in higher eukaryotic organisms are variable across different tissues and physiological conditions, the tRNA demand changes dynamically (Rak et al., 2018). Likewise, it has been shown that different functional sets of genes have specific codon usage preferences further increasing the dynamics of tRNA supply and demand (Gingold et al., 2014).

#### 1.3.1.4 Quantification of tRNA abundance

Genomic tRNA gene copy numbers have been used as a proxy of tRNA abundance, with the more frequently used codons being those corresponding to high gene copy number tRNAs. This approximation is somewhat established, but recent studies have shown that tRNA abundance is not static but adapts to cellular dynamics (Dittmar et al., 2006; Gingold et al., 2014; Goodarzi et al., 2016; Torrent et al., 2018). Therefore, quantification methods beyond genomic copy numbers are necessary to detect tRNA abundance variation.

High throughput tRNA quantification represents a technical challenge, due to the high abundance of tRNA modifications and strong tRNA structure. Typically, microarray methods have been used in which tRNAs are hybridized to probes and spotted on a solid surface. More recently, high throughput sequencing methods with higher resolution have been developed (Orioli, 2017). For this thesis project, hydro-tRNA-sequencing has been performed. This method is based on an alkaline hydrolysis treatment that facilitates deep sequencing of complex secondary structures such as tRNAs by generating shorter fragments with less structure that would otherwise interfere with cDNA synthesis necessary for the library preparation for sequencing (Gogakos et al., 2017).

#### 1.3.1.5 Variability in tRNA abundance

Cells must tailor and fine-tune protein levels in order that functionally specialized proteins are produced in the appropriate conditions (such as when it needs to divide or differentiate). In the last years, tRNA expression has been shown to be dynamic and, hence, to contribute to the condition-specificity of gene expression. So far, most of the studies correlating tRNA supply with codon usage and translation efficiency have been carried out in bacteria or unicellular eukaryotes. For instance, it has been shown that yeast selectively regulates tRNA expression during stress conditions in order to produce the proteins required during this state (Torrent et al., 2018). Among the first studies in human, it was established that there are significative differences of tRNA expression between eight tissues analyzed (Dittmar et al., 2006). Later, a first report linking tRNA variability and cancer showed that breast cancer cells selectively upregulate specific tRNAs (Pavon-Eternod et al., 2009). Important novel insights came from one of the studies (which is of crucial importance for the project described in Chapter 2 of this thesis), which showed that cells with different biological characteristics express transcripts with a specific codon usage and that the tRNA pool will be adapted to translate this set of codons with high efficiency (Gingold et al., 2014). They investigated the tRNA abundance of patient tissue samples and human cell lines in different conditions (proliferation, differentiation and starvation) and found that tRNA expression profiles are distinct between these conditions. One of the most interesting findings is that the codon usage of genes related to proliferation versus differentiation/cell cycle arrest correlates with the matching tRNA isoacceptor abundance. In line with these results, it has been shown that two tRNAs (tRNAGluUUC and tRNAArgCCG) are upregulated in metastatic breast cancer cells and consequently lead to an increase of translation efficiency of pro-metastatic transcripts enriched with the two corresponding codons. (Goodarzi et al., 2016). Altogether, cells combine the expression of tRNAs with the codon usage of specific transcripts to optimize or to restrict their translation in different cellular contexts (Figure 1-12).

#### 1.3.1.6 tRNA biology and RAS in cancer

As discussed above, tRNA biology has been linked to tumorigenesis, with a higher abundance of tRNAs in cancer cells likely to provide sufficient tRNA supply for increase protein synthesis in proliferating cells. In addition, specific tRNAs are upregulated during uncontrolled cell growth, hence leading to important changes in expression in cancer cells. Interestingly,

signaling pathways controlled by oncogenes and tumor suppressors have been shown to play an important role in the synthesis of tRNAs through the regulation of the RNA polymerase III (White, 2004). For instance, TORC1, MYC, and RAS controlled pathways have been demonstrated to regulate protein synthesis favoring their ability to promote cell proliferation and drive oncogenic processes. The RAS/MAPK pathway has been shown to stimulate the activity of translation initiation factors (Roux and Topisirovic, 2012). Also, another possible mechanism for the activation of protein synthesis mediated by RAS has been recently described in Drosophila cells, where RAS inhibits MAF1 (Sriskanthadevan-Pirahas et al., 2018). MAF1 is a phospho-protein localized at the nucleus where it represses RNA polymerase III. It is possible that RAS/MAPK activation leads to MAF1 phosphorylation and prevents its nuclear accumulation. Hence, oncogenic RAS signaling might increase tRNA levels and contribute to cancer progression. Nevertheless, individual mechanism by which oncogenes fine-tune tRNA expression still lack understanding.

#### 1.3.1.7 Translation and mRNA stability

Although, many studies have focused on the impact of codon usage on translation efficiency, it has been also shown its influence on mRNA stability (Hanson and Coller, 2017). Indeed, it has been observed that mRNA stability changes depend on the codon usage in *Saccharomyces cerevisiae*, suggesting a connection between transcript levels and translation. More specifically, stable transcripts are enriched in codons corresponding to abundant tRNAs. On the other hand, transcripts with non-frequently used codons have a lower mRNA stability (Presnyak et al., 2015). The yeast helicase Dhh1 is responsible for triggering RNA decay mechanism of the transcripts with slow-moving ribosomes (Radhakrishnan et al., 2016). Recently, codon-dependent mRNA stability has been also described in different types of humans cells (Wu et al., 2019) (Figure 1-12).

Overall, these observations suggest considering tRNA pools as well as codon usage as important contributors to an active and dynamic regulation of gene expression, rather than being only basic modules of protein synthesis (Figure 1-12).

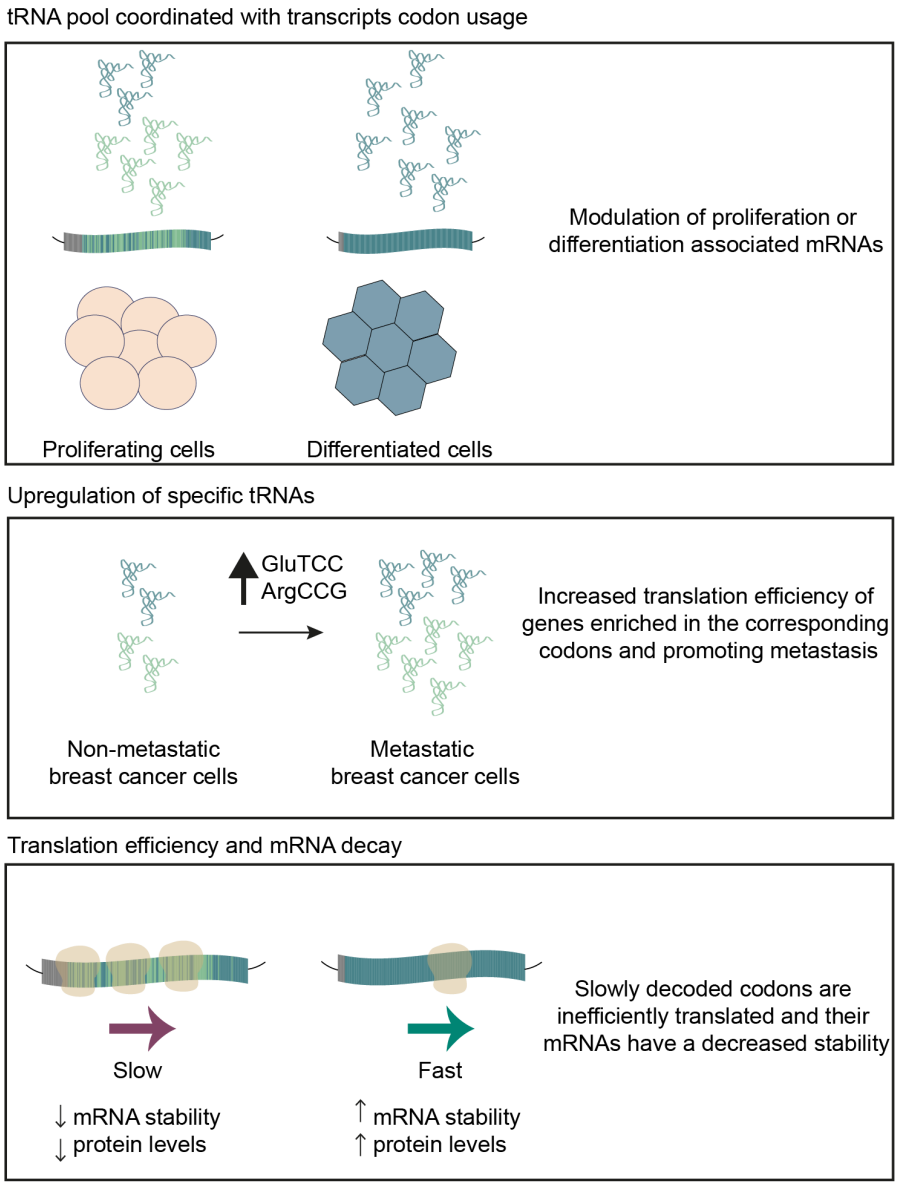



**Figure 1-12 Effects of different tRNA pools and codon choice discussed in this thesis**

*Changes in tRNA expression are linked to proliferation or cancer via the adaptation of the tRNA pool to the codon usage of proliferation-specific mRNAs or oncogenic mRNAs. tRNA molecules are represented with their three-dimensional structure.*

### 1.3.2 Protein degradation

Protein abundance within a cell is determined by the balance between protein synthesis and protein degradation, also known as protein turnover. Proteins are in a continuous dynamic state of synthesis and degradation (Schoenheimer, 1942), which is necessary for maintaining proteostasis. The degradation of proteins is regulated individually and can change under distinct cellular contexts (Lahtvee et al., 2017). Half-lives of proteins vary widely, for instance, many rapidly degraded proteins are regulatory proteins, such as transcription factors, needed to act fast after a specific extracellular signal. The half-life of RAS has been reported to be 12 hours (Shukla et al., 2014). On the other hand, housekeeping proteins such as histones will have a half-life of several weeks, or collagens of 70 years. The protein degradation process also allows to destroy damaged and toxic proteins. In eukaryotic cells there are two distinct pathways by which proteins get degraded: the ubiquitin-proteasome pathway and the lysosomal-mediated proteolysis.

The lysosome is an organelle containing proteases that will cleave proteins into individual amino acids (Xu and Ren, 2015). The proton pumps in the lysosomal membrane allow to transport protons inside the lysosome in order to generate an acidic environment required for an efficient proteolysis. The transport of proteins to the lysosome is done by a variety of means. For example, endocytic vesicles from the cell surface can fuse to the lysosome; this is, for example, a mechanism for degradation of cell surface receptors and thus, for the downregulation of incoming signals.

The second major pathway of protein degradation in eukaryotic cells involves selective protein degradation by the proteasome (Reinstein and Ciechanover, 2006). The proteasome is a large, cylindrical, and multi-subunit protein complex found in the cytosol which degrades proteins that have been tagged with ubiquitin. To accomplish this, E1, E2, and E3 enzymes will covalently attach ubiquitin to a lysine on the protein to be degraded. Additional molecules of ubiquitin are then added and polyubiquitinated proteins are recognized and degraded by the proteasome. This process requires energy in the form of ATP in order to recognize and bind ubiquitinated proteins.

## 1.4 Structure of this thesis

In the next chapters I am going to present the research I have performed in collaboration with colleagues to investigate the determinants of RAS protein abundance.

### 1.4.1 Chapter 2

One yet unresolved question in RAS cancer research that has captured my curiosity is: Why is the oncogene KRAS more often mutated than NRAS and HRAS despite their high amino acid sequence and functional similarity?
As described in this introduction, different explanations have been considered ranging from differences in interactions with other proteins to preferential expression or localization. However, it is intriguing that in the case of KRAS it is found that despite the high amino acid similarity to HRAS and NRAS it has a very different codon usage. Here, I have investigated whether this codon bias plays a role in cell context-specific expression and, thus, in the mutation prevalence of KRAS in cancer.
Chapter 2 corresponds to a manuscript that has been deposited in bioRxiv (doi:10.1101/695957) and has been adapted to be included in this thesis. The manuscript is currently under revision.
The bioinformatic analysis have been done with the help of Marc Weber, Xavier Hernandez-Alias and Martin Schaefer.

### 1.4.2 Chapter 3

RAS signal transduction is achieved by the interaction with proteins with multiple outcomes. The biological relevance of these interactions has been extensively study. Here, we focus in one level of regulation that has not yet been investigated: How are RAS-effector interactions influencing the stability of RAS and therefore its protein abundance?
Specifically, we show that oncogenic mutants and effector domain mutants display variable KRAS protein abundance. We investigate how changes in protein degradation rates caused by differential changes in binding affinities to the effectors might be playing a major role.
This work has been done in collaboration with Christina Kiel and Leandro Radusky.

# Chapter 2

# 2 Proliferation codon usage facilitates oncogene translation

## 2.1 Summary

Tumors evolve under selection for gene mutations that give a growth advantage to the cancer cell. Intriguingly, some cancer genes are more often found mutated in tumors than their closely related family members. Here, we find that for seven oncogene families (RAS, RAF, RAC, RHO, FGFR, COL and AKT), the most prevalent mutated members in cancer have a codon usage characteristic of genes involved in proliferation. For example, while KRAS is found mutated in a large number of patients of different cancer types, HRAS mutations are found only in a small percentage of tumors. To date, the molecular basis for this bias is still unclear. We now report that the codon usage of KRAS is more adapted to be efficiently translated in proliferative cells than the codon usage of HRAS. We also show that the translation efficiency of KRAS varies between cell lines in a manner related to their tRNA expression. Therefore, we propose that codon bias related to cell proliferation contributes to the prevalence of mutations in certain members of oncogene families. Altogether, our study demonstrates that a dynamic translational program contributes to shaping the expression profiles of oncogenes in different cell contexts and that this may account for the mutation bias among the genes.

## 2.2 Introduction

Cancers arise due to mutations that confer a selective growth advantage on cells (Nowell, 1976). These mutations can occur in oncogenes, which when activated by mutations contribute to the cancer proliferation phenotype. Interestingly, oncogenes often have closely related family members that are less frequently mutated in cancer.

The RAS family is a striking example. Activating mutations in KRAS are among the most common mutations in human cancers (Lawrence et al., 2014). KRAS belongs to a family of three genes, the other two being HRAS and NRAS. The encoded proteins share a high sequence identity of 85% and hence similar structure and biochemical properties (Hobbs et al., 2016). However, the reasons for the drastic variation in mutation incidence between the RAS genes remain enigmatic. Significant effort is being invested in studying the molecular differences between the three family members and specifically to understand what is special about KRAS (Apolloni et al., 2002; Drosten et al., 2017; Haigis et al., 2008; Koera et al., 1997; Lampson et al., 2013; Omerovic et al., 2008; Potenza et al., 2005; Prior et al., 2003; Quinlan and Settleman, 2009; Yan et al., 1998). An intriguing observation is that even though RAS proteins are very similar, the codon usage is different, with only 15% of codon identity (Lampson et al., 2013). The nucleotide sequence of KRAS is enriched in rare codons (decoded by low-abundant tRNAs) in comparison to HRAS. This has been linked to a poor translation efficiency of KRAS and a high efficiency for HRAS (Lampson et al., 2013). It has been speculated that these differences in codon usage relate to the imbalance of the mutation frequency within the RAS family: the constitutively activated form of the highly translated HRAS might lead to an over-activation of the MAPK pathway, ultimately leading to oncogene-induced senescence (Bodemann and White, 2013; Pershing et al., 2015).

Codon usage and tRNA abundance are important parameters for fine-tuning protein synthesis. The functional influence of codon optimality and tRNA levels on the efficiency of protein production remains a topic of intense debate (Hanson and Coller, 2017). In recent years, studies have shown that tRNA levels are not static but dynamically regulated in different cellular contexts, leading to changes of the translation efficiency of transcripts depending on their codon composition (Bornelöv et al., 2019; Gingold et al., 2014; Goodarzi et al., 2016; Torrent et al., 2018). In mammalian cells, changes in tRNA abundance have been reported across different cell states, and specifically between healthy and cancer cells (Gingold et al., 2014; Goodarzi et al., 2016). Interestingly, Gingold et al (Gingold et al., 2014) showed that a specific subset of tRNAs are upregulated in proliferating cells, while they are downregulated in differentiated or arrested cells. Additionally, they show that genes that are necessary for cell division have a codon usage

adapted to the tRNA repertoire in proliferative cells. Thus, changes in the expression of specific tRNAs could regulate an entire functional class of genes, for instance proliferative genes, to favor cell growth. Would a cancer cell take advantage of this translational program to modulate the expression of cancer genes to its own growth advantage? Could it be that a dynamic regulation of RAS translation efficiency determines the uneven mutation frequencies across RAS genes? Will this be a general phenomenon in other cancer gene families?

To answer the above questions, we first identify 8 protein families of three members, RAS, RAF, RAC, RHO, FOXA, FGFR, COL and AKT, with high protein sequence similarity and at least one protein being relevant for cancer. We find that in all but one family the codon usage signature of the most frequently mutated gene is characteristic of proliferation-related genes in comparison to its homologous family members. We then study the RAS family composed of KRAS, HRAS and NRAS in detail. We measure how proliferation and quiescent cell states induce codon-dependent changes in KRAS protein levels. Finally, we find that different tRNA expression profiles between cell lines correspond to differences in KRAS protein levels. This work supports the existence of different translational programs such as the up regulation of proliferative tRNAs that have the potential to boost the protein synthesis of oncogenes. Thus, our results suggest that dynamic changes in this fundamental cellular process may contribute to cancer and specifically to the prevalence of mutations in certain genes as compared to their closely related family members.

## 2.3 Results

### 2.3.1 Codon usage of cancer genes

To explore whether the differences in mutation frequency between RAS genes are also observed in other gene families, we perform a genome-wide survey in a pan-cancer data set from The Cancer Genome Atlas (TCGA) to identify gene families with variation in mutation incidence in cancer. To define families, we cluster sets of proteins based on protein sequence similarity. We restrict the analysis to families containing at least one known cancer driver gene (Lawrence et al., 2014). We identify 8 families including the RAS family. We consistently observe one gene more frequently mutated (non-synonymous mutation counts) in comparison to the other genes of the

family (Figure 2-1, Table 2-1). Especially, for the RAS, RAF and RAC families we observe at least a two-fold variation in the mutation count number (fold change between the family member with the lowest mutation count and the highest). For the RHO, FOXA, FGFR and COL families we observe between 1.30 to 1.95-fold change. However, this effect is mild for the AKT family with only a 1.22-fold change.

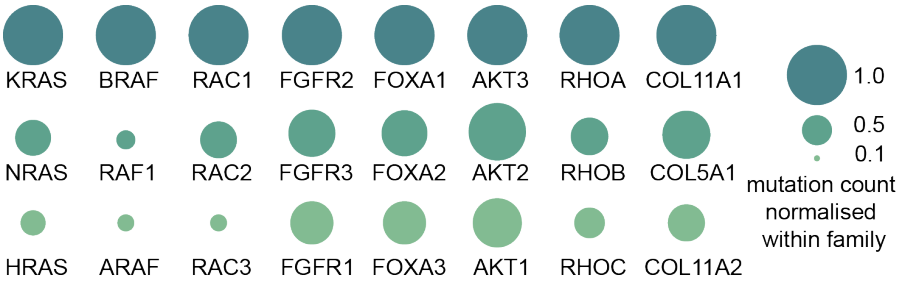



**Figure 2-1 Selected cancer gene families.**

*Gene triplets with divergent mutation frequencies in cancer.*

As previously described, RAS genes have a high amino acid sequence identity (85%) but differ in their codon usage (15% codon identity) (Lampson et al., 2013). The same observation applies to the 7 other families we selected (Figure 2-2). This raises the question of whether differences in the mutation count could be related to variation in codon usage in addition to potential biochemical differences on the protein level. Therefore, we investigate whether the codon usage of these genes could be related to a specific translation program. Previously, it has been described (Gingold et al., 2014) the average codon usage bias in different gene functional groups and observed that genes in two cellular programs, differentiation and proliferation, preferentially use different synonymous codons. Additionally, they found that tRNAs induced during proliferation correspond to the codons that are enriched in the functional set of proliferation genes.

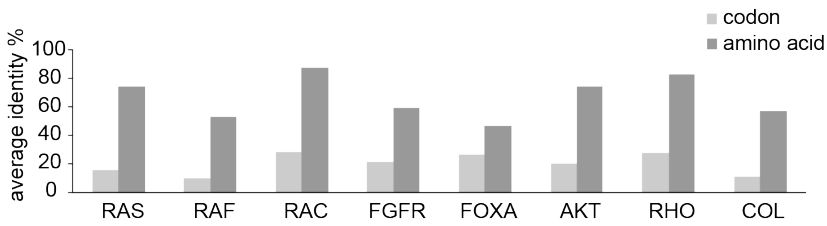



**Figure 2-2 Average codon and amino acid identity of the eight cancer gene families.**

*Amino acid and codon identity are reported as an average obtained from the alignment of the three family members of each family.*

| CCDS | gene | mutation count TCGA | mutation count normalised within family | protein sequence identity to representative % | similarity cutoff % |
|---|---|---|---|---|---|
| | | | | | |
| CCDS877.1 | NRAS | 175 | 0.36 | 85.19 | 74.74 |
| CCDS7698.1 | HRAS | 86 | 0.18 | 100.00 | 74.74 |
| CCDS8702.1 | KRAS | 491 | 1.00 | 84.66 | 74.74 |
| | | | | | |
| CCDS5863.1 | BRAF | 527 | 1.00 | 100.00 | 48.69 |
| CCDS35232.1 | ARAF | 42 | 0.08 | 56.98 | 48.69 |
| CCDS2612.1 | RAF1 | 53 | 0.10 | 56.79 | 48.69 |
| | | | | | |
| CCDS13945.1 | RAC2 | 15 | 0.38 | 92.19 | 89.08 |
| CCDS11798.1 | RAC3 | 9 | 0.23 | 92.71 | 89.08 |
| CCDS5348.1 | RAC1 | 40 | 1.00 | 100.00 | 89.08 |
| | | | | | |
| CCDS2795.1 | RHOA | 46 | 1.00 | 100.00 | 79.25 |
| CCDS1699.1 | RHOB | 18 | 0.39 | 83.16 | 79.25 |
| CCDS854.1 | RHOC | 12 | 0.26 | 91.71 | 79.25 |
| | | | | | |
| CCDS9994.1 | AKT1 | 36 | 1.00 | 100.00 | 76.94 |
| CCDS31076.1 | AKT3 | 44 | 0.27 | 82.88 | 76.94 |
| CCDS12552.1 | AKT2 | 42 | 0.55 | 81.29 | 76.94 |
| | | | | | |
| CCDS31298.1 | FGFR2 | 122 | 1.00 | 100.00 | 60.93 |
| CCDS43730.1 | FGFR1 | 63 | 0.52 | 67.06 | 60.93 |
| CCDS3353.1 | FGFR3 | 75 | 0.61 | 64.36 | 60.93 |
| | | | | | |
| CCDS13147.1 | FOXA2 | 31 | 0.58 | 57.02 | 46.54 |
| CCDS12677.1 | FOXA3 | 28 | 0.53 | 50.86 | 46.54 |
| CCDS9665.1 | FOXA1 | 53 | 1.00 | 100.00 | 46.54 |
| | | | | | |
| CCDS6982.1 | COL5A1 | 203 | 0.64 | 100.00 | 65.88 |
| CCDS43452.1 | COL11A2 | 121 | 0.38 | 71.27 | 65.88 |
| CCDS53348.1 | COL11A1 | 317 | 1.00 | 75.08 | 65.88 |

**Table 2-1 Mutation and protein similarity data from the eight cancer gene families.**

*The first column indicates the Consensus Coding Sequence (CCDS).*

To test if functional adaptation to these cellular programs could have shaped the codon usage in the selected gene families, we examine how the codon usage of the genes correlates to the codon usage of proliferation-related and

differentiation-related genes. We use a similar approach to Gingold et al (Gingold et al., 2014) by applying PCA to the relative codon usage frequencies of all individual genes, in order to visualize how the codon usage of the genes of the 8 families correlate with the codon usage of pro-proliferative genes. By computing the projection of all major gene sets in the Gene Ontology (GO) classification, we reproduce the results of Gingold et al (Gingold et al., 2014) revealing two distinct functional poles at the extremes of the codon usage main projected axis, the first principal component (PC1). At one extreme, for negative values of PC1, we find a strong enrichment of gene sets that are descendants of the "cell cycle" term (16 out of the top 30, Fisher exact test two-sided $p < 2.2e-17$). At the other extreme, for positive values of PC1, a majority of the gene sets are descendants of the "multicellular organism development" or "cell differentiation" terms (14 out of the top 30, Fisher exact test two-sided $p < 5.8e-6$). This observation, together with the previously described changes in tRNAs in proliferative versus non-proliferative cells (Gingold et al., 2014), shows that the two poles of codon usage correspond to two cellular translation programs. We next calculate the average codon usage of each coding sequence of the selected cancer gene families and project it in the PCA plane as well (Figure 2-3). We observe that the transcript of KRAS is composed of codons more frequently used by genes involved in proliferation in comparison to HRAS (Figure 2-3). This seems to be a general phenomenon as the codon usage of the most frequently mutated family member corresponds better to the codon usage of pro-proliferative genes than their cognate family members, except for the FGFR and FOXA families. FGFR1 and FGFR2 have an inverse relationship, with the gene less often mutated being the one having a pro-proliferation codon usage. In cancer, amplifications are the most common alterations of FGFR1 whereas FGFR2 harbors activating mutations (Sobhani et al., 2018). Thus, in our analysis it is difficult to discern which of the two is more oncogenic. FOXA genes show a codon usage in the opposite pole of the pro-proliferation codon usage. In this family the cancer driver gene FOXA1 can take the role of a tumor suppressor (Barbieri et al., 2012; Schroeder et al., 2014), typically inactivated by mutations in contrast to the other families where the most frequently mutated gene is an oncogene with activating mutations. This suggests that the usage of proliferation-associated codons in cancer genes is a characteristic property of oncogenes.

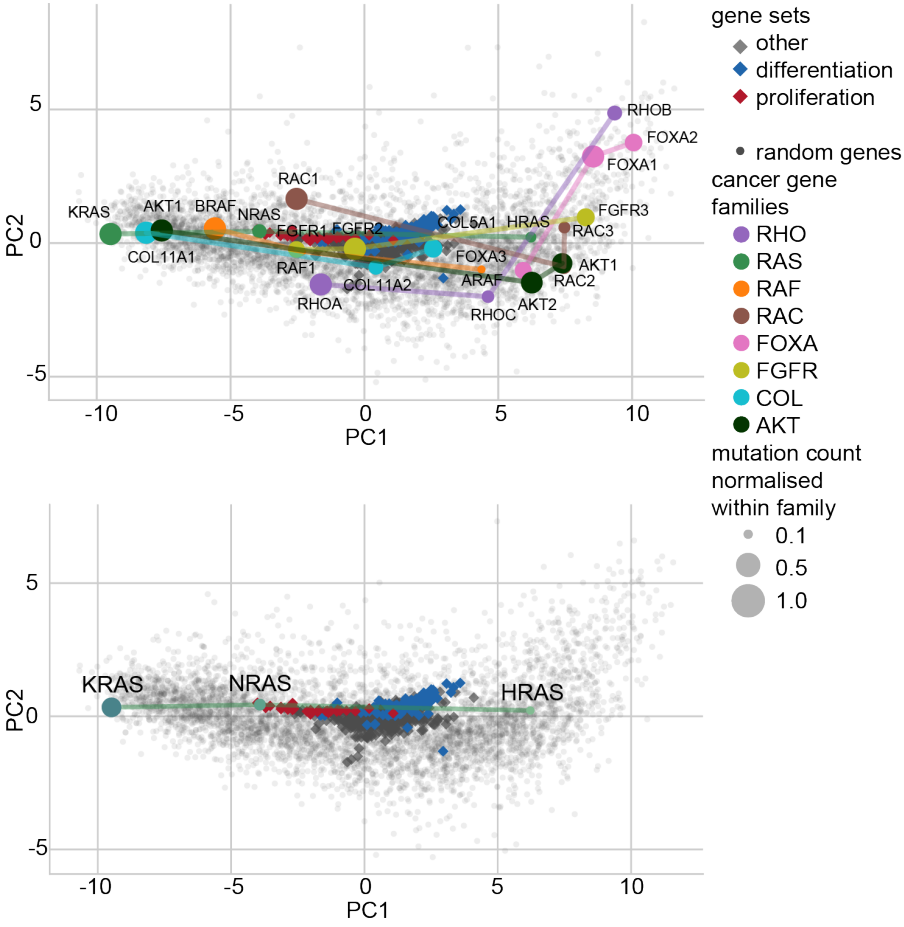



**Figure 2-3 PCA projection of the human codon usage.**

*The location of each gene is determined by its codon usage. Distribution of GO gene sets along the main codon usage axis reveals the two functional poles, "proliferation" (negative PC1) and "differentiation" (positive PC1). The positions of the all studied genes (upper panel) and of only RAS genes (lower panel) and their normalized mutation count are shown.*

Next, we seek to assess the significance of the correlation pattern between codon usage and mutation frequency. Our main observation is that the gene member that is the most frequently mutated is the one that presents a codon usage most adapted to the proliferation codon usage pole (negative pole of PC1). Thus, we expect that PC1 and mutation frequency are negatively correlated. For the 63 gene families that do not contain any cancer driver gene (non-cancer gene families), we assume that there is no specific relationship

between codon usage and mutation frequency, such that the correlation should be randomly distributed around zero. We also assume that the pattern is more significant when, within a cancer gene family, we observe both a large variation in codon usage and in mutation frequency. Thus, we compare the distribution of the covariance of PC1 and mutation frequency for cancer gene families to the background gene families (Figure 2-4). The covariance tends to be large, and thus gives more weight to the families that present a large variation in codon usage and in mutation frequency. Families with little variation in either codon usage or mutation frequency, on the other hand, present a smaller covariance. We observe that the covariance of cancer gene families is significantly more negative than the background families (Wilcoxon-Mann-Whitney (W.M.W.) test $p<0.018$). In particular, 7 out of 8 families (RAS, RAF, COL, RAC, RHO, AKT and FGFR) present a pattern of negative correlation, with the families showing the highest covariance being RAS, RAF, RAC, RHO and COL.

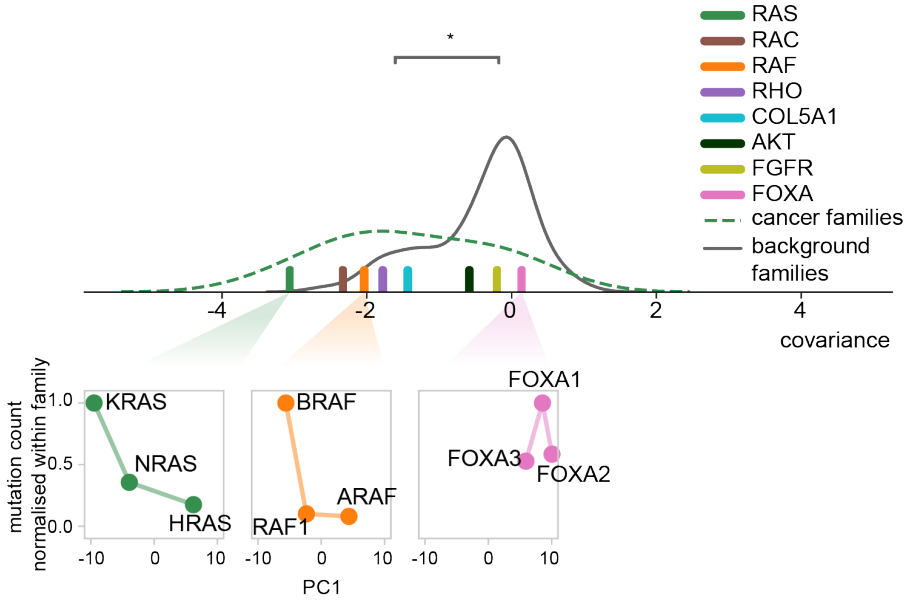



**Figure 2-4 Covariance mutation count and PC1.**

*Distribution of the covariance of mutation count normalized within family and PC1 (lines are kernel density estimates as a guide to the eye). The covariance of cancer gene families is significantly more negative than the background families (W.M.W. test $p<0.018$). All families but one (FOXA) have a negative covariance. See also Figure S1C.*

### 2.3.2 Codon usage-specific changes of KRAS protein abundance under different cell states

The above analysis suggests that oncogenes with a codon usage signature characteristic of proliferation-related genes will be more expressed under a proliferative cell state. To test this hypothesis, we decide to work with the RAS family.

In order to determine if KRAS protein abundance changes in different cell states in a codon-dependent manner, we establish a series of manipulated cells that co-express KRAS wild-type ($KRAS_{WT}$) and a protein sequence-identical KRAS variant but with a codon usage similar to HRAS ($KRAS_{HRAS}$) (Lampson et al., 2013). The two proteins have different N-terminal tags that allow us to distinguish between the two versions of KRAS by their size (FLAG and 3xHA).

A bidirectional symmetrical promoter controls the simultaneous expression of the two genes. This design provides us with a controlled expression system to assess exclusively codon-dependent changes in protein abundance while reducing the impact of other factors (e.g. transcriptional efficiency or biochemical properties of the protein). Moreover, both genes are in the same plasmid and are therefore integrated into the genome with equal stoichiometry (Figure 2-5).

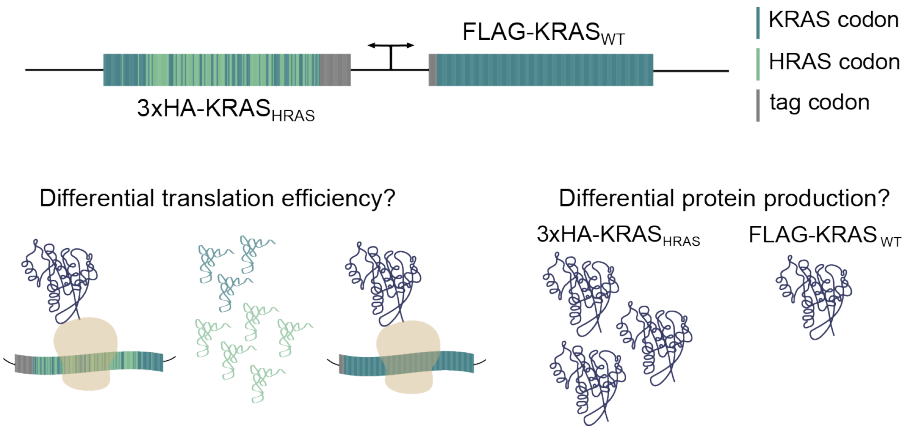



**Figure 2-5 Experimental design.**

*The construct co-expresses two genes coding for the same KRAS protein, differentiated by size with two different tags. HRAS-specific codons are represented in green. KRAS-specific codons and identical codons between KRAS and HRAS are represented in blue. KRAS protein is represented in dark purple. Distinct tRNAs pools appear in green and blue.*

Gingold et al (Gingold et al., 2014) reported changes in tRNA profiles of BJ/hTERT fibroblast in different cell-states: a quiescent state when the cells are starved and a proliferative state when the cells are not starved. Therefore, we first co-express $KRAS_{WT}$ and $KRAS_{HRAS}$ in BJ/hTERT fibroblasts and quantify the protein ratio of $KRAS_{WT}/KRAS_{HRAS}$ in these two different cell-states. We observe that the ratio increases by more than two-fold when the cells are proliferating (Figure 2-6). The observed fold change suggests that KRAS codons are more efficiently translated during proliferation than HRAS codons.

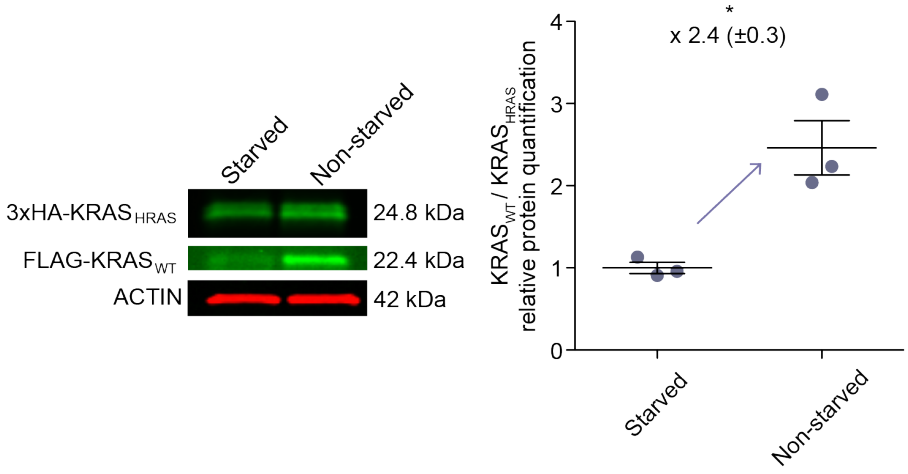



**Figure 2-6 $KRAS_{WT}$ and $KRAS_{HRAS}$ protein expression during starvation and non-starvation.**

*Western blot analysis of the levels of $KRAS_{WT}$ and $KRAS_{HRAS}$ in starved and non-starved BJ/hTERT cells. The protein ratio $KRAS_{WT}/KRAS_{HRAS}$ increases from quiescent to proliferative state. Results are representative of three independent experiments with three technical replicates each. Values are relative to starved condition. Error bars represent SEM. * $p < 0.05$ (unpaired Student t test).*

We also measure the ratio at the transcript level and, interestingly, we find the same effect as observed at the protein level: the ratio between $KRAS_{WT}$ and $KRAS_{HRAS}$ is increased by more than two-fold in proliferation versus starvation (Figure 2-6). Previous studies have shown in different species that codon optimality has a high impact on transcript stability (Boël et al., 2016; Presnyak et al., 2015; Wu et al., 2019). An interesting hypothesis is that the dynamics of ribosomal elongation influences mRNA decay. Ribosome translocation is slower through non-optimal transcripts and promotes mRNA decay, mediated by the DEAD-box protein Dhh1p in *S. cerevisiae*

(Radhakrishnan et al., 2016). Thus, codon content directly modulates both translation efficiency and mRNA stability. Our study suggests that $KRAS_{WT}$ is composed of codons that are optimal for its expression in proliferative cells but that are non-optimal in starved cells. Therefore, to determine if changes in KRAS transcript abundance are due to differences in translation efficiency and not transcriptional regulation, we prevent translation by deleting the ribosome binding site and the ATG start codon. We first test to confirm that there is no protein expression when cells are established with the non-productive expression cassette (Figure 2-7).

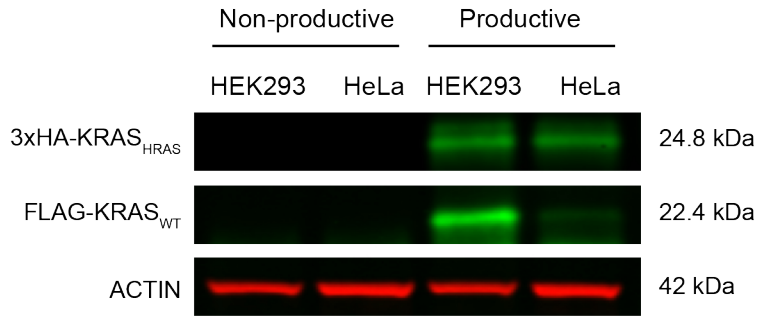



**Figure 2-7 Comparison of expression between the productive and non-productive (without RBS and ATG) expression cassette.**

*No $KRAS_{WT}$ or $KRAS_{HRAS}$ expression is observed when translation is suppressed.*

After blocking the translation of the two genes we observe that the difference of $KRAS_{WT}/KRAS_{HRAS}$ between non-starved and starved is not significant at the transcript level (Figure 2-8). In short, $KRAS_{WT}/KRAS_{HRAS}$ changes are mainly due to a differential translation efficiency (that also increases the corresponding mRNA level) between a quiescent state and a proliferative state. Our results provide new evidence supporting the dynamic translational efficiency by cell-state-specific codon usage of transcripts.

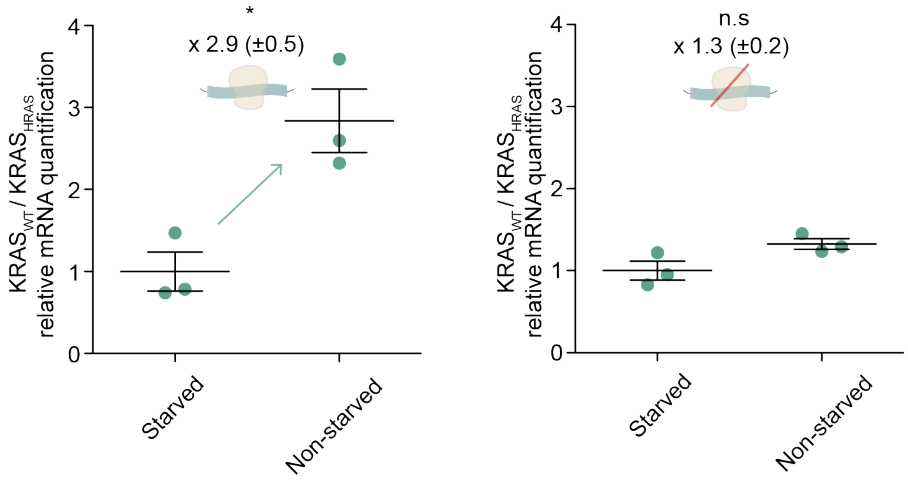



**Figure 2-8 Transcript abundance with and without translation.**

*Left panel: The transcript ratio $KRAS_{WT}/KRAS_{HRAS}$ increases between the two cell states. Right panel: Translation inhibition with RBS and ATG site removal decreases the effect on transcript level. Results are representative of three independent experiments with three technical replicates each. Values are relative to starved condition. Error bars represent SEM. * $p < 0.05$ (unpaired Student t test).*

### 2.3.3 Specific differences in tRNA levels explain differences in KRAS abundance between cell lines

To investigate if the condition-specific translation efficiency is mediated by differential tRNA expression, we explore the effect of cell line-specific tRNA abundances on KRAS expression. A previous study (Fu et al., 2018) has already reported a cell line-specific expression of $KRAS_{WT}$ and $KRAS_{HRAS}$. We therefore hypothesize that the tRNA content of different cell lines varies and may influence the translation efficiency in a codon-dependent manner. To test our hypothesis, we first establish two additional cell lines (HEK293 and HeLa) to co-express $KRAS_{WT}$ and $KRAS_{HRAS}$. We verify that the expression results in changes of both protein and mRNA when comparing BJ/hTERT, HEK293 and HeLa (Figure 2-9).

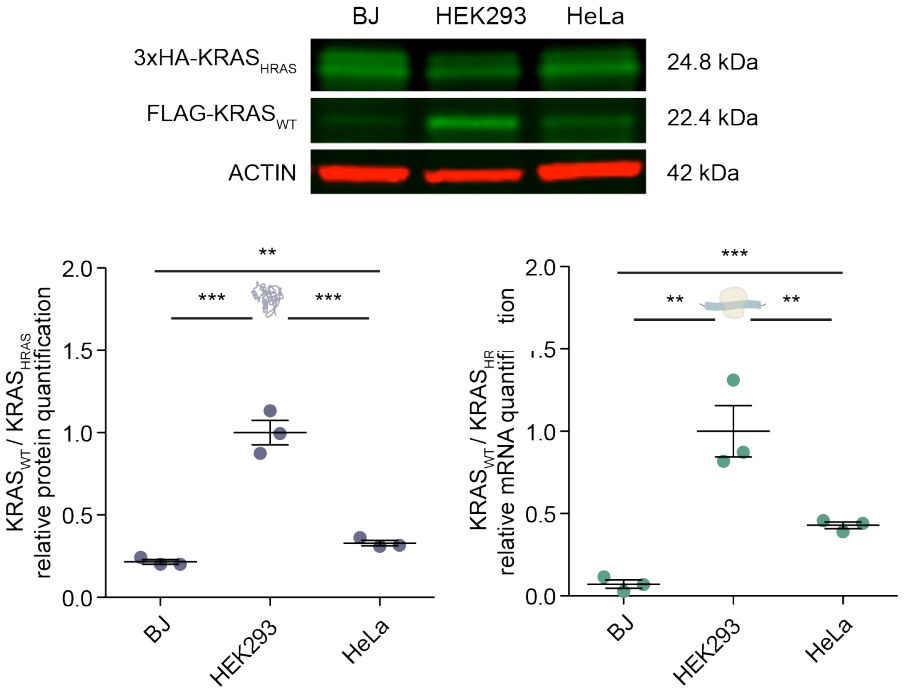



**Figure 2-9 KRAS$_{WT}$ and KRAS$_{HRAS}$ protein expression in BJ/hTERT, HEK293 and HeLa.**

*Western blot and mRNA quantification of the levels of KRAS$_{WT}$ and KRAS$_{HRAS}$ in BJ/hTERT, HEK293 and HeLa cells. The protein and the transcript ratio KRAS$_{WT}$/KRAS$_{HRAS}$ vary between the different cell lines. Results in are representative of three independent experiments with three technical replicates each. Values are relative to HEK293. Error bars represent SEM. ** $p < 0.01$; *** $p < 0.001$ (unpaired t test).*

Of these three cell lines HEK293 exhibits the highest proliferative rate (Figure 2-10) and higher abundance of KRAS$_{WT}$ in comparison to HeLa and BJ/hTERT. We observe the same effect on protein level when switching the position of the tags (Figure 2-11), showing that FLAG and 3xHA are not influencing our observation.

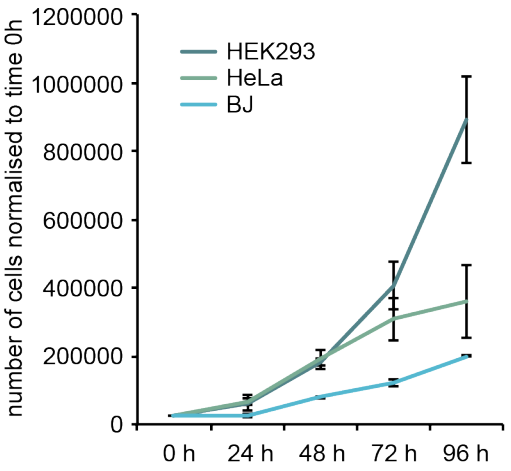



**Figure 2-10 HEK293, HeLa and BJ/hTERT cell count over 96 hours.**

*Results are representative of two independent experiments with three technical replicates each. Values are relative to starved condition. Error bars represent SEM.*

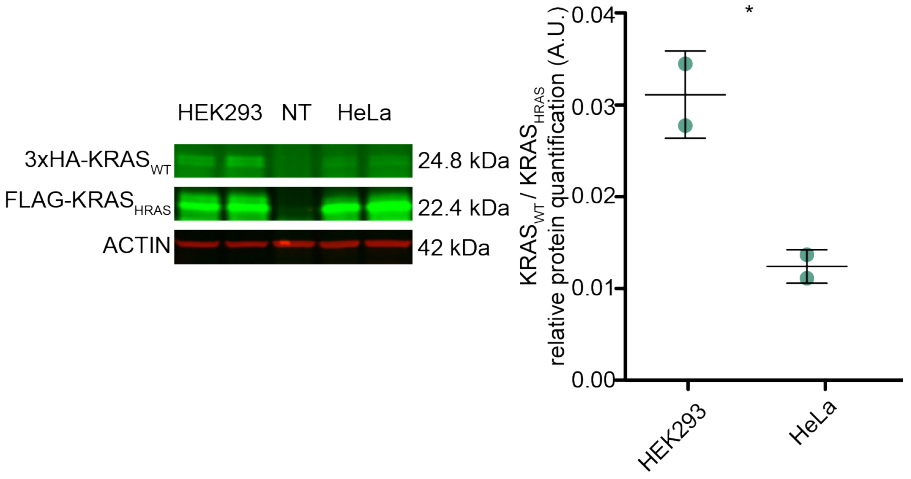



**Figure 2-11 Control with inverted tags in the expression construct.**

*Western blot analysis of the levels of KRAS$_{WT}$ and KRAS$_{HRAS}$ with inverted tags in HEK293 and HeLa cells. The protein ratio KRAS$_{WT}$/KRAS$_{HRAS}$ varies between the different cell lines similarly, regardless of the tag. Error bars in represent SEM of two independent experiments *, $p < 0.05$.*

As before, the removal of the ribosome binding site and start codon leads to similar transcript levels for the three cell lines (Figure 2-12), indicating that translation is an important determinant of mRNA stability.

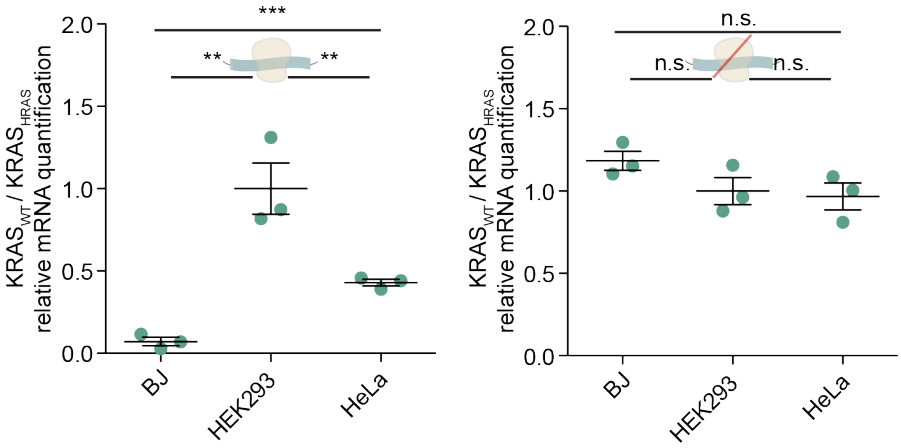



**Figure 2-12 Comparison of transcript abundance with and without translation.**

*Right panel: Translation inhibition decreases the differential effect observed in cell lines on transcript level. Results in are representative of three independent experiments with three technical replicates each. Values are relative to HEK293. Error bars represent SEM. ** $p < 0.01$; *** $p < 0.001$ (unpaired t test).*

The above observations suggest that the effect of codon bias may be differentially regulated in different cell types. If translational efficiency is different in each cell type, we hypothesize that it should match the cell type's tRNA anticodon abundance. More specifically, we expect that the relative synonymous codon frequencies (relative to the amino acid) of $KRAS_{WT}$ match better to the relative abundances of cognate tRNAs in HEK293 than in HeLa. To associate the amount of tRNAs with codon usage, we perform hydro-tRNA sequencing (Gogakos et al., 2017) and quantify tRNA expression in HEK293 and HeLa cells (Table 2-2). We find 14 tRNAs showing significant differences ($q < 0.05$, t test) between the two cell lines (Figure 2-13).

| tRNA | HEK293_rep1 | HEK293_rep2 | HEK293_rep3 | HeLa_rep1 | HeLa_rep2 | HeLa_rep3 |
|---|---|---|---|---|---|---|
| AlaAGC | 9962.2 | 11428.7 | 11205.6 | 12746.7 | 12454.9 | 14826.1 |
| AlaCGC | 3490.2 | 3754.4 | 3554.9 | 357.0 | 330.8 | 328.7 |
| AlaGGC | 21.0 | 13.5 | 18.7 | 9.5 | 16.1 | 4.2 |
| AlaTGC | 7379.8 | 7674.9 | 7562.5 | 4377.4 | 4385.2 | 3944.7 |
| ArgGCG | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| ArgACG | 10343.5 | 9117.6 | 10840.4 | 6969.7 | 6052.9 | 7539.7 |
| ArgCCG | 2700.8 | 2426.2 | 2939.1 | 1920.0 | 1899.7 | 2001.7 |
| ArgCCT | 5402.7 | 8586.5 | 7207.6 | 3724.1 | 6638.6 | 3902.8 |
| ArgTCG | 2034.4 | 2001.9 | 2386.4 | 1836.4 | 1768.8 | 2313.6 |
| ArgTCT | 6306.5 | 5594.4 | 6414.1 | 3619.7 | 3640.9 | 3764.6 |
| AsnATT | 5.1 | 10.4 | 17.4 | 1.9 | 2.3 | 2.1 |
| AsnGTT | 13090.8 | 16099.5 | 21450.9 | 9193.5 | 9657.0 | 9160.3 |
| AspATC | 24.4 | 19.3 | 27.7 | 7.6 | 9.2 | 12.6 |
| AspGTC | 26273.9 | 24881.7 | 28560.6 | 19589.2 | 20883.0 | 20313.9 |
| CysACA | 474.3 | 306.7 | 330.4 | 87.4 | 80.4 | 108.9 |
| CysGCA | 10490.3 | 15346.1 | 11942.5 | 2522.0 | 2758.8 | 1809.0 |
| GlnCTG | 21953.0 | 21092.5 | 22154.3 | 9660.7 | 9392.9 | 9964.3 |
| GlnTTG | 17742.6 | 17024.1 | 16039.7 | 7953.4 | 7348.4 | 6924.2 |
| GluCTC | 212242.8 | 231932.2 | 195640.4 | 335733.8 | 337233.0 | 325915.0 |
| GluTTC | 197175.5 | 173224.0 | 175162.1 | 160249.6 | 163551.4 | 157672.8 |
| GlyCCC | 11768.2 | 12061.8 | 13963.1 | 11871.3 | 12069.0 | 12252.9 |
| GlyGCC | 30083.1 | 30585.4 | 29345.8 | 26114.5 | 26129.5 | 31499.0 |
| GlyTCC | 9894.2 | 9672.1 | 11095.5 | 9163.1 | 8965.6 | 10393.6 |
| GlyACC | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| HisGTG | 48612.9 | 47180.5 | 57134.8 | 41290.2 | 41630.4 | 37828.5 |
| HisATG | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| IleAAT | 2937.6 | 2980.7 | 4189.3 | 2288.4 | 1913.5 | 2374.4 |
| IleGAT | 1275.6 | 1060.6 | 1261.8 | 1770.0 | 1435.7 | 1622.7 |
| IleTAT | 1210.4 | 1138.7 | 1310.8 | 697.0 | 1123.3 | 695.1 |
| LeuAAG | 673.8 | 857.0 | 1094.3 | 670.4 | 673.1 | 577.9 |
| LeuCAA | 7878.5 | 8663.6 | 9257.8 | 6396.2 | 6193.0 | 9688.0 |
| LeuCAG | 1611.6 | 3351.4 | 3949.1 | 415.9 | 493.9 | 1329.6 |
| LeuTAA | 6173.9 | 6574.2 | 6291.1 | 7793.9 | 8333.9 | 7200.5 |
| LeuTAG | 11826.0 | 10746.7 | 11054.9 | 8933.4 | 9721.3 | 9388.6 |
| LeuGAG | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| LysCTT | 67913.3 | 69311.7 | 71313.0 | 68304.8 | 68410.0 | 72396.9 |
| LysTTT | 21823.3 | 24140.3 | 26064.7 | 15675.1 | 14379.9 | 15921.2 |
| MetCAT | 16837.1 | 15467.4 | 19496.7 | 19887.3 | 18009.3 | 18333.2 |
| PheGAA | 8657.1 | 9254.5 | 9499.4 | 19834.2 | 17816.3 | 17927.0 |
| PheAAA | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| ProAGG | 448.8 | 676.3 | 575.2 | 193.7 | 174.6 | 255.4 |
| ProCGG | 588.2 | 679.4 | 553.3 | 355.1 | 340.0 | 364.3 |
| ProGGG | 8.5 | 7.3 | 7.1 | 7.6 | 6.9 | 6.3 |
| ProTGG | 26353.8 | 19392.6 | 22165.2 | 19384.1 | 17694.6 | 20173.7 |
| SerACT | 1292.0 | 2528.8 | 2368.4 | 279.2 | 356.1 | 519.3 |
| SerAGA | 1226.9 | 2755.8 | 2622.2 | 136.7 | 108.0 | 226.1 |
| SerCGA | 723.6 | 1134.5 | 1448.6 | 222.2 | 204.4 | 355.9 |
| SerGCT | 1871.2 | 3093.7 | 3795.1 | 723.6 | 882.1 | 1704.3 |
| SerGGA | 0.6 | 7.3 | 6.4 | 1.9 | 0.0 | 0.0 |
| SerTGA | 2168.1 | 2426.2 | 2903.0 | 2187.8 | 1860.7 | 2150.3 |
| ThrAGT | 3406.3 | 4078.8 | 4885.6 | 2296.0 | 1993.9 | 2759.6 |
| ThrCGT | 1531.2 | 1559.3 | 1872.4 | 1644.6 | 1589.6 | 1708.5 |
| ThrTGT | 9160.3 | 8187.7 | 10433.3 | 5372.5 | 5347.7 | 5841.7 |
| ThrGGT | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| TrpCCA | 4354.3 | 3643.5 | 3967.7 | 4375.5 | 4153.2 | 4614.7 |
| TyrATA | 2.3 | 6.2 | 3.2 | 3.8 | 4.6 | 4.2 |
| TyrGTA | 11824.8 | 14634.9 | 15322.8 | 13274.7 | 13270.4 | 14001.2 |
| ValAAC | 5603.9 | 5780.2 | 5601.2 | 7590.7 | 7943.4 | 8666.2 |
| ValCAC | 38588.9 | 45481.7 | 36630.1 | 30831.8 | 30710.0 | 30638.4 |
| ValTAC | 73241.7 | 61083.4 | 62310.2 | 75743.6 | 73957.5 | 72281.8 |
| ValGAC | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| SupTCA | 384.8 | 350.9 | 274.4 | 431.1 | 535.2 | 412.5 |
| SupTTA | 74.8 | 55.7 | 47.7 | 66.5 | 52.8 | 52.3 |
| iMetCAT | 6032.8 | 5115.4 | 6215.0 | 6120.8 | 6197.6 | 5695.1 |
| SeCTCA | 14344.8 | 13296.8 | 15840.7 | 6711.4 | 6801.7 | 7227.8 |
| UndetNNN | 477.1 | 442.6 | 373.6 | 379.8 | 413.5 | 402.0 |
| READS | 1764677.0 | 1920686.0 | 1552522.0 | 526566.0 | 435331.0 | 477603.0 |

**Table 2-2 Hydro-tRNA sequencing expression data (rpm).**

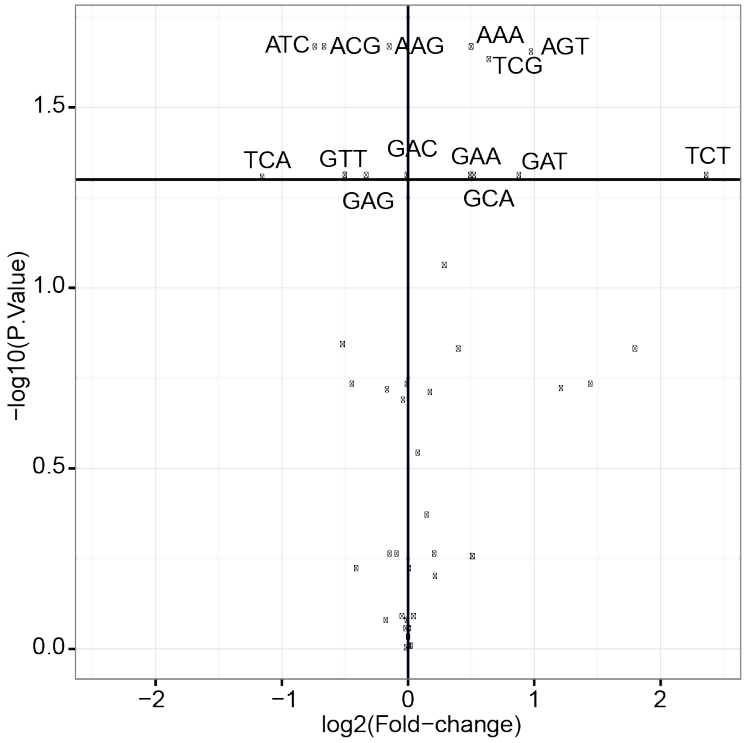



**Figure 2-13 Volcano plot showing relative tRNA differential expression in log2 fold change between HEK293 and HeLa.**

Six of them are expressed higher in HEK293 and match codons enriched in the coding sequence of $KRAS_{WT}$ (TCT, AAA, AGT, GAT, GAA, GCA). One exception occurs with $tRNASer^{CGA}$, which is expressed higher in HEK293, but the associated codon TCG is not enriched in $KRAS_{WT}$. On the other hand, 5 tRNAs expressed significantly higher in HeLa correspond to codons enriched in HRAS and therefore in $KRAS_{HRAS}$ (ATC, GAG, GAC, ACG and AAG). Only two tRNAs higher in HeLa ($tRNASer^{TGA}$ and $tRNAVal^{AAC}$) do not match codons enriched in $KRAS_{HRAS}$ (Figure 2-14).

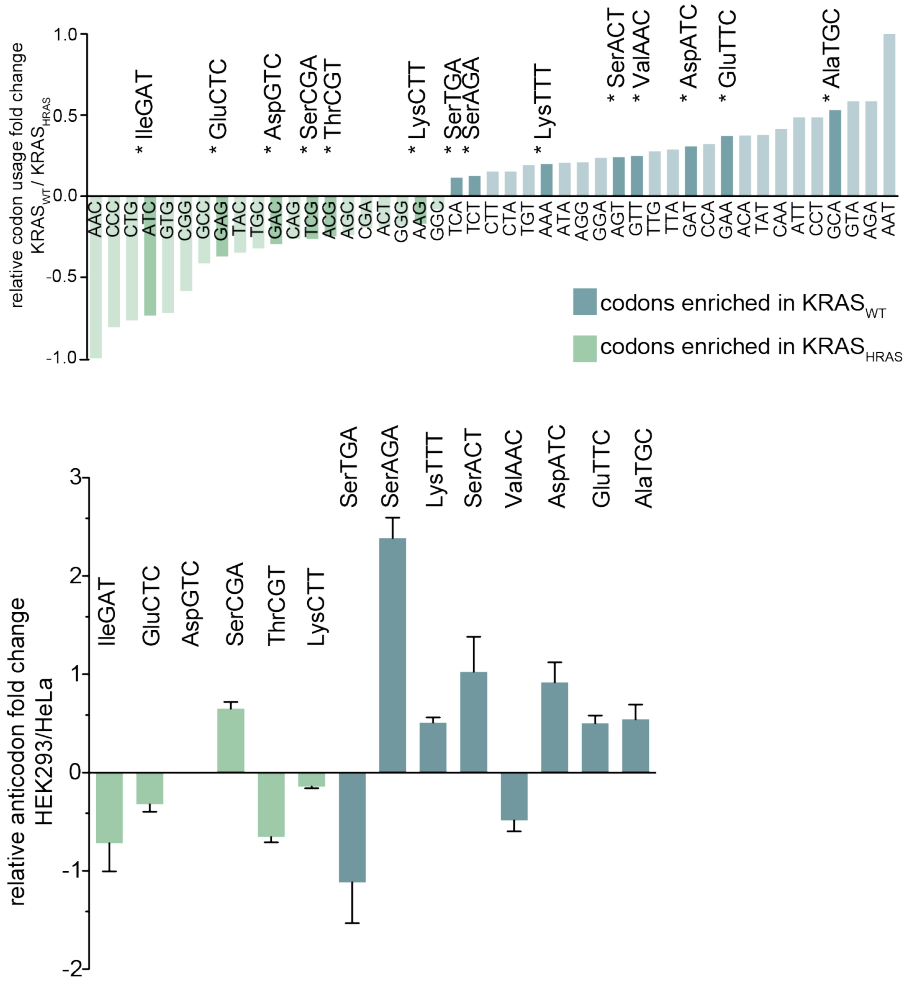



**Figure 2-14 tRNA levels associated to KRAS$_{WT}$ and KRAS$_{HRAS}$ codon usage.**

*Upper panel: Fold change of the relative codon usage (pseudocount +1) between KRAS$_{WT}$/KRAS$_{HRAS}$. The codons that are not changing in amount between KRAS$_{WT}$ and KRAS$_{HRAS}$ are not represented. tRNAs differentially expressed between HEK293 and HeLa are highlighted. Lower panel: Fold change of tRNA expression between HEK293 and HeLa is represented for the cognate tRNAs of the codons enriched in either KRAS$_{WT}$ (blue) or KRAS$_{HRAS}$ (green). Error bars represent SEM of three independent hydro-tRNAseq experiments. Differences between HEK293 and HeLa were assessed by a multiple t test using a permutation based FDR cut-off of $p<0.05$. * $p < 0.05$.*

To sum up, we find 11 out of 14 tRNAs matching the expected codons in KRAS$_{WT}$ and KRAS$_{HRAS}$. Therefore, the difference in tRNA supply between cell lines could explain the observed variation of KRAS$_{WT}$ and KRAS$_{HRAS}$

protein levels. In a previous study in which different codons of KRAS had been changed, it was observed that certain replacements resulted in significant increases in KRAS expression reaching in a cumulative manner the levels of HRAS (Lampson et al., 2013). Among them were the changes GCA to GCC, AAA to AAG and ATT to ATC, which correspond to the anticodons of tRNAs differentially expressed between HEK293 and HeLa. The changes GAA to GAG and CCT to CCC did not display protein abundance changes (Figure 2-15).

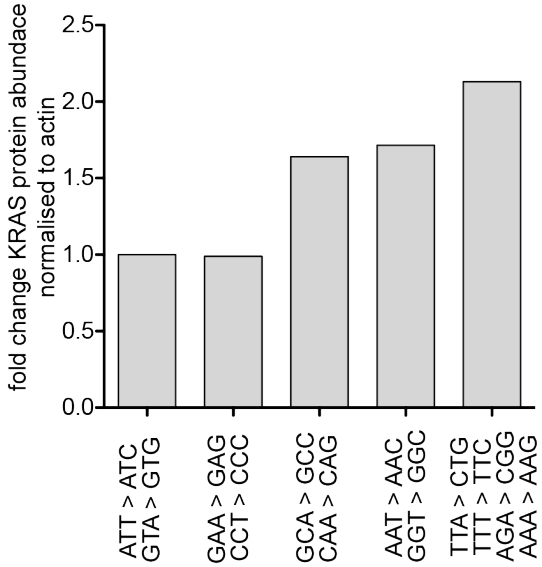



**Figure 2-15 Immunoblot data (from Lampson et al., 2013).**

*Immunoblot data taken and quantified from Lampson et al (Lampson et al., 2013) where KRAS codons were progressively converted.*

Additionally, we investigate if the codons corresponding to the significantly changing tRNAs are also found enriched in the most prevalent oncogenes of the RAF, RAC, RHO and COL families. Overall, the codons enriched in $KRAS_{WT}$ and having their matching tRNAs significantly increased in HEK293 are also enriched in the oncogenes BRAF, RAC1, RHOA and COL11A1 in comparison to their less mutated family member. Conversely, the codons enriched in $KRAS_{HRAS}$ and the matching tRNAs in HeLa, are also enriched in the less frequently mutated genes, RAF1, RAC3, RHOC and COL11A2 (Figure 2-16).

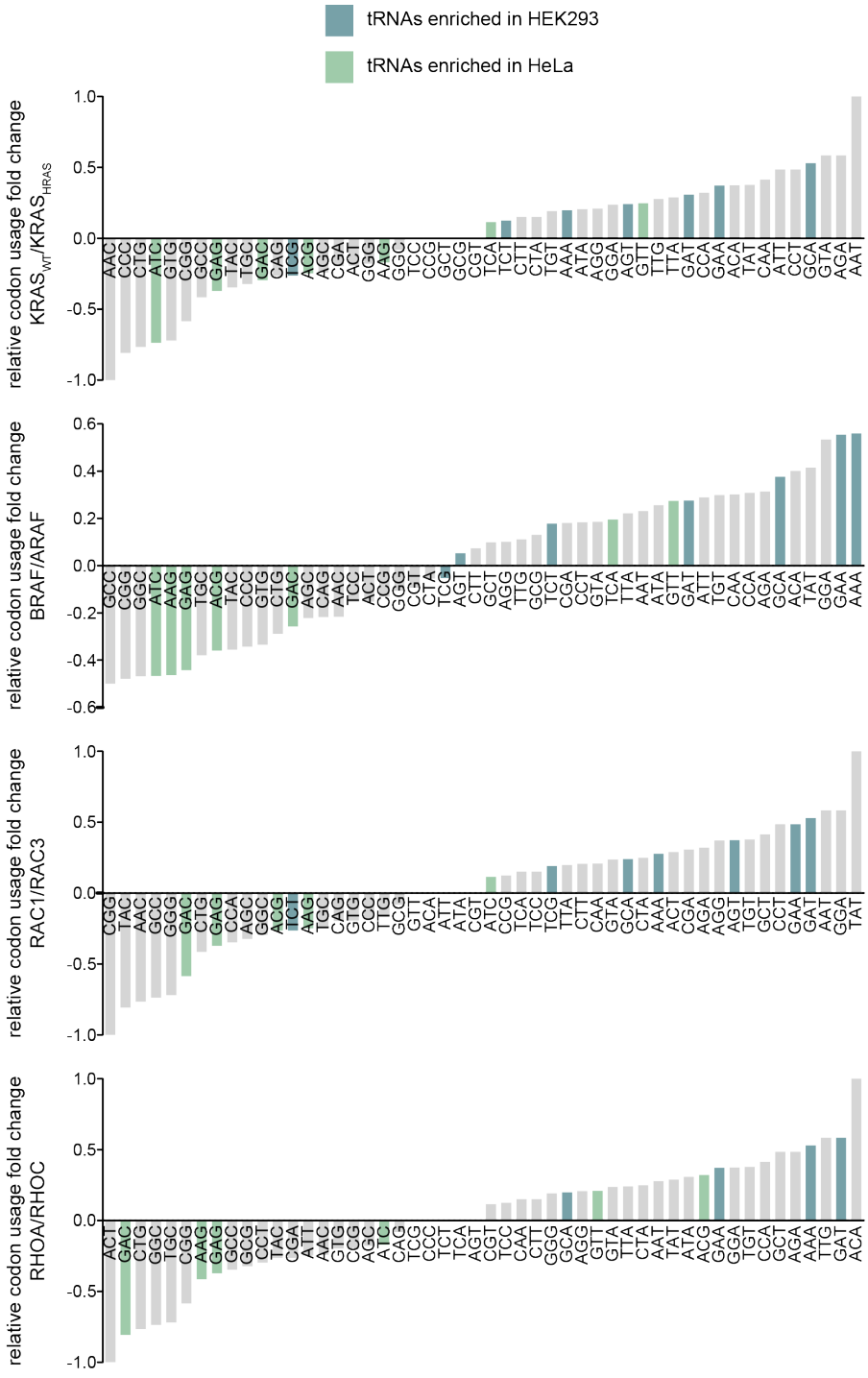



Figure continues on next page


**Figure 2-16 Relative codon usage between the most frequently and the less frequently mutated gene.**

*Fold change of the relative codon abundance (pseudocount +1) between the most frequently and the less frequently mutated gene for the 4 families displaying the highest negative covariance together with the RAS family. tRNAs differentially expressed between HEK293 (blue) and HeLa (green). Generally, the tRNAs enriched in HEK293 and matching the codons enriched in KRAS$_{WT}$ are enriched in BRAF, RAC1, RHOA and COL11A1 and vice versa with the tRNAs enriched in HeLa.*

Finally, we confirm that 6 out of 7 tRNAs more highly expressed in the most proliferative cell line HEK293, correspond to proliferation-related codons (Figure 2-17). Altogether, our results support a dynamic translational program, where specific changes in tRNA abundance can shape the expression of proliferation-related transcripts.

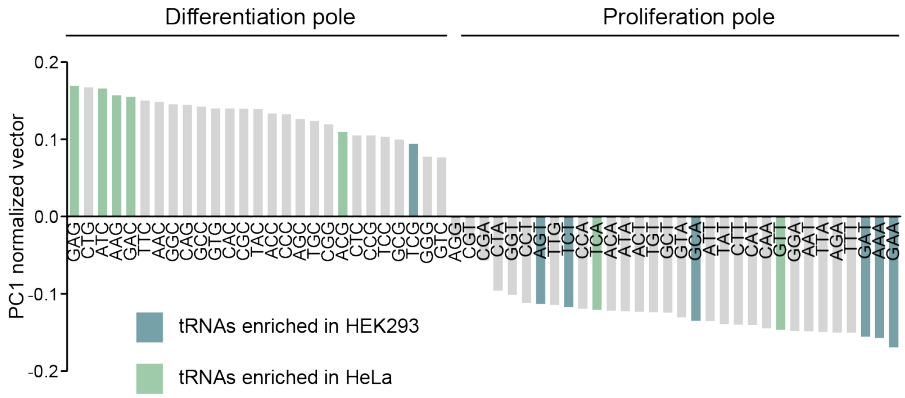



**Figure 2-17 Proliferation- versus differentiation-related codons.**

*Codons are ordered following their value in the first component axis (PC1 axis). Note that the scale of the values is arbitrary, only the relative values are important (direction of the vector in the multidimensional space). Negative values indicate negative PC1 toward the proliferation pole, positive values toward the differentiation pole.*

## 2.4 Discussion

Codon usage and tRNA abundance are crucial for efficient and accurate translation of mRNAs into proteins. Previous studies have found that tRNA repertoires are dynamic in a manner that facilitates selective translation of specific transcripts (Bornelöv et al., 2019; Gingold et al., 2014; Newman et al., 2016; Supek, 2016; Torrent et al., 2018). Here, we investigate if there is an oncogenic translation program shaping the abundance of cancer driver genes. We describe protein families with strong differences in codon usage and mutation frequencies within the family. The observed codon bias reveals a proliferation-specific codon usage of the more prevalent family members in cancer. Specifically, the families RAS, RAF, RAC, RHO and COL exhibit the largest negative covariance between mutation frequency and proliferation-associated codon usage. This raises the question of whether these transcripts are more effectively translated in proliferative cells than their closely related family members.

We focus on the example of the RAS family and we experimentally show that the translation efficiency of $KRAS_{WT}$ is upregulated in proliferative cells in comparison to the translation efficiency of $KRAS_{HRAS}$. Additionally, we find that translation efficiency is a determinant of transcript abundance. This observation has been previously described in *H. sapiens* (Wu et al., 2019), *S.*

*cerevisiae* (Presnyak et al., 2015) and *E. coli* (Boël et al., 2016). Here, we consistently show that the changes in $KRAS_{WT}$/$KRAS_{HRAS}$ transcript abundance between cell types and cell states decrease when translation is suppressed.

Activating mutations of oncogenes are a product of selection during tumor initiation for an ideal level of signaling. It is plausible that selection acting on a gene depends on the level of expression of that gene. Pershing et al observed that replacing KRAS codons with HRAS codons in one exon renders mice more resistant to lung tumors and decreases the amount of KRAS mutations (Pershing et al., 2015). This supports our hypothesis that translation efficiency might contribute to mutation frequency differences between genes.

RAS abundance is an important determinant of MAPK signaling which is tightly connected to cancer growth. Importantly, it has been shown that cells initially expressing low levels of oncogenic RAS only progressed into malignant lesions after RAS levels increased (Ferbeyre, 2007; Sarkisian et al., 2007). In line with this model, it is tempting to speculate that mutated RAS increases its own translation by triggering cell proliferation (Figure 2-18).

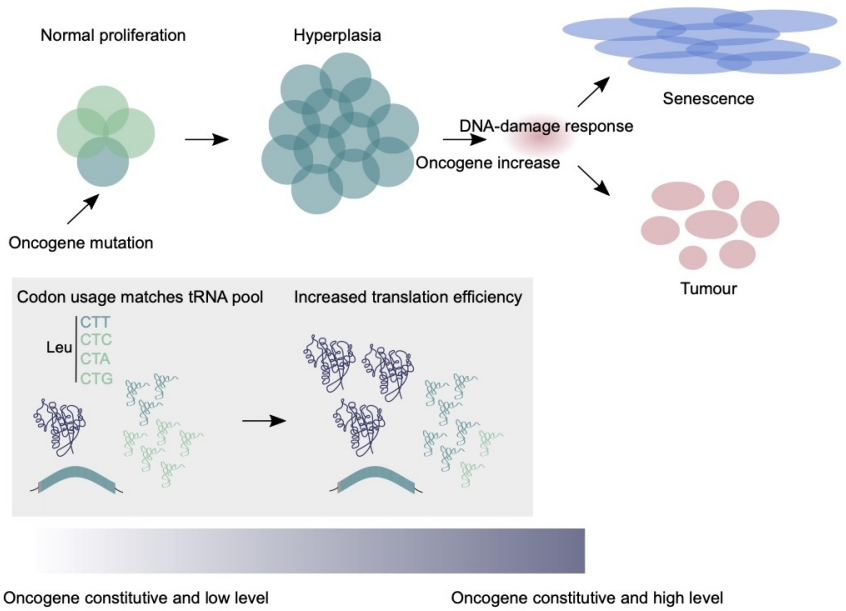



**Figure 2-18 Based on the three-stage carcinogenesis model for RAS-induced tumors proposed (Sarkisian et al., 2007).**

*Adapted figure from "Barriers to RAS transformation", Ferbeyre G, Nat Cell Biol, 2007 (Ferbeyre, 2007). Oncogenic mutations lead to a constitutive activation of the oncogene that promotes cell proliferation. Low levels of oncogene are not transforming. An increase of oncogene translation efficiency will increase oncogene levels. High oncogene levels and activity would transform cells if cells evade senescence.*

Our observations agree with recently identified alterations in transcript-specific translation that emerge as drivers of cellular transformation. For example, it has been shown that up-regulation of specific tRNAs (tRNAGluTTC and tRNAArgCCG) in metastatic cells leads to an increase in the amount of certain proteins, specifically EXOSC2 and GRIPAP1 that play an important role in metastasis (Goodarzi et al., 2016). Indeed, we have seen similar results here for KRAS. Consistent with previous reports (Dittmar et al., 2006; Pavon-Eternod et al., 2009), we observe that specific tRNAs vary between different cell lines, which could explain the differences in KRAS expression between HEK293 and HeLa. One of them, tRNAGluTTC, is also upregulated in metastatic breast cancer cells as mentioned above. Moreover, we find differentially expressed tRNAs corresponding to codons previously reported to change KRAS protein levels when synonymously mutated

(Lampson et al., 2013). Taken together, our results suggest that in order to increase KRAS translation efficiency it is not necessary to change the expression of multiple tRNAs but just that of a few specific ones. Particularly, we observe that codons corresponding to the tRNAs playing a role in these changes, are also codons enriched in the oncogenes BRAF, RAC1, RHOA and COL11A1. Our results suggest that certain tRNAs could be used as markers of oncogene-specific translation. Determination of tRNA abundance of different cell types may reveal previously unseen connections between translation and oncogene prevalence in cancer. It would also be interesting to investigate how tRNA modifications could also influence oncogene translation. Furthermore, Supek et al (Supek et al., 2014) show that selection acts on somatic synonymous mutations of oncogenes in tumor evolution. In many cases they are associated with changes in oncogene splicing in tumors. It would be interesting to further investigate if some of the recurrent synonymous mutations in those oncogenes correspond to changes towards enriching their coding sequence in proliferation-related codons, ultimately yielding to a greater translation efficiency.

The question remains as to the physiological role of this family codon bias. One possible explanation is that the protein levels of KRAS, BRAF, RAC1, RHOA and COL11A1 need to be tightly controlled. For example, during embryogenesis tRNA levels have been shown to vary in mouse brain and liver (Schmitt et al., 2014). Indeed, KRAS (Johnson et al., 1997), RAC1 (Corbetta et al., 2005; Roberts et al., 1999; Sugihara et al., 1998) and RHOA (Hakem et al., 2005; Liu et al., 2001) are the only family members embryonically lethal in homozygous null mice. On the other hand, BRAF (Wojnowski et al., 1997) and COL11A1 (Li et al., 1995) homozygous mutants are also lethal together with RAF1 (Wojnowski et al., 1998) and COL5A1 (Wenstrup et al., 2004). Thus, cancer cells could take advantage of a developmental translation regulation to boost the translation of oncogenic transcripts to their own growth advantage.

Taken together, our work not only addresses a fundamental aspect of RAS biology but also provides insight into the controversial issue of how codon bias can influence protein expression. Collectively, our findings demonstrate that codon-driven translational efficiency can modulate protein expression of oncogenes in different cell contexts.

## 2.5 Appendix

In addition, I provide an Appendix that contains further results which were not included in the initial manuscript submission.

In this project we also work with the colorectal cancer cell line, HCT116. These cells harbor an oncogenic mutation in the gene KRAS (mutation G13D). We perform the same expression experiments as presented in Section 2.3.3 (Figure 2-19). Both the protein and the transcript ratios of $KRAS_{WT}/KRAS_{HRAS}$ are lower in HCT116 cells similarly to HeLa cells (Figure 2-19A, B). However, opposite to the other cell lines, when the translation of the KRAS transcripts is inhibited we still observe that the ratio $KRAS_{WT}/KRAS_{HRAS}$ is lower in HCT116 cells (Figure 2-19C). We hypothesize, that HCT116 cells due to their mutational state where KRAS is mutated, might develop a safety mechanism in order to downregulate KRAS transcripts such as the expression of microRNAs targeting KRAS. However, we have failed to find a microRNA known to be expressed in HCT116 cells and targeting KRAS. In order to explore this possibility, we could imagine to change by stretches the sequence of $KRAS_{WT}$ (with for example the codons of HRAS) and analyze the expression. This type of experiment could give us a hint whether there is a microRNA targeting KRAS and which is the sequence it is recognizing.

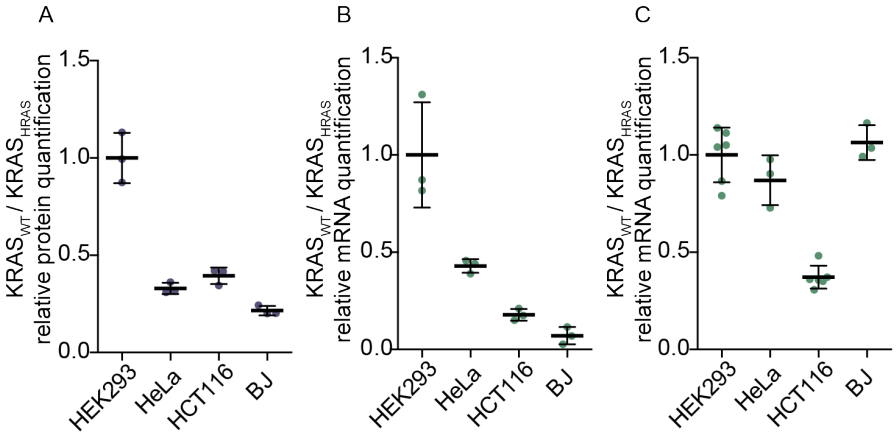



**Figure 2-19 Comparison with HCT116.**

*Western blot and mRNA quantification of the levels of KRAS$_{WT}$ and KRAS$_{HRAS}$ in HEK293, HeLa, HCT116, and BJ/hTERT cells. The protein (A) and the transcript (B) ratio KRAS$_{WT}$/KRAS$_{HRAS}$ vary between the different cell lines, whereas the transcript ratio when the construct is non-productive (C) is similar for all cell lines except HCT116. Results in are representative of three independent experiments with three technical replicates each. Values are relative to HEK293. Error bars represent the SD.*

Apart from analyzing families of triplets, we also investigate cancer gene families composed of two members. Interestingly, for some of them we observe a similar pattern as for the RAS family where the most mutated gene has a codon usage characteristic of genes involved in proliferation; for instance: RXRG, MAP2K1, PTPN11, CCND2, TBX3, ANK3 and ALK (Figure 2-20). However, in the case of gene pairs we do not find a significant negative covariance as we observe in Figure 2-4 when we compare with background genes.

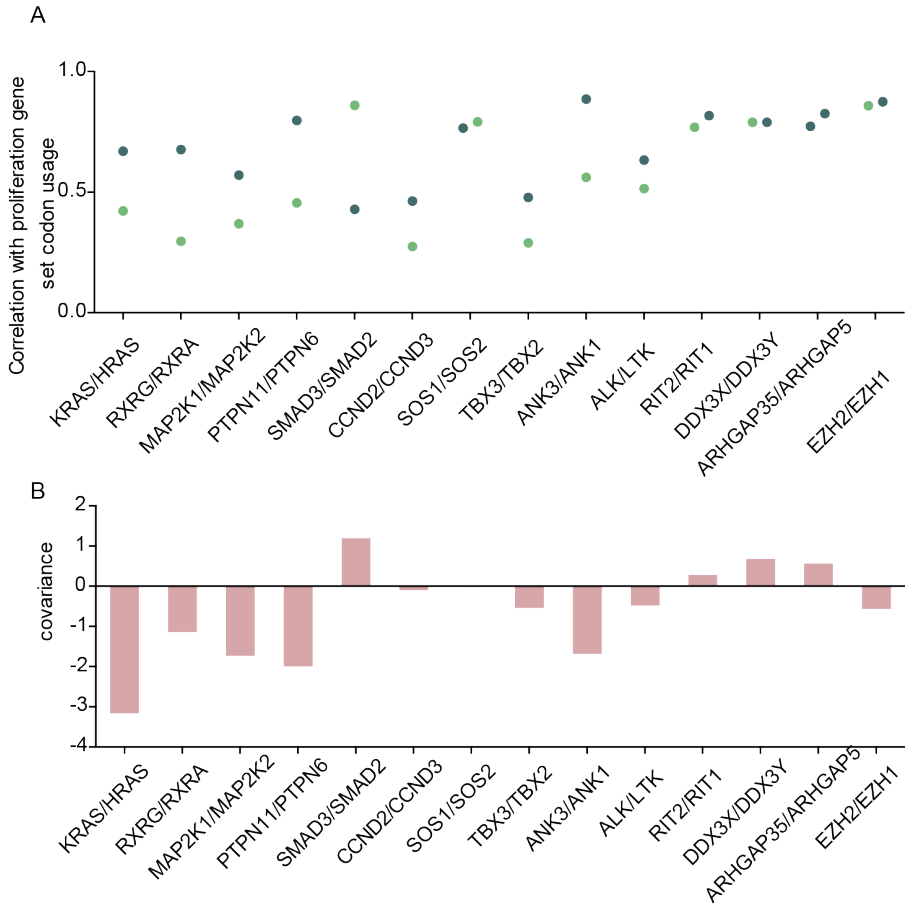



**Figure 2-20 Cancer gene pairs covariance codon usage and mutation counts.**

*A) Represents the correlation between the average codon frequency of a proliferation-related gene set and the codon frequency of each coding gene sequence. In blue the genes with higher mutations counts, in green gene with lower mutation counts. For comparison, we use the example of KRAS and HRAS. B) The covariance between mutation frequency and PC1 for each family.*

Finally, we also correlate the codon frequencies of $KRAS_{HRAS}$ with the average codon frequencies of proliferation- and differentiation-related genes (Figure 2-21). We observe that the pattern of $KRAS_{HRAS}$ is similar to HRAS, i.e. their codon usage correlates with the differentiation gene set better than KRAS. Regarding, the proliferation-related codon usage the differences between KRAS and $KRAS_{HRAS}$ are smaller.

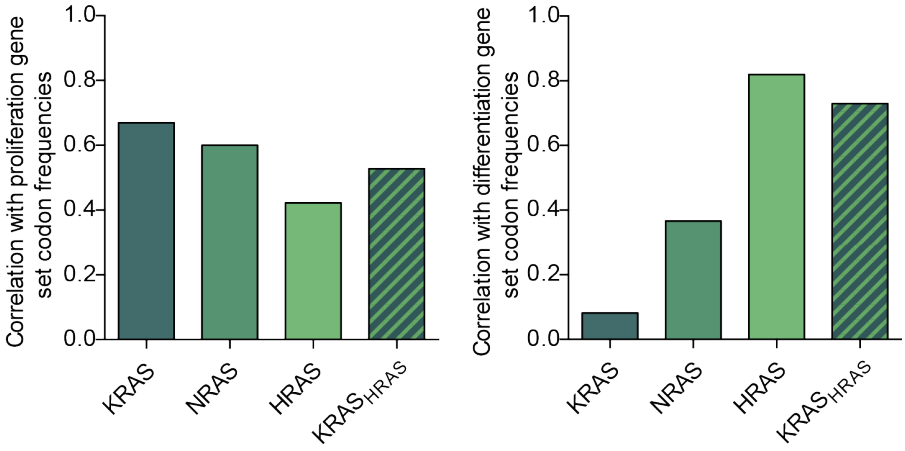



**Figure 2-21 Correlation of average codon frequencies of proliferation and differentiation gene set with the codon frequencies of RAS genes.**

*We correlate the reported average codon frequencies of two different gene sets (proliferation and differentiation) reported in TableS1 from Gingold et al., 2014 with the codon usage of RAS genes.*

## 2.6 Methods

### 2.6.1 Data Sources

**Paralogs Ensembl**

To define gene families, we retrieved protein sequence similarity and family membership information from Ensembl. As we observed that Ensembl's family classification often contained outliers with much lower sequence similarity compared to the other proteins, we applied another, more stringent filter: for each family we computed the similarity distribution of all members to a consensus member. We then removed all family members with a similarity of less than the mean similarity minus one standard deviation. We only considered families with at least three members.

**TCGA**

Mutation data was obtained from The Cancer Genome Atlas (TCGA). We retrieved somatic mutations in coding regions for 20 cancer types: Bladder Urothelial Carcinoma, Breast invasive carcinoma, Cervical squamous cell carcinoma and endocervical adenocarcinoma, Colon adenocarcinoma, Glioblastoma multiforme, Head and Neck squamous cell carcinoma, Kidney renal clear cell carcinoma, Kidney renal papillary cell carcinoma, Acute

Myeloid Leukemia, Brain Lower Grade Glioma, Liver hepatocellular carcinoma, Lung adenocarcinoma, Lung squamous cell carcinoma, Pancreatic adenocarcinoma, Pheochromocytoma and Paraganglioma, Prostate adenocarcinoma, Skin Cutaneous Melanoma, Stomach adenocarcinoma, Thyroid carcinoma, and Uterine Corpus Endometrial Carcinoma comprising a set of 5,960 samples.

**Cancer gene catalogue**
We considered cancer driver genes to be those genes that had a significant ($q < 0.01$) number of non-silent mutations in at least 1 out of 21 cancer types in 4,742 patients as defined in Lawrence et al (Lawrence et al., 2014).

**Coding Sequences**
The coding sequences of *Homo sapiens* were downloaded from the Consensus CDS (CCDS) project (ftp://ftp.ncbi.nlm.nih.gov/pub/CCDS/) release 2016/09/08. In the case of non-cancer genes, one unique canonical coding sequence was arbitrarily chosen for each protein based on Uniprot mapping to the CCDS. In the case of genes in the selected cancer gene families, the canonical coding sequence was chosen following the corresponding protein defined as canonical in Uniprot.

**GO gene sets**
Gene ontology was downloaded as MySQL dump of the amiGO database release 2017/01, and human gene annotations were downloaded from amiGO database release 2018/01/04. We defined GO gene sets as follows: for each GO term, we retrieved all descendant GO terms (with any kind of relationship type) and assigned all associated genes. We selected all GO terms with a minimal distance to the root "biological process" term shorter or equal to 3, and at least 30 associated genes, resulting in a total of 708 gene sets. Note that there is a lot of overlap between these GO gene sets, with a protein appearing on average in 44 sets.

### 2.6.2 Computational analysis

**Codon usage PCA**
We applied principal component analysis (PCA) to the relative synonymous codon frequencies (Sharp and Li, 1987) of all individual human coding sequences. Note that, contrary to other studies such as the one in Gingold et al (Gingold et al., 2014), we defined our PCA projection based on the codon

usage distribution of individual genes, not of gene sets. By doing so, our projection is independent from the gene ontology annotations. In addition, PCA based on the average codon usage of gene sets may suffer from bias due to the fact that GO gene sets are highly overlapping. Thus, the codon usage of a specific gene may contribute to several gene set data points, which may in turn distort the real variation in codon usage. When computing the PCA of individual genes, we first excluded single codon families (AUG, UGG). In the case of coding sequences that lacked codons of a specific family (6.7% of total), we impute values with the average codon frequency across all genes. We applied the PCA projection to GO gene sets, by computing the mean of relative codon frequencies of all genes in the set.

**Quantification of tRNA expression**

tRNAseq mapping was performed using a specific pipeline for tRNAs (Hoffmann et al., 2018). The basic pipeline was adapted to paired-end sequencing data. Moreover, given that hydro-tRNA-seq yields short sequences, all reads over 10 nt were included after BBDuk adapter trimming. Isoacceptors were quantified as reads per million (RPM), summing up all reads mapping to isodecoders that share the same anticodon. Ambiguous reads mapping to genes of different isoacceptors were discarded. The human reference genome GRCh38 (GenBank 2339568) was used.

**Relative codon usage**

We correlated the relative codon usage of $KRAS_{WT}$ and $KRAS_{HRAS}$ (calculated by dividing each codon value by the sum of the codon values of a given amino acid). In order to be able to calculate the fold change we added a pseudo count to all values (+1). For the 4 other families we calculated this fold change of codon usage between the most mutated gene and the less mutated gene from the same family. We first performed a sequence alignment using TranslatorX (Abascal et al., 2010) to be able to compare only the codons that align between the two sequences. Finally, we calculated the relative codon usage and the fold change in the same way as done for the comparison between $KRAS_{WT}$ and $KRAS_{HRAS}$.

**Differential tRNA anticodon abundance**

We exclude anticodons for which there are no corresponding tRNA genes ($Arg^{GCG}$, $Gly^{ACC}$, $His^{ATG}$, $Leu^{GAG}$, $Phe^{AAA}$, $Thr^{GGT}$ and $Val^{GAC}$) based on the tRNA gene prediction from the *H. sapiens* genome GRCh38/hg38 using

tRNAscan-SE (Chan et al.). Next, we calculate the relative anticodon abundance: dividing each anticodon rpm value by the sum of the anticodons rpm values for a given amino acid). Differential relative expression analysis was performed using t-test, where p-values were FDR-corrected, with $p < 0.05$ as a cutoff.

### 2.6.3 Sample preparation and experimental procedures

**Cell lines**

The cell lines included in this study were comprised of: HeLa, HEK293, HCT116 and fibroblast BJ/hTERT (used in Gingold et al.(Gingold et al., 2014), kindly provided by the author Disa Tehler). Cells were maintained at 37 °C in a humidified atmosphere at 5% $CO_2$ in DMEM 4.5g/L Glucose with UltraGlutamine media supplemented with 10% of Tet-free FBS (Clontech) and 1% penicillin/streptomycin.

**Expression vector design**

$KRAS_{HRAS}$ was obtained from pBABE-Puro-KRas* (Addgene#46745). For conditional-gene overexpression experiments, $KRAS_{WT}$ and $KRAS_{HRAS}$ were cloned into a modified version of the XLone-GFP vector (Randolph et al., 2017) (Addgene#96930). The modification consisted of replacing the promoter of XLone-GFP with a bidirectional TRE3G promoter (Clontech) allowing the simultaneous expression of both KRAS genes. We use a FLAG tag and 3xHA to distinguish FLAG-$KRAS_{WT}$ and 3xHA-$KRAS_{HRAS}$ by size. We also inverted the tags FLAG-$KRAS_{HRAS}$ and 3xHA-$KRAS_{WT.}$ The vector was co-transfected in different cell lines with the plasmid pCYL43 (Wang et al., 2008) containing the PiggyBac transposase. Cells were selected with blasticidin (HeLa: 5µg/mL, HEK293: 15µg/mL, BJ/hTERT: 5µg/mL). Gene expression was induced with doxycycline (HeLa: 100ng/mL, HEK293: 12ng/mL, BJ/hTERT: 500ng/mL).

**Serum starvation assay**

BJ/hTERT were grown in starvation media (1% Tet-free FBS) or non-starvation media (10% Tet-free FBS) for 48 hours. The expression of both $KRAS_{WT}$ and $KRAS_{HRAS}$ was measured after doxycycline induction overnight.

**Cell lines assay**

Established HeLa, HEK293, HCT116 and BJ/hTERT cells were induced with doxycycline and the expression was measured after overnight incubation.

**Cell growth**

The cells were seeded at a density of 25000 cells per well in a 12-well plate and the counts were performed with Countess cell counting chamber slides and the Countess automated cell counter (ThermoFisher). The counts were carried out every 24 hours.

**mRNA quantification**

RNA isolation was performed with RNeasy kit (Qiagen). $KRAS_{WT}$ and $KRAS_{HRAS}$ transcript abundances were quantified by RT-qPCR (Power SYBR Green RNA-to-CT 1-Step Kit, ThermoFisher). See Table 2-3 for the primers used for RT-qPCR. As both genes are in the same expression cassette, for each sample, the Ct values for $KRAS_{WT}$ were normalized to the $KRAS_{HRAS}$, $\Delta Ct = (CtKRAS_{WT} - CtKRAS_{HRAS})$ and represented as $2^{-\Delta Ct}$.

| Primer | Sequence |
|---|---|
| FLAG-$KRAS_{WT}$ fwd | 5'-CAAGGACGACGATGACAAG-3' |
| FLAG-$KRAS_{WT}$ rev | 5'-GAGAATATCCAAGAGACAGGTT-3' |
| GADPH fwd | 5'-GAGTCAACGGATTTGGTCGT-3' |
| GADPH rev | 5'-TTGATTTTGGAGGGATCTCG-3' |
| 3HA-$KRAS_{HRAS}$ fwd | 5'-CCTGACTATGCGGGCTATC-3' |
| 3HA-$KRAS_{HRAS}$ rev | 5'-GGGTCGTATTCGTCCACAA-3' |

**Table 2-3 RT-qPCR primers**

**Quantitative protein blots**

Cells were lysed using an M-PER buffer (ThermoFisher) supplemented with anti-proteases. Protein concentration was measured using a BCA Protein Assay Kit (Pierce). Equal amounts of each sample were mixed with 1x Laemmli buffer and boiled for 5 min. Samples were separated using 12% polyacrylamide gels (BioRad). Transfer was performed using the iBlot system (Invitrogen). Membranes were treated with Li-COR Odyssey blocking buffer for 1 hr at RT, then incubated with primary antibody (1:1000) in 0.2% Tween-20/Li-COR odyssey blocking buffer overnight at 4ºC. Following three 5 min washes in TBS-T, the membrane was incubated with secondary antibodies (1:10000) in 0.2% Tween-20/Li-COR Odyssey

blocking buffer for 45 min at RT. Following three 5 min washes in TBS-T, the membrane was scanned using the Li-COR Odyssey Imaging System. We used the following primary antibodies: anti-pan-RAS (Abcam, ab52939) and anti-β-actin (Sigma, A2228) and were detected using a goat anti-rabbit (Abcam, ab216773) or goat anti-mouse (Abcam, ab216776) IgG antibody conjugated to an IRdye at 800CW and 680CW, respectively. Visualization and quantification were done using ImageJ and Image Studio Lite (LI-COR).

**Hydro-tRNA sequencing**

Total RNA from HEK293 and HeLa was extracted using the miRNeasy Mini kit. For each sample, 20 μg of total RNA was treated following the protocol of hydro-tRNAseq (Gogakos et al., 2017).Total RNA was resolved on 15% Novex TBE Urea gels (ThermoFisher) and size-selected for 60-100 nt fragments. The recovered material was then alkaline hydrolyzed in 100mM sodium carbonate and 100mM sodium bicarbonate, followed by an incubation at 60ºC for 10 minutes. The resulting RNA was de-phosphorylated with Antartic Phosphatase (New England Biolabs) at 37ºC for 1 hour. De-phosphorylated RNA was purified and re-phosphorylated with Polynucleotide kinase (New England Biolabs). PNK-treated RNAs were purified and similar to small RNA sequencing library preparation, adaptor-ligated, reverse-transcribed and PCR-amplified for 14 cycles. The resulting cDNA was purified and sequencing was done on Illumina HiSeq 2500 platform in 50bp paired-end format. Raw data have been deposited in the ArrayExpress database (Kolesnikov et al., 2015) at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-8144.

# Chapter 3

# 3 Implications of point mutations on KRAS protein abundance

## 3.1 Summary

As discussed in the previous chapters, RAS genes show particular mutation patterns. For example, the most frequently mutated member of the RAS family varies between different cancer types. We showed how changes in protein abundance might contribute to the prevalence of KRAS mutants. This might also apply to the type of substitution.
In this chapter we investigate the molecular reasons behind the observation that RAS mutants show a mutation-specific protein abundance. Oncogenic mutations block KRAS in a constitutively active state and thus, enable the continuous interaction of KRAS with its effectors. This is in contrast with the wild type form, which switches between the active and inactive state. Here, we investigate how protein interactions contribute to KRAS protein levels and specifically we investigate mutation-specific protein degradation rates.

## 3.2 Introduction

In many oncogenes non-synonymous mutations are enriched at specific amino acid positions (hotspot mutations). KRAS mutations occur mainly at codons 12, 13 and 61 (Prior et al., 2012). Even though all of these mutations are activating, they are not equally transforming (Seeburg et al., 1984). Biochemical analyses have shown mutation-specific effects on RAS nucleotide binding, GTP hydrolysis and effector interactions, that cause differences in signaling outputs (Hunter et al., 2015; Smith et al., 2013).

RAS has multiple downstream effectors, e.g. RAF1, PI3Kα, RALGDS, PLCε, and RASSF5 (see Table 1-1) (Bunney and Katan, 2006; Castellano and Downward, 2011; Kiel et al., 2004; Ramocki et al., 1998; Stieglitz et al.,

2008; Wittinghofer and Herrmann, 1995). Structural analyses revealed that the interaction interface is the same for the effectors (see Figure 1-3). All bind to the same domain in RAS, the so called effector binding domain ranging from residues I21 to Y64 (Nakhaeizadeh et al., 2016). This includes the change of conformation upon activation of the structural elements switch I (residues 30-38) and switch II (residues 59-67) which is essential for all effector interactions. As presented in more detail in the introduction of this thesis (see Section 1.2.2.1), early studies have shown that distinct amino acids (e.g., T35, E37, and Y40) dictate effector specificity (White et al., 1995). Remarkably, germline mutations found in the effector binding domain cause developmental disorders such as the RASopathies and some of these mutations occur at positions P34 and I36 (Jindal et al., 2015) which correspond to the effector binding domain. They have been shown to affect the strength of the interaction with RAF1 (Gremer et al., 2011). Additionally, oncogenic mutations in codons 12, 13 and 61 have been shown to have variable affinities for the effector RAF1 (Hunter et al., 2015).

The sum of all possible RAS effectors is higher than the amount of RAS (HRAS, KRAS and NRAS), therefore, as they all bind to a common domain, these interactions are mutually exclusive. Specifically, the competition for binding will be high if all the interactions occur in the same subcellular compartment. The differences in binding affinities, in protein concentration, and the localization of the proteins binding to RAS, might affect protein complex formation leading to changes of the signaling outputs. For example, changes in protein expression contribute to competition with a competitive advantage for proteins highly expressed. Indeed, it has been observed that this is a regulatory mechanism contributing to cell type-specific MAPK signaling (Kiel et al., 2013).

Working with different RAS mutations we made the observation that the protein abundance of exogenously expressed KRAS mutants varies depending on the substitution. In particular, we found that oncogenic mutations exhibit higher protein levels in comparison to $KRAS_{WT}$, suggesting that oncogenic mutations cause a change of the activation level as well as on the protein level. This could be due to a decrease of degradation rates of the activated RAS protein upon binding to an effector. To test this hypothesis, we use designed RAS effector domain mutants that selectively enable binding to a subset of five downstream effector proteins or block all effector binding. The predictions are based on structural data and experimentally validated. We

observe that differential binding to the effectors causes changes in the protein abundance of RAS. Finally, we try to elucidate to which extend differential protein degradation of the mutants plays a role in the mutant-specific protein abundance of RAS.

Additionally, we describe other cellular mechanism playing a role in our observation and discuss the possibility that multiple factors contribute to the increased abundance of oncogenic KRAS.

## 3.3 Results

### 3.3.1 $KRAS_{WT}$ and oncogenic KRAS have different protein abundance.

The initial idea to this project came when we observed that exogenously expressed $KRAS_{WT}$ and $KRAS_{G12D}$ display different protein abundances in mammalian cells. In a previous study, we have expressed the two variants under an inducible promoter and we have shown that different levels of inducer are needed in order to reach similar protein levels of WT and G12D (Beltran-Sastre et al., 2015). The Glycine 12 is located at the P-loop (Figure 1-2) and is especially sensitive to amino acid substitutions. Replacement of G12 with Aspartate (G12D) or any other residue sterically interferes with the GTP hydrolysis in the presence of GAPs. Thus, G12D mutant is no longer susceptible to the negative control by GAPs and spend a much longer period in the activated state in comparison to WT (Malumbres and Barbacid, 2003). Even though the observation of a high protein abundance of G12D applies to HRAS as well (NRAS was not tested), we decided to work with KRAS due to its more prominent role in oncogenesis. Here, the expression cassette contains $KRAS_{HRAS}$ instead of $KRAS_{WT}$ (see Chapter 2) as in most cell types $KRAS_{HRAS}$ has a high protein abundance that facilitates the detection and protein quantification by immunoblotting. For simplicity we refer to $KRAS_{HRAS}$ as KRAS in this Chapter.
First, we exogenously express with transient transfection FLAG-$KRAS_{WT}$ and FLAG-$KRAS_{G12D}$ using a doxycycline-inducible expression vector. We consistently observe differences in protein abundance with the oncogenic mutant G12D being more abundant than wild type KRAS in HeLa cells ($p < 0.05$; t test; Figure 3-1).

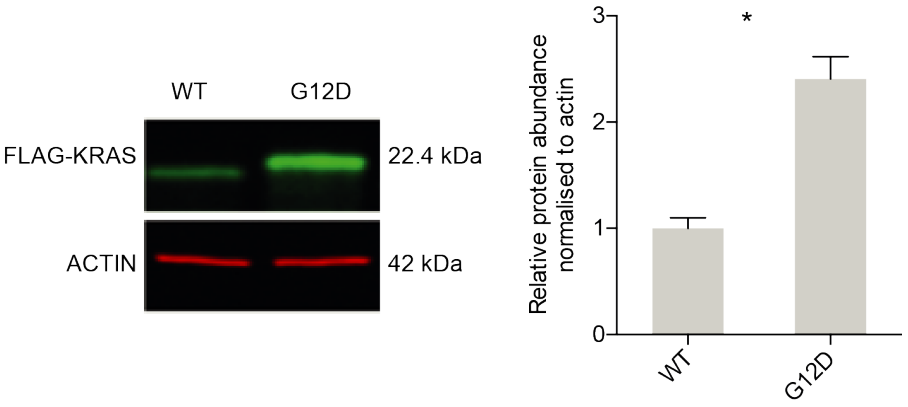



**Figure 3-1 Example of $KRAS_{WT}$ and $KRAS_{G12D}$ protein abundance in HeLa cells.**

*Western blot and protein quantification of FLAG-$KRAS_{WT}$ and FLAG-$KRAS_{G12D}$ protein levels in HeLa cells. Error bars represent the SD of three independent experiments with two technical replicates each. * $p < 0.05$ (unpaired Student t test).*

This observation applies to different cell types and expression systems. For example, mouse intestinal organoids $APC^{-/-}$ established with retroviruses to express FLAG-$KRAS_{WT}$ and FLAG-$KRAS_{G12D}$ show a difference in protein levels as well (Figure 3-2).

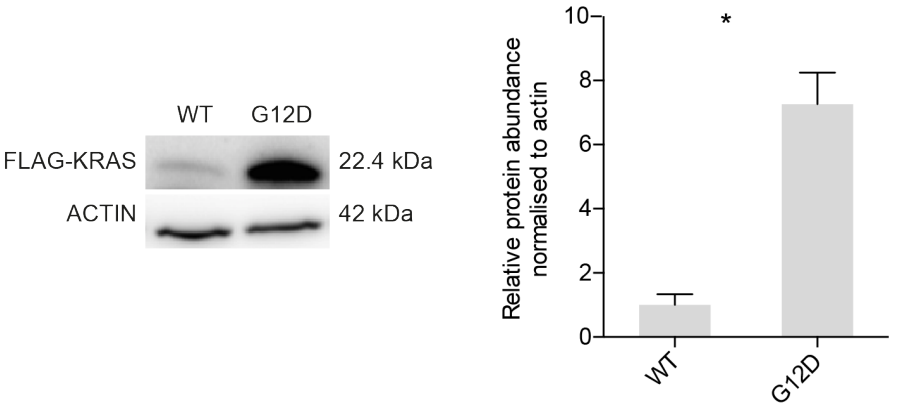



**Figure 3-2 $KRAS_{WT}$ and $KRAS_{G12D}$ protein abundance in mouse intestinal organoids.**

*Western blot and protein quantification of the over-expression of FLAG-$KRAS_{WT}$ and FLAG-$KRAS_{HRAS}$ in intestinal $APC^{-/-}$ organoids. Error bars represent the SD of two biological replicates (organoids derived from two different mice) with two technical replicates each. * $p < 0.05$ (unpaired Student t test).*

In addition, we take the same approach with intestinal organoids derived from $APC^{-/-}$-$KRAS_{G12D}$ mouse strain and $APC^{-/-}$-$KRAS_{WT}$. The expression of KRAS in these organoids occurs from the endogenous locus. The protein quantification shows a higher abundance of RAS when the $KRAS_{G12D}$ is expressed in comparison to the $KRAS_{WT}$ (Figure 3-3). However, the antibody used does not allow to distinguish the endogenous KRAS from endogenous HRAS and NRAS. Therefore, we cannot determine if the observed increase of protein levels between WT and G12D organoids is specific for KRAS or also occurs for HRAS and NRAS. In addition, we quantify the abundance of BRAF (Figure 3-3), a downstream signaling protein of RAS proteins, and compare its protein level between organoids expressing $KRAS_{WT}$ or $KRAS_{G12D}$. We observe that the abundance of endogenous BRAF is not increased in $KRAS_{G12D}$ organoids, suggesting that the effect on protein abundance is specific to KRAS.

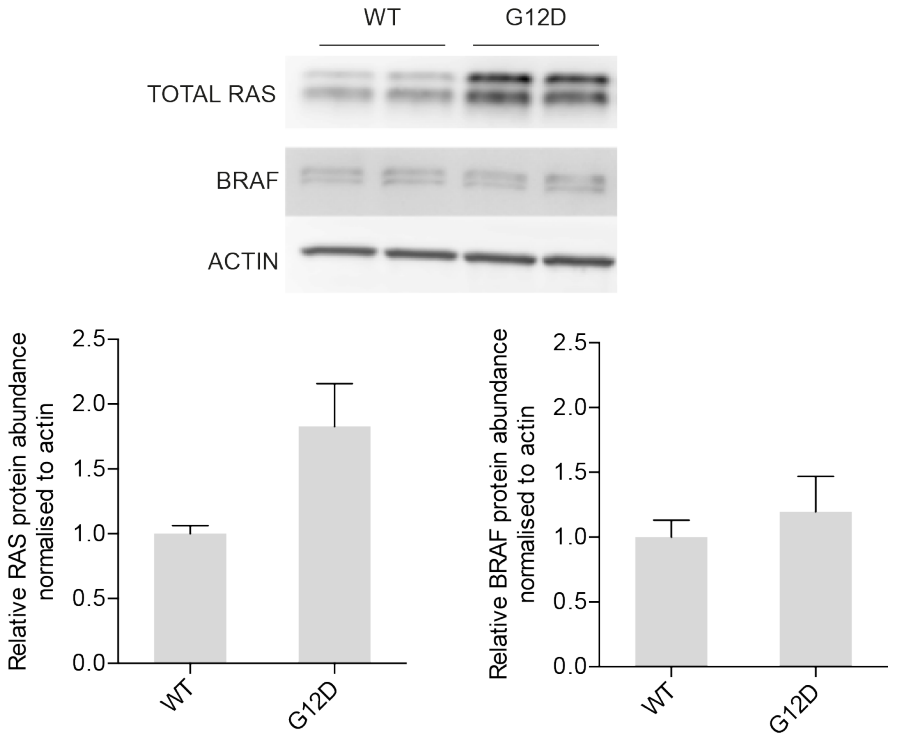



**Figure 3-3 KRAS expression in $APC^{-/-}$ $KRAS_{WT}$ and $APC^{-/-}$;$KRAS_{G12D}$ mouse intestinal organoids.**

*Western blot and protein quantification of RAS and BRAF in $APC^{-/-}$ and $APC^{-/-}$;$KRAS_{G12D}$ intestinal organoids. Error bars represent SD of two biological replicates (organoids derived from two different mice) with two technical replicates each.*

Over-expression of other oncogenic mutants of KRAS shows that their expression is also higher in comparison to the wild type (Figure 3-4). In summary, we observe a consistent increase of KRAS protein levels in different expression systems and cell types when KRAS harbors an oncogenic mutation.

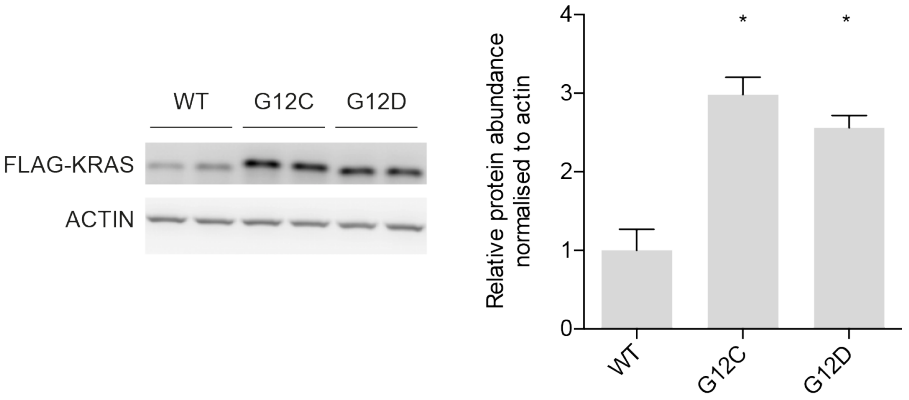



**Figure 3-4 Comparison between FLAG-KRAS WT, G12D and G12C in HeLa cells.**

*Western blot and protein quantification of FLAG-KRAS$_{WT}$, FLAG-KRAS$_{G12C}$ and FLAG-KRAS$_{G12D}$ expression in HeLa cells. Error bars represent the SD of two independent experiments with two technical replicates each.*

Consequently, based on the above observations, an oncogenic mutation is not only affecting the activation level of KRAS, but also its protein level. This suggest that protein abundance changes might be an additional level of perturbation of signaling pathways controlled by KRAS. Therefore, we aim to understand which mechanisms are leading to the mutant-specific changes in KRAS protein abundance.

### 3.3.2 Investigating possible mechanisms leading to different protein abundance.

Protein abundances can be described as the balance between protein synthesis and degradation. The former is determined by transcription, mRNA degradation, translation and protein degradation. Therefore, the observed difference of KRAS protein levels could be due to a number of factors:
Difference in protein synthesis:

- Oncogenic KRAS cause a global increase of protein synthesis.

- Differences in mRNA abundance (here, various mechanisms could explain an increase of the mRNA of KRAS).

Difference in protein degradation:

- G12D mutations stabilize KRAS.
- Lower protein degradation of $KRAS_{G12D}$ is due to its higher activity and thus binding to effectors that might protect from degradation.
    - Mutation causes delocalization of RAS to a compartment with lower degradation rates.
    - Mutation prevents binding of the degradation machinery.

#### 3.3.2.1 Changes in mRNA levels

A previous study based on data from The Cancer Genome Atlas (TCGA) has shown that KRAS gene expression in samples harboring KRAS mutations is higher in comparison to their $KRAS_{WT}$ counterparts in many but not all tumor types (Stephens et al., 2017). To test if we can reproduce this with our experimental set up, we measure the differences in mRNA levels between $KRAS_{WT}$ and $KRAS_{G12D}$ in HEK293 cells. We observe that $KRAS_{G12D}$ has a higher mRNA abundance in comparison to its WT form (Figure 3-5). This observation supports the previous finding that variation in protein synthesis related to mRNA levels contributes to the differences in protein abundance.

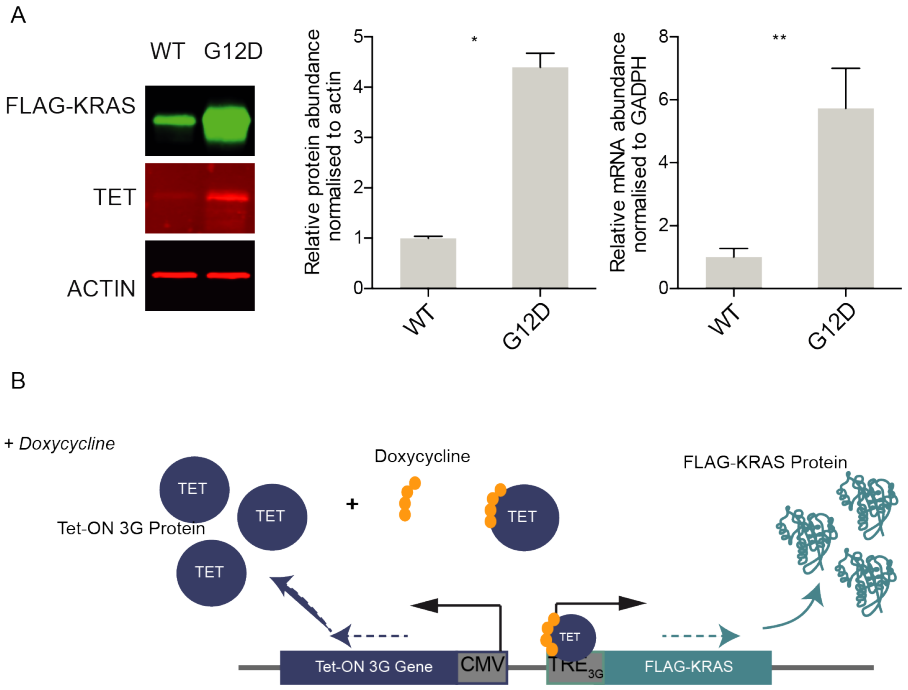



**Figure 3-5 Protein and mRNA levels of FLAG-KRAS$_{WT}$ and FLAG-KRAS$_{G12D}$ in HEK293 cells.**

*A) Comparison of the quantifications of protein levels and mRNA levels of FLAG-KRAS$_{WT}$ and FLAG-KRAS$_{G12D}$. Error bars represent SD of two (for protein) and three (for mRNA) independent experiments with two (for protein) and three (for mRNA) technical replicates each. * $p < 0.05$, ** $p < 0.01$ (unpaired Student t test). B) Expression vector system based on a doxycycline-inducible TRE3G promoter. The transcription of the TET gene is controlled by a CMV promoter. Once produced, the TET protein is able to bind to the TRE3G in presence of doxycycline and induce the expression of FLAG-KRAS.*

Surprisingly, we observe that the TET protein that upregulates the expression of KRAS in our construct and is expressed with the same vector (see Section 2.6, Figure 3-24), is also upregulated with FLAG-KRAS$_{G12D}$ both on transcript and protein level (Figure 3-6). Therefore, it could be that that the changes observed on protein levels are a bias of transfection, i.e. more molecules of the plasmid containing FLAG-KRAS$_{G12D}$ and TET have been transfected, and hence, we observe a higher abundance of both proteins. However, this hypothesis seems unlikely as we carefully measure and transfect the same amount of DNA for the WT and G12D construct. Also, the higher abundance of oncogenic KRAS is consistently observed in all our

transfections as well as when using a different expression system (e.g. retroviral infection in intestinal organoids, see Section 2.6).

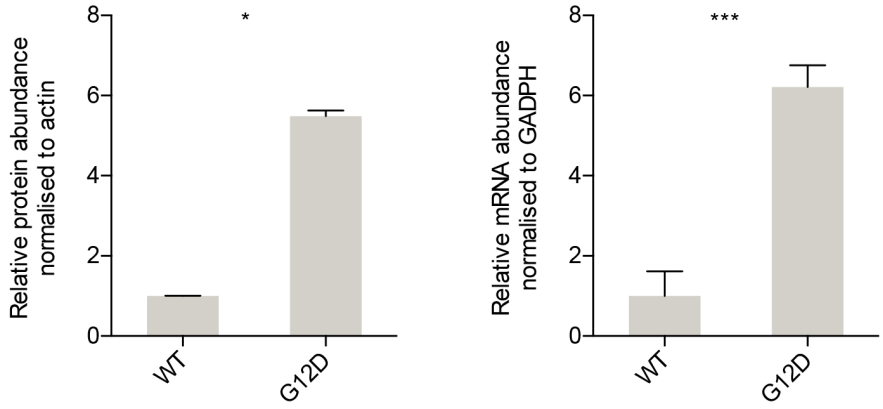



**Figure 3-6 Protein and mRNA levels of TET when $KRAS_{WT}$ and $KRAS_{G12D}$ are expressed in HEK293 cells.**

*Comparison of the quantifications of protein levels and mRNA levels of TET in cells expressing FLAG-$KRAS_{WT}$ or FLAG-$KRAS_{G12D}$. Figure related to results in Figure 3-5.Error bars represent SD of two (for protein) and three (for mRNA) independent experiments with two (for protein) and three (for mRNA) technical replicates each. * $p < 0.05$, *** $p < 0.005$ (unpaired Student t test).*

Notably, the expression of TET is controlled by a CMV promoter that has been shown to be over-activated by the MAPK pathway (Rodova et al., 2013). Therefore, it is possible that the expression of active KRAS activates the CMV promoter by activating the MAPK pathway, which in turn will induce more KRAS expression creating a positive feedback loop. Currently, we are testing a different expression vector in which TET is controlled by an EF1 promotor, which does not contain the same motif that has been described to cause the over-activation of CMV (see Figure 3-23 for preliminary results). Therefore, it remains to be elucidated to which extend KRAS expression is biased by this expression system. Nevertheless, it is clear that the $KRAS_{G12D}$ is upregulated as we have observed this in organoids where the expression system is not dependent on a CMV promoter and TET expression (Figure 3-2).

In the following, we will investigate and discuss the possibility that in addition to the apparent changes in mRNA levels, protein levels are directly affected by changes in the synthesis or degradation rates of KRAS.

#### 3.3.2.2 KRAS-mediated stimulation of protein synthesis

We wondered if a general increase of protein synthesis as a consequence of G12D induced RAS signaling could contribute to the differences in KRAS abundance. Indeed, a recent study has shown that RAS signaling promotes tRNA synthesis in Drosophila and also that this stimulation is required for RAS to promote cell growth (Sriskanthadevan-Pirahas et al., 2018). We would expect a general upregulation of protein and therefore we quantify the effect on other cellular proteins. The results above show that for example, the protein actin is not increased when $KRAS_{G12D}$ is expressed (Figure 3-1 to Figure 3-5). Additionally, we have quantified the total protein when both WT and G12D are expressed (Figure 3-7). We observe that there are no significant differences on global protein levels, suggesting that the higher amount of $KRAS_{G12D}$ is not caused by a general over-activation of protein synthesis in the cell.

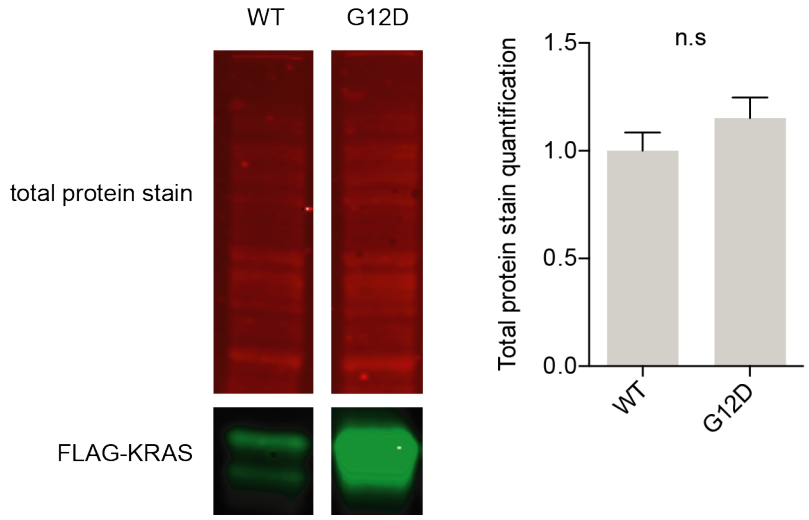



**Figure 3-7 Total protein quantification of HEK293 cell lysates with FLAG-KRAS$_{WT}$ and FLAG-KRAS$_{G12D}$ expression.**

*Western blot and protein quantification of total protein in HEK293 cells expressing KRAS$_{WT}$ and KRAS$_{G12D}$. A fluorescence-based protein stain is used for total protein detection and quantification as fluorescent signals are proportional to sample loading. Error bars represent the SD of two technical replicates.*

This result should be reconfirmed by another experimental approach such as protein quantification by mass spectrometry. Additionally, the quantification should be done distinguishing regulatory proteins from housekeeping proteins. The reason is that an increase of protein synthesis could potentially affect regulatory proteins to a stronger degree as compared to housekeeping proteins. Regulatory proteins respond to different stimuli and therefore have a larger dynamic range of expression. On the other side, the expression levels of housekeeping proteins are less dynamic and under stronger control in order to maintain baseline cellular functions and cellular homeostasis. Therefore, it would be interesting to quantify protein abundance of regulatory proteins only in cells expressing KRAS$_{WT}$ and KRAS$_{G12D}$.

### 3.3.2.3 Effect of G12D mutation on KRAS stability

Protein stability is directly related to the protein free energy. Some proteins are very stable, while others are more easily perturbed by point mutations which can lead to misfolding of the protein. The contribution of each position and amino acid is different, some might have a small effect to the folding free energy while others are crucial (Fersht and Serrano, 1993). The observed

difference in protein abundance between $KRAS_{WT}$ and $KRAS_{G12D}$ could be related to a change in protein stability. Therefore, we calculated the change in stability caused by the change from Glycine to Aspartic acid.
To do so, we use the *in silico* tool FoldX, which is co-developed and maintained by our group (Schymkowitz et al., 2005). FoldX allows to evaluate the direct effect of a point mutation on the overall stability of the protein based on 3D high-resolution crystallographic structures. It outputs the result as the difference in energy between the mutant and the wild-type structure ($\Delta\Delta G$ in kcal $mol^{-1}$).
Different RAS structures are available in the Protein Data Bank (PDB). We select five KRAS and HRAS crystallographic structures in the GTP-bound state and calculate the stability change with FoldX when introducing the mutation G12D (
Table 3-1).

| PDB id | Description | Stability change |
|---|---|---|
| 5VQ2 | $KRAS_{WT}$ in complex with GTP | 0.31 $kcal.mol^{-1}$ |
| 6GOD | $KRAS_{WT}$ in complex with GPPNHP | 0.68 $kcal.mol^{-1}$ |
| 6MBQ | $KRAS_{WT}$ in complex with GPPNHP | 0.29 $kcal.mol^{-1}$ |
| 4EFL | $HRAS_{WT}$ in complex with GPPNHP | 0.18 $kcal.mol^{-1}$ |
| 2RGE | $HRAS_{WT}$ in complex with GPPNHP | 0.38 $kcal.mol^{-1}$ |

**Table 3-1 Stability change by G12D mutations.**

*Five different 3D structures are used to predict the stability change when Glycine 12 is mutated to Aspartic acid. The stability change is calculated with FoldX (Schymkowitz et al., 2005).*

The standard deviation of the error reported for FoldX calculations is 0.46 kcal $mol^{-1}$, therefore any energies in the interval between -0.46 and +0.46 kcal $mol^{-1}$ are considered as not having an effect on protein stability (Schymkowitz et al., 2005). Therefore, our calculations suggest that the G12D mutation introduced in different 3D structures is not increasing protein stability. In one case (PDB 6GOD) we predict even a slightly destabilizing effect with 0.68 $kcal.mol^{-1}$. Hence, we conclude that G12D likely has no effect on increasing

the stability of RAS that could explain the increase of protein abundance observed in cells.

#### 3.3.2.4 Effector binding protects from degradation

Next, we ask whether protein connectivity influences protein turnover, reasoning that an active binding state might protect KRAS protein from fast degradation. This is related to previous reports, where highly interacting proteins have been shown to have prolonged half-lives (Cambridge et al., 2011; Martin-Perez and Villén, 2017). The KRAS oncogenic mutations cause a constitutive activation and thus, KRAS is able to continuously interact with its binding partners and to activate the corresponding signaling cascades. Therefore, we hypothesize that increased activation and hence interaction with effectors might protect KRAS from degradation. To test this hypothesis, we mutate KRAS with a very well-studied effector binding mutation D38A, which has been shown to decrease the ability of KRAS to interact with its different effectors. The residue Asp38 makes important contacts with the RAS binding domains of the effectors. These contacts are perturbed by the mutation to Alanine reducing the binding affinity (e.g. affinity to RAF1 has been shown to be reduced by 72-fold (Herrmann et al., 1995). Therefore, opposite to the oncogenic G12D mutant, D38A prevents RAS-mediated activation of these pathways. We observe that $KRAS_{D38A}$ protein level is lower in comparison to $KRAS_{WT}$. In addition, we compare protein levels between $KRAS_{G12D}$ and $KRAS_{D38A\text{-}G12D}$ mutant and observe again a decrease in the protein level of the effector domain KRAS mutant D38A (Figure 3-8).

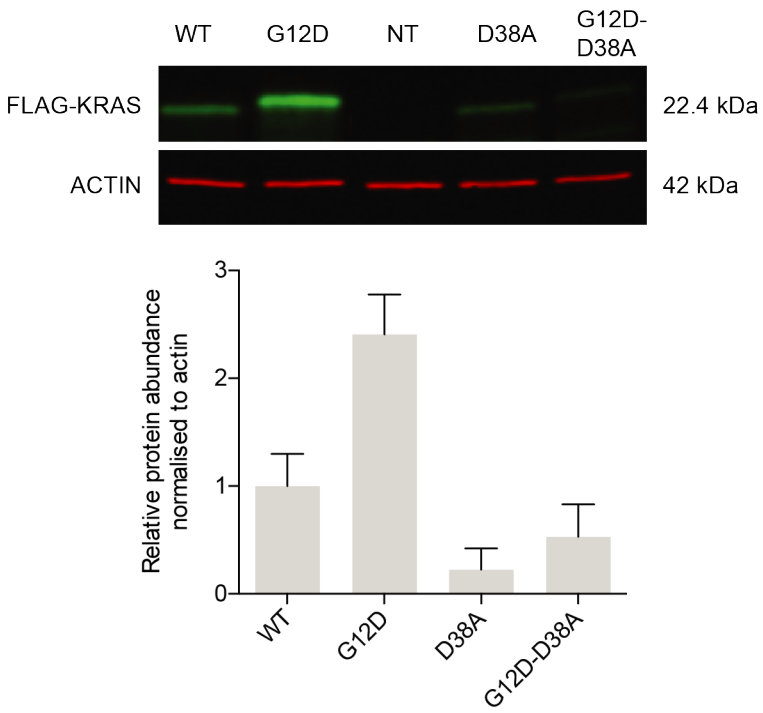



**Figure 3-8 Protein abundance comparison between effector binding mutant D38A, WT and G12D.**

*Quantification of protein abundance of exogenously expressed FLAG-KRASWT, D38A, G12D, and G12D-D38A in HeLa cells. Error bars represent the SD of two independent experiments with two technical replicates each.*

Therefore, we hypothesize that increased KRAS activation and interaction with effectors would lead to an avoidance of KRAS degradation, whereas the suppression of binding to effectors by the D38A mutations would lead to a faster degradation of the protein. We observe that both D38A and G12D-D38A mutations lead to a lower KRAS protein abundance, suggesting that KRAS interactions with effectors play a role in KRAS protein abundance.
To verify this hypothesis, we perform a degradation experiment where we measure protein abundance at different time points after the addition of a protein synthesis inhibitor (cycloheximide) (Figure 3-9). If the higher interaction of $KRAS_{G12D}$ with its effectors protects KRAS from degradation, we would expect a slower degradation in comparison to the least abundant KRAS mutant (D38A).

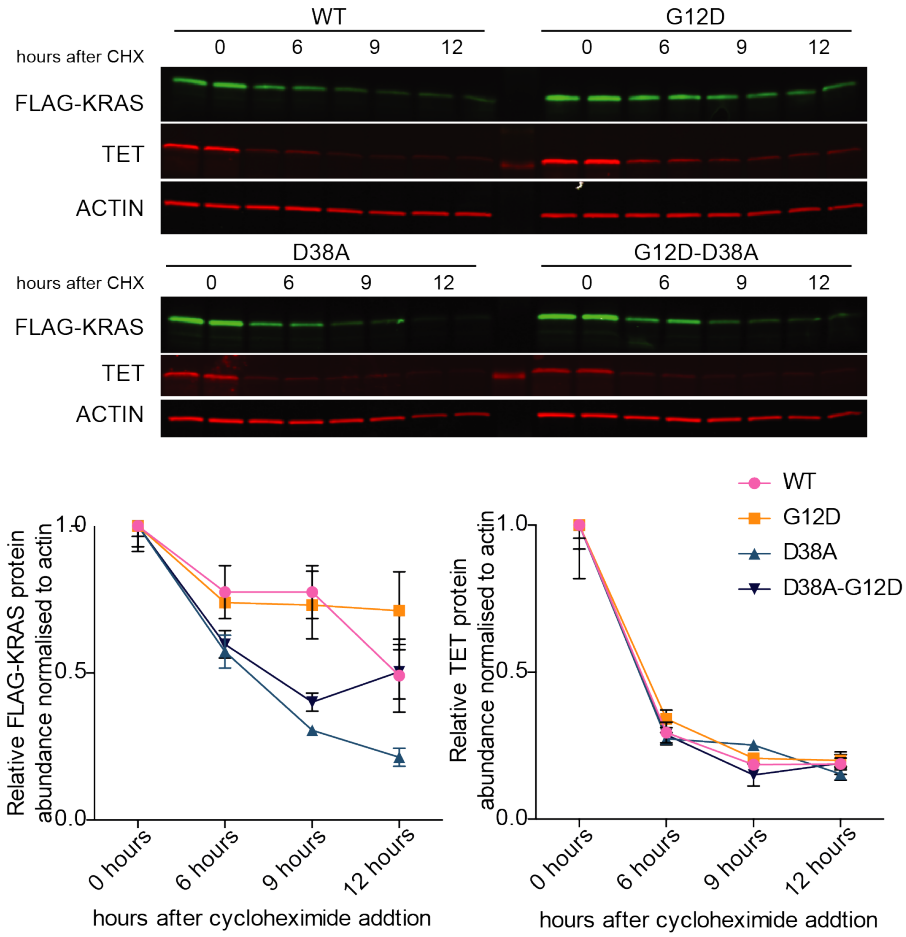



**Figure 3-9 Quantification of FLAG-KRAS WT and mutants at four different time points after cycloheximide addition.**

*Western blot and protein quantification (left panel) of $KRAS_{WT}$, $KRAS_{G12D}$, $KRAS_{D38A}$, and $KRAS_{D38A\text{-}G12D}$ at different times points after cycloheximide addition. Quantification of protein TET (right panel) when the different KRAS variants are expressed. Error bars represent the SD of two independent experiments with two technical replicates each.*

We observe that D38A and D38A-G12D mutants are faster degraded than the most highly expressed mutant G12D (Figure 3-9). Specifically, we find that the abundance of G12D is significantly different from the one of D38A at 9 and 12 hours after cycloheximide degradation (* $p < 0.05$, unpaired Student t test). To control that this effect is specific to the exogenously expressed KRAS mutants and not to other cellular proteins, we quantify additionally the abundance of the protein TET that is expressed from the same transfected

plasmid carrying KRAS. We observe that the degradation is similar independently of the variant of KRAS. This experiment shows that different mutant-specific protein degradation rates could play a role in the mutant-specific protein abundance we have observed.

Taking advantage of the expertise of the group in RAS protein design, we decide to use rationally designed KRAS mutants that lack some interactions with certain effectors. In this way, we will first test whether mutants have a specific protein abundance depending on their interactions and second it might be possible to understand if there is a specific interactor that plays a major role in protecting KRAS from degradation.

### 3.3.3 Designed RAS mutants

The *in silico* design of RAS mutants was previously done using available crystal structures of RAS in complex with different effectors. RAS is a signaling hub interacting with binding partners through the same interface. Five RAS-effector complex structures from the PDB database were used: RAS in complex with the RAS binding domain of PLCε, RALGDS, RAF1, PI3Kp110g, and RASSF5 (pdb ids: 1LFD, 1GUA, 1HE8, 3DDC, and 2C5L) (Figure 3-10).

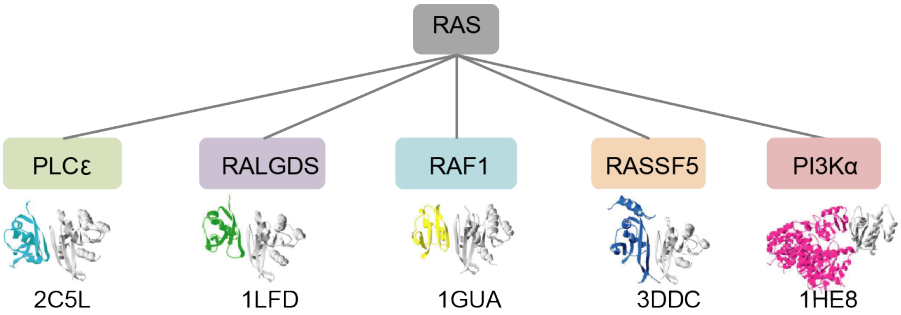



**Figure 3-10 RAS-effector structures.**

*RAS-effector structures used for in silico design with FoldX and their PDB IDs.*

For the *in silico* design of RAS mutants, the protein design tool FoldX was used. FoldX is a force field for energy calculations that not only provides a quantitative estimation of the importance of the amino acids contributing to the stability of proteins, as mentioned above, but also of protein complexes. Despite high similarity of the physical interface site, the detailed contributions of each amino acid of RAS involved in complex formation

might be different depending on the binding partner (Nakhaeizadeh et al., 2016; White et al., 1995). Thus, designed mutations of these amino acids might enable the selective interaction with partners. We have determined the predicted changes in interaction energies between RAS mutants and the different effectors (Figure 3-12A). Additionally, we have included the classical effector domain mutants Y40C and E37G. These ones are key residues for the binding specificity of RAS to different effectors (White et al. 1995; Khosravi-far et al. 1996) (See Section 1.2.2.1). It is necessary to keep in mind that the FoldX-based predictions do not consider the importance of the water molecules in the interaction interface. The biomolecular interface between RAS and the effectors contains important networks of water molecules that contribute to the binding energies of the complexes. Therefore, the results we include the implementation of a tool that predicts the effect of water-mediated interactions on the stability of the complexes (Figure 3-12B). For example, the residue Y40 has a water-mediated interaction with D38 (Figure 3-11). The substitution with a phenylalanine would perturb the stability. This is reflected in the prediction when combining FoldX with an additional tool that allows introducing the water molecules in the 3D structure (See Methods 3.6.1).

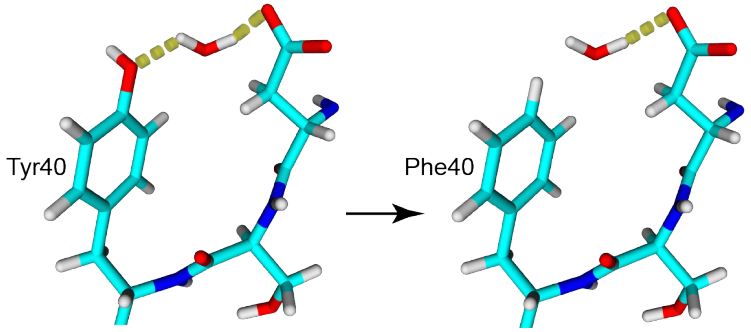



**Figure 3-11 Water-mediated interaction of RAS residues.**

*Example of a water-mediated interaction between Tyr40 and Asp38 that might perturb the stability of the complex (pdb: 3DDC).*

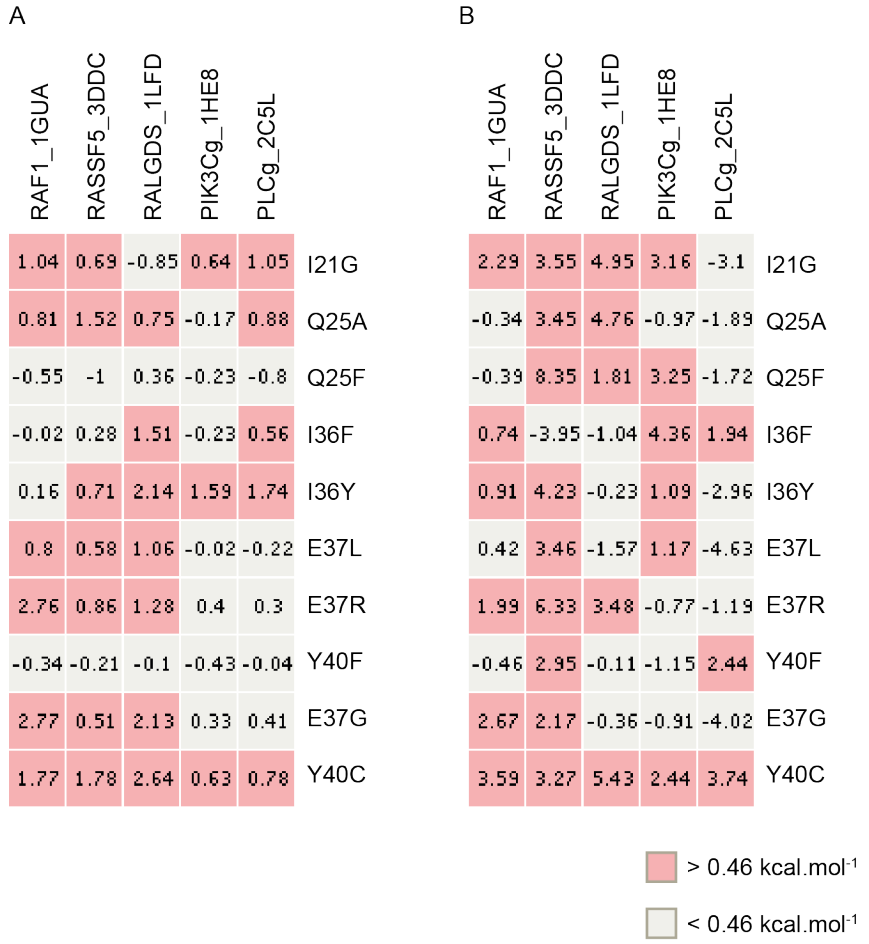



**Figure 3-12 Designed RAS mutants binding different downstream effectors.**

*Binary heatmap displaying the predicted values of interaction energy calculated by FoldX. A) The results correspond to predictions not considering water-mediated interactions. B) The results correspond to predictions considering the effect of water hydrogen bonds on the interaction interface. Values are represented in kcal.mol$^{-1}$. Red color highlights the perturbation of RAS-effector interaction.*

The selected mutants have predicted binding preferences for different partners and for the next part of this Chapter we based our comparisons on the results with water molecules (Figure 3-12B).

The predictions obtained with FoldX for the classical effector domain mutants (E37G and Y40C) agree with the literature. For example, the mutant E37G has been reported to hinder the interaction with RAF1 and PI3K, but to maintain the one with RALGDS and RASSF5, this is reproduced in our *in*

*silico* predictions. On the other hand, in agreement with FoldX, Y40C has been shown to eliminate binding with RASSF5 and RALGDS. Also, Y40C has been widely used because it selectively activates PI3K, but we obtain the opposite prediction. A possible explanation is that this interaction was measured using the class I PI3K p110α subtype (Rodriguez-Viciana et al., 1997) whereas the crystalized complex corresponds to p110γ, which was later shown to have a reduced binding to Y40C (Pacold et al., 2004). Therefore, the binding capacities of both E37G and Y40C mutants agree with the previously reported *in vitro* interactions, suggesting that the predictions of the structure-based designed mutants are correct.

### 3.3.4 Experimental validation of designed RAS mutants

#### 3.3.4.1 Validation of interactions

In order to test the interactions of the effectors with RAS mutants we use two different techniques. First we measure the affinities between RAS and the different effectors with Microscale Thermophoresis (MST), a method based on the mobility of biomolecules in a temperature gradient (Wienken et al., 2010). Here, we use purified HRAS mutants and effectors and compare the binding affinities with $HRAS_{WT}$ (See Methods 3.6.2). We use HRAS due to the availability of protocols to produce and purify recombinant HRAS. The effector domains of HRAS and KRAS are identical, therefore, we would not expect differences in *in vitro* measurements. Regarding the effectors, we failed to produce PI3K, thus the measurements have been done with RALGDS, RASSF5, RAF1 and PLCε. The results regarding these experiments are found Section 3.5.2 (Figure 3-22) and will be discussed below.

On the other hand, we have used a method that gives the possibility to monitor binding in a physiological context. This method has been recently developed and is based on the small luciferase NanoLuc. Briefly, NanoBRET is a bioluminescence resonance energy transfer (BRET)-based interaction assay that can detect interactions between proteins in mammalian cells (Machleidt et al., 2015). KRAS mutants are expressed with an N-terminal NanoLuc fusion and the effectors RAF1, PIK3CA, RALGDS, and RASSF5 are tagged in the C-terminal with HaloTag (Figure 3-13). Due to cloning difficulties the effector PLCε is not included in the following experiments. The advantage of using this NanoBRET is first, the expression and binding assessment of KRAS takes place in living cells. Second, the use of full-length effectors, in

comparison to only using the RAS binding domains in the crystal structures or other *in vitro* interaction measurements such a microscale thermophoresis. On the other hand, the drawback of this technique is the use of tags (NanoLuc and HaloTag) that could interfere with the protein-protein interactions.

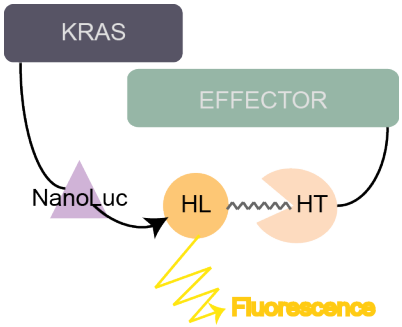



**Figure 3-13 Overview of the NanoBRET assay principle.**

*The energy is transferred from the luciferase of NanoLuc-KRAS (energy donor) to the labeled HaloTag(HT)-effector with a fluorophore ligand (HL) (energy acceptor) upon interaction between KRAS and the effector.*

NanoLuc-$KRAS_{mutant}$ and HaloTag-effector are co-expressed in HEK293 cells and their interaction is monitored calculating the ratio between the emissions of the acceptor and the donor. To avoid the effect of the interaction of KRAS with upstream interactors and just focus on the differential binding effect of the mutations with the downstream effectors, we decide to introduce to all KRAS mutants a second mutation consisting of G12D that blocks KRAS in its active state.

First, we assess the two methods with the WT forms of RAS (KRAS for NanoBret and HRAS for MST). In Figure 3-14-B we report the binding affinities ($K_d$ in μM) measured by MST with four effectors and compare with published data. We observe that our MST measurements for $HRAS_{WT}$-effector interactions completely overlap with the already published data. Next, we assess the method NanoBRET with NL-$KRAS_{G12D}$ (Figure 3-14-A). The binding partner used as negative control is P53 which is not known to interact with KRAS. The values are normalized to RAF1 as is known to be the effector with the highest binding affinity (80-100nM). We are able to confirm the KRAS-RAF1 and the KRAS-RASSF5 interactions. However, the results for testing the binding of KRAS to RALGDS and PIK3CA are not as clear as the standard deviation is overlapping with the negative control. We can think that the C-terminal tag in RALGDS and PIK3CA are not optimal

for testing the interaction with KRAS, and therefore, we should switch the HaloTag to the N-terminal. On the other hand, we compare the NanoBRET results of KRAS-effectors interactions with the binding affinities obtained with MST experiments (Figure 3-14-B, see also Figure 3-22). We observe that there is a correlation between the different methods; i.e. RAF1 and RASSF5 have binding affinities for HRAS in the nanomolar range, whereas PIK3CA and RALGDS are in the millimolar range, which agrees with the higher mBRET signal that we obtain for KRAS-RAF1 and KRAS-RASSF5 in comparison to KRAS-PIK3CA and KRAS-RALGDS. Therefore, we could hypothesize that the dynamic range of NanoBRET does not allow to distinguish low-affinity interactions from the negative control.
Regarding the mutation D38A, we observe that the binding specificity is lost as expected, and we obtain overlapping values with the negative control (Figure 3-14-B).

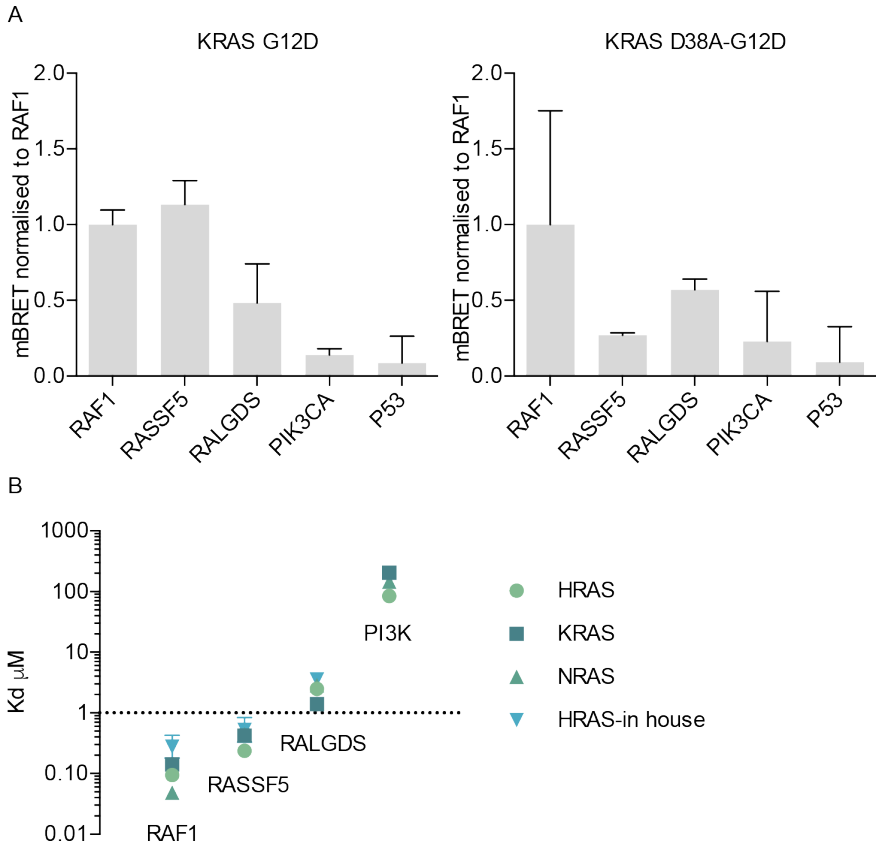



**Figure 3-14 RAS interactions with effectors measured with different methods.**

*A) NanoBRET assay monitoring the interactions between $KRAS_{G12D}$ and $KRAS_{D38A-G12D}$ with four different RAS effectors and P53 as negative control. Error bars represent two independent experiments with two technical replicates each. B) Binding affinities for HRAS, KRAS, NRAS obtained from Nakhaeizadeh et al. 2016 and from in-house experiments performed with Microscale Thermophoresis (see Section 3.5.2, Figure 3-22. Error bars represent two or three independent experiments of thermophoresis.*

Next, we perform the NanoBRET experiments with the designed mutants and the two classical effector domain mutants E37G and Y40C. Again, we confirm that the interactions with PIK3CA and RALGDS are difficult to quantify. Due to the high level of experimental noise, there are no significant differences between the interactions. Therefore, it is not possible to make conclusions about the interactions with these two effectors. On the other hand, for RAF1 and RASSF5 we observe significant differences between the binding capacities of the mutants (Figure 3-15).

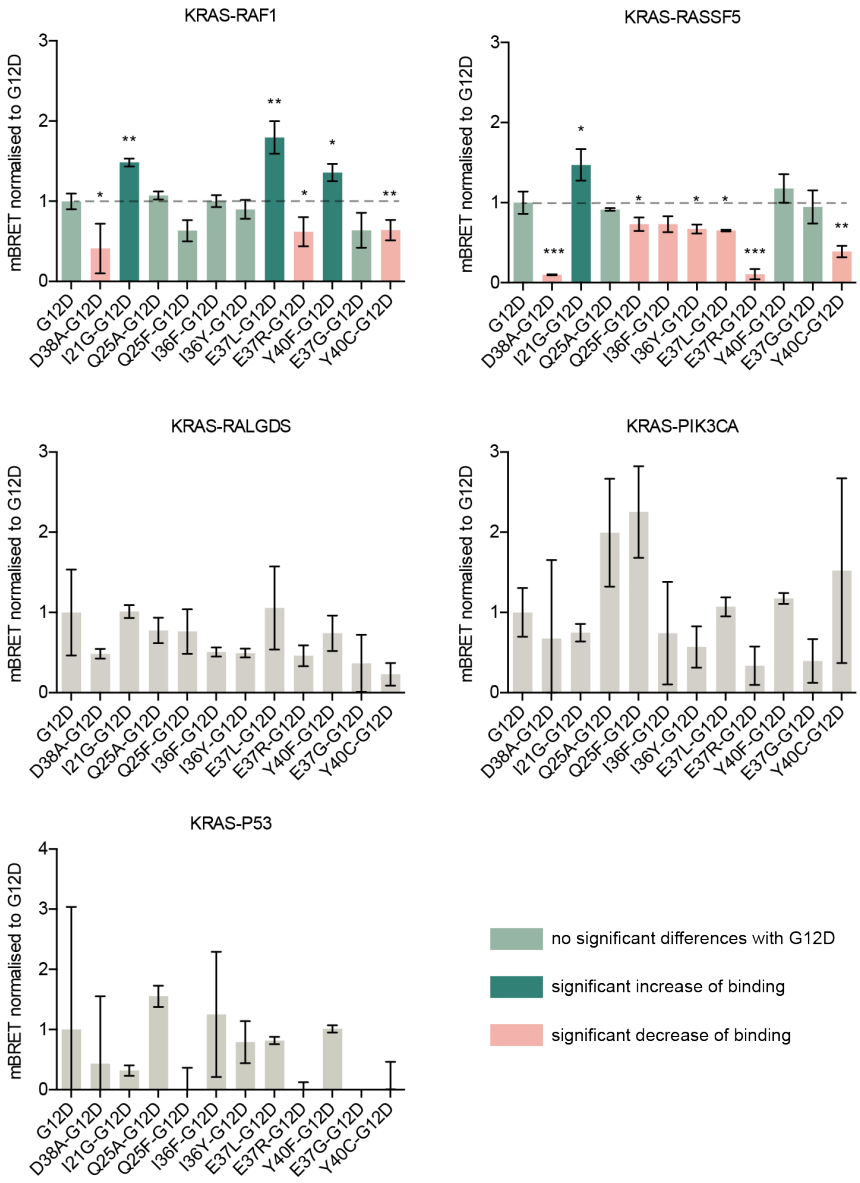



**Figure 3-15 NanoBRET results with KRAS$_{mutants}$-effectors interactions.**

*NanoBRET assay results for effector mutants with four different effectors and the negative control p53. Green bars highlight the mutants that bind to the effectors. Darker green corresponds to the mutants that significantly bind better to the effectors in comparison to G12D. Light red highlights significant decrease of interaction in comparison to G12D. Results are representative of three independent experiments with two technical replicates each. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$ (unpaired Student t test).*

We compare the results obtained with MST and NanoBRET with the FoldX predictions. We observe that for MST, 25 out 31 measurements are in agreement with the *in silico* predictions and 11 out 20 interactions for NanoBRET (
Table 3-2). The differences might be caused by a combination of experimental errors as well as the use of full-length effectors (for NanoBRET) in comparison to the crystallization of RAS with only the RAS binding domains of the effectors. Regarding, the classical mutants (E37G and Y40C) only measured with NanoBRET, we observe an overlap with FoldX predictions for all but one interaction (E37G with RAF1).

| | RAF1-RAS | | | RASSF5-RAS | | | RALGDS-RAS | | PLCe-RAS | |
|---|---|---|---|---|---|---|---|---|---|---|
| | FoldX | BRET | MST | FoldX | BRET | MST | FoldX | MST | FoldX | MST |
| I21G | | | | | | | | | | |
| Q25A | | | | | | | | | | |
| Q25F | | | | | | | | | | |
| I36F | | | | | | | | | | |
| I36Y | | | | | | | | | | |
| E37L | | | | | | | | | | |
| E37R | | | | | | | | | | |
| Y40F | | | | | | | | n.a | | |
| E37G | | | n.a | | | n.a | | n.a | | n.a |
| Y40C | | | n.a | | | n.a | | n.a | | n.a |

**Table 3-2 Comparison of predicted and measured interactions.**

*Comparison of FoldX predictions with MST and NanoBRET experiments. Red color represents non-binding whereas white represents binding.*

#### 3.3.4.2 Mutant-specific protein abundances

Next, we exogenously express all designed KRAS mutants in HEK293 with the inducible expression system (Figure 3-24). We also include a second mutation (G12D) to all the designed KRAS mutants as we have done with the expression constructs in NanoBRET experiments. Remarkably, the exogenous expression of the different mutants shows that there are KRAS mutant-specific protein levels. These results support the hypothesis that the interaction with the effectors play an important role in regulating KRAS protein levels. Interestingly, different substitutions in the same amino acid can lead to opposite levels of KRAS (Figure 3-16). For instance, the

abundance of Y40F-G12D is higher in comparison to Y40C-G12D, as well as E37L-G12D in comparison to E37R-G12D and E37G-G12D.
Interestingly, the mutants displaying a high protein abundance (comparable to the abundance of the G12D mutant) have shown to bind RAF1 in NanoBRET experiments (see Figure 3-15). This is the case for I21G, Q25A, Q25F, E37L, and Y40F (except E37G). Regarding the abundances of I36F and I36Y, which also bind to RAF1 in NanoBRET, we are not able to detect the protein by immunoblotting. The residue I36 corresponds to the epitope recognized by the RAS antibody we use and therefore, when the residue is mutated, we are not able to quantify the abundance with this antibody. We use the antibody recognizing FLAG tag in lysates from intestinal organoids (see Appendix, Figure 3-20). Interestingly, we observe that the growth of the organoids agrees with the abundance of the respective mutant, suggesting that the level of KRAS plays an important role on cell growth (see Appendix, Figure 3-21).

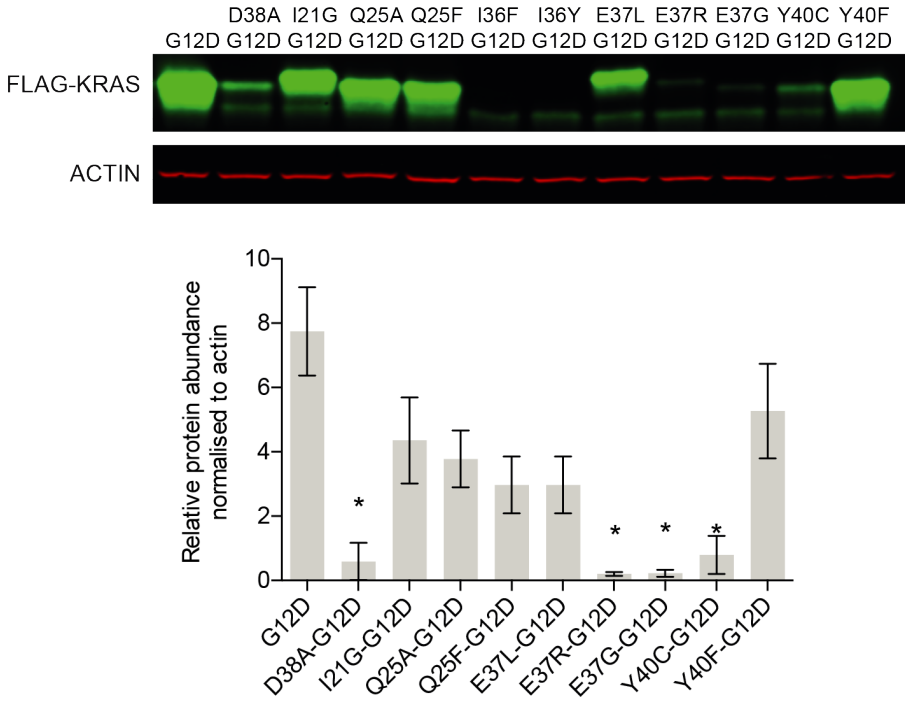



**Figure 3-16 Expression of FLAG-KRAS$_{mutants}$ in HEK293 cells.**

*Western blot and protein quantification of KRAS$_{mutants}$ in HEK293 cells. Error bars represent the SD of three independent experiments with two technical replicates each. * $p < 0.05$ (unpaired Student t test).*

### 3.3.4.3 Mutant-specific protein degradation

Next, we perform the same previous degradation experiment to test the possible different degradation between KRAS mutants. For a first test we choose the mutants Y40C-G12D and Y40F-G12D, that show a high difference in abundance. We carry out the same experiment presented previously in section 3.3.2.4 (cycloheximide blocking) and quantify the protein abundance differences over time between Y40C-G12D and Y40F-G12D (Figure 3-17).

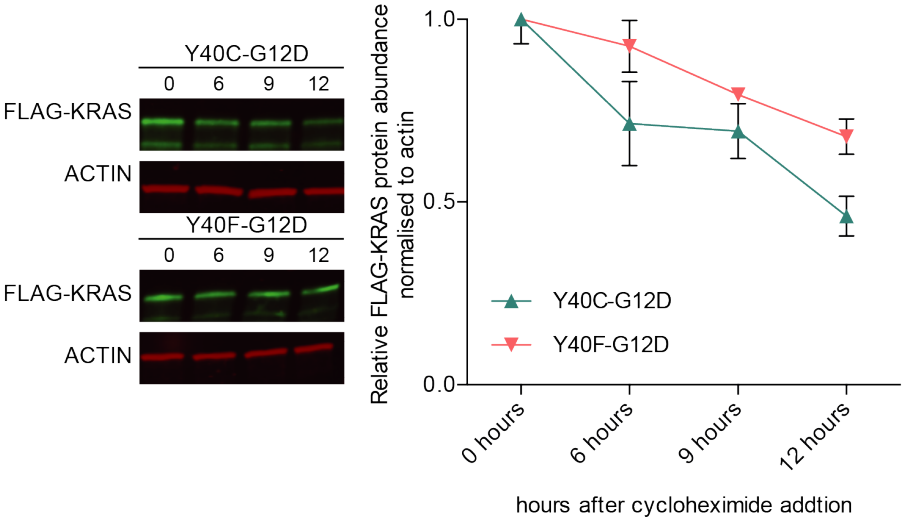



**Figure 3-17 Degradation experiment with Y40C-G12D and Y40F-G12D mutants.**

*Western blot and protein quantification of $KRAS_{Y40C-G12D}$, and $KRAS_{Y40F-G12D}$ at different times points after cycloheximide addition. Error bars represent the SD of two independent experiments with two technical replicates each.*

We observe that the less abundant mutant (Y40C) is significantly more degraded than Y40F (* $p < 0.05$, unpaired Student t test). However, the difference observed at 12 hours is approximatively 1.45-fold in comparison to ~5-fold change of protein abundance (see Figure 3-16). Ideally, we would like to support these results with an alternative method to measure protein degradation. Indeed, we would expect to have higher differences in degradation rates if protein degradation was the only mechanism playing a role in the different mutant-specific KRAS abundances. Therefore, we do not exclude that other molecular mechanisms contribute to mutant-specific protein abundances.

### 3.3.5 KRAS-RAF1 interaction increases KRAS protein abundance.

Additionally, we decide to test the effect of the co-expressing of KRAS together with RAF1. Indeed, we observe that mutants that bind to RAF1 display a higher protein abundance, so we hypothesize that the interaction might prevent protein KRAS degradation in comparison to the mutants that are not able to bind RAF1. Interestingly, we observe an increase of KRAS when RAF1 and PIK3CA are over-expressed (Figure 3-18).

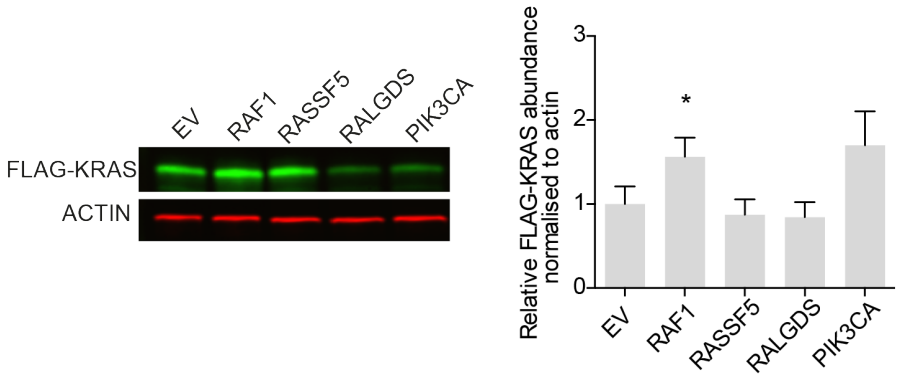
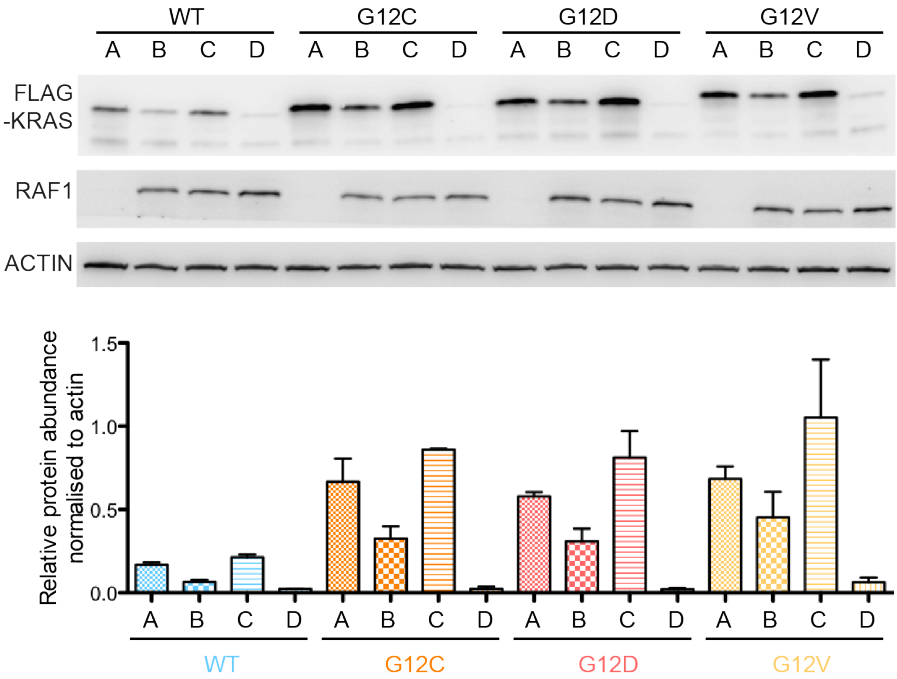



**Figure 3-18 Effectors co-expression with FLAG-KRAS$_{WT}$.**

*Quantification of protein levels FLAG-KRAS$_{WT}$ co-expressed one by one with four different effectors and empty vector (EV) as control. Error bars represent SD three independent experiments. * $p < 0.05$, (unpaired Student t test).*

Next, we repeat the same experiment but using RAF1 with a kinase dead mutation (K375M), which does not transduce downstream signals. Additionally, we express KRAS in the presence of RAF1$_{R89L\text{-}K375M}$ which harbors (in addition to the kinase activity-inhibition mutation) a mutation preventing the binding of RAF1 to KRAS (Figure 3-19).




A: KRAS only

B: KRAS + $RAF1_{K375M}$ (kinase dead mutant)

C: KRAS + $RAF1_{R89L,K375M}$ (kinase dead mutant and not binding to KRAS)

D: $KRAS_{D38A}$ + $RAF1_{K375M}$ (kinase dead mutant)

**Figure 3-19 Co-expression of FLAG-KRAS with kinase dead RAF1.**

*Quantification of protein levels FLAG-$KRAS_{WT/G12C/G12D/G12V/D38A}$ expressed either alone or co-expressed with $RAF1_{K375M}$ and $RAF1_{K375M-R89L}$. Error bars represent SD two independent experiments.*

Surprisingly, we observe that KRAS levels decrease when co-expressed with inactive RAF1 ($RAF1_{K375M}$). The level of protein is recovered when RAF1 harbors a mutation that do not allow the binding to KRAS. These results suggest that not only the physical interaction with RAF1 contributes to increase KRAS levels, but also the activity of the complex. We could hypothesize that the cell would target the degradation of inactive protein complexes such as KRAS-$RAF1_{K375M}$. We should perform localization experiments that could give a hint about the localization of the inactive KRAS-$RAF1_{K375M}$ that might be migrate to endosomes to be degraded for example in the lysosome (Lu et al., 2009).

## 3.4 Discussion

In this chapter we have investigated why KRAS protein abundance is mutant-specific. Protein levels in mammalian cells are controlled by multiple molecular events that have an effect on transcription rates, mRNA degradation rates, as well as mRNA translation and protein degradation rates (Schwanhäusser et al., 2011). In this study, we focus mainly on understanding if protein degradation rates are different between KRAS mutants. First, we consistently observe that the oncogenic KRAS G12D mutant has a higher abundance than its wild type form when exogenously expressed in different cell lines using a TetOn3G inducible expression system. This observation is reproduced in mouse intestinal organoids where with use a retroviral infection combined with a Cre-Lox-mediated expression or with expression from the endogenous KRAS locus. Altogether, suggest that the higher protein abundance of the mutant $KRAS_{G12D}$ is not the result of a technical bias linked to the expression system.

$KRAS_{G12D}$ is blocked in an active state, permanently available to interact with its effectors in comparison $KRAS_{WT}$ which is switching between the active and the inactive state. Therefore, we have reasoned that the larger amount of protein-protein interactions for $KRAS_{G12D}$ with the effectors might protect it from degradation, leading to the higher protein abundance observed. In line with our hypothesis, previous work has shown that active proteins and proteins involved in protein complexes are less prone to be degraded and have longer half-lives (Cambridge et al., 2011; Martin-Perez and Villén, 2017). We find that $KRAS_{D38A}$, a mutant hindering the interaction with the effectors, displays a lower protein abundance in comparison to $KRAS_{WT}$ and $KRAS_{G12D}$. Moreover, we find that $KRAS_{G12D}$ and $KRAS_{D38A}$ degradation rates are significantly different suggesting that the interactions with the effectors play an important role in KRAS protein stability.

In support of this, we have found that RAF1 might play an important role in protecting KRAS from degradation in comparison to the other effectors. Interestingly, we have measured that RAF1 has the highest binding affinity among the effectors, which could contribute to a larger amount of KRAS-RAF1 complex formation. On the other hand, when RAF1 is not able to transduce signals (kinase dead), we have observed a decreased of KRAS. This is agreement with the studies showing that active protein complexes are more stable than inactive proteins (Cambridge et al., 2011; Martin-Perez and

Villén, 2017). Indeed, KRAS and RAF1 have been shown to interact and signal from the late endosomes, their inactivation could lead to rapid targeting to the lysosome (Lu et al., 2009). Future experiments will be performed to gain a deeper understanding of this process. First, the determination of the major degradation pathways of KRAS mutants aiming to know whether KRAS mutants are targeted to the lysosomal compartment or to the proteasome by quantifying protein abundance changes upon treatment with proteasomal or lysosomal inhibitors. Second, we will perform localization assays in order to know how the differential binding of KRAS mutants to the effectors might influence their localization within the cell and therefore, their stability.

Next, we have worked with *in silico* designed RAS mutants predicted to have differential affinity for five RAS binding partners. Having a closer look to the crystal structures of RAS in complex with the effectors we realized that a water-mediated network of hydrogen bonds further stabilizes the interaction interface. Therefore, we have implemented a method that combined with FoldX considers the possible effects of mutations on water networks at the interface and in the stability of the complex. We have validated these predictions using two different methods MST and NanoBRET. The first one has allowed us to measure the binding affinities between purified HRAS and effectors proteins. The binding affinities obtained for the HRAS wild type correlate perfectly with the reported affinities found in the literature, suggesting that MST is an accurate method to monitor the interactions of the designed RAS mutants. A possible drawback to this method is the use of the RAS binding domain of each effector instead of full-length effectors, which could impact on the affinities. Moreover, one of the effectors was not assessed (PI3K) due to its resistance and insolubility during protein purification. On the other hand, we have expressed HaloTag-effectors in order to monitor the interactions with NanoLuc-KRAS within HEK293 cells. This method allows to monitor the interactions in a more physiological context with the expression of full-length proteins within living cells. However, the disadvantage is that the tag NanoLuc has a very similar size (19.1 kDa) to KRAS (21 kDa) which might perturb the interactions with the effectors. In addition, the levels of expression might not be physiological leading to unspecific interactions. The interactions measured with MST have a higher overlap with the FoldX predictions (80%) than with NanoBRET (55%). A possible explanation is that both crystal structures and MST measurements

have been done with the RAS binding domains of the effectors instead of full-length effectors. Overall, using complementary techniques can lead to a relatively realistic picture of the biological interactions.

Remarkably, we have observed that the protein abundances of the designed mutants are mutant-specific, supporting our hypothesis that KRAS-effector interactions play a role in KRAS stability. We have not only observed these differences in HEK293 but also in intestinal organoids. Mutant-specific abundances vary between the cell line and the organoids, these could be due to the different cell context (e.g. different levels of binding partners). Interestingly, the cell growth phenotype of the organoids generally agrees with the expression levels of KRAS, highlighting the importance of considering the protein abundance in addition to the protein activity in KRAS mutants related studies.

We have measured the degradation rates of two mutants in HEK293 (Y40C-G12D and Y40F-G12D) showing opposite protein abundance and we have observed corresponding protein degradation rates, i.e. Y40C has a lower protein abundance and a faster degradation than Y40F. These results confirm that mutant-specific protein levels are linked to the process of protein degradation. We will perform the same degradation experiments with the other designed mutants to quantify all mutant-specific degradation rates. In addition, as mentioned above, it would be interesting to use a different method to measure protein degradation to confirm our observations. Actually, continued incubation with cycloheximide might not reflect the actual protein half-live under normal growth conditions due to the cytotoxic effects of prolonged protein synthesis inhibition. Also, the method is not well adapted for long-lived proteins, for instance, if key proteins of the degradation pathway have a shorter half-live than KRAS (~12 hours), there might be a stabilization of targeted proteins for degradation, which further complicates the accurate measurement of protein half-live. Among alternative approaches to monitor protein degradation, pulse-chase analysis involves minimal disruption of normal cell growth. Typically, proteins are labeled with a radioactive precursor and during the chase period an excess of nonradioactive precursors are added to prevent more protein radiolabeling. At different time points cells are lysed and the radiolabeled protein of interest is quantified by immunoblotting. This would allow us to confirm if protein degradation is the

main mechanism leading to mutant-specific protein abundance or whether protein synthesis might have an important contribution as well.

Finally, we have observed that $KRAS_{G12D}$ and $KRAS_{WT}$ protein abundances correlate with the corresponding transcript abundances. Surprisingly, TET protein and transcript levels correlate with KRAS abundance and we hypothesize a possible feedback loop in our TetON3G expression system. It has been shown that the activation of MAPK leads to a stimulation of the transcription controlled by the CMV promoter. This highlights the importance of selecting the appropriate expression vector in comparative gene expressions. To control for this effect, we have exchanged the CMV for an EF1 promoter and we are determining the transcript levels of TET when $KRAS_{WT}$ or $KRAS_{G12D}$ are expressed. Additionally, we will measure $KRAS_{WT}$ and $KRAS_{G12D}$ when express with a non-inducible system and independent of the CMV promoter. Nevertheless, the CMV promoter has not been used for the expression of the mutants in intestinal organoids, supporting that the difference in protein abundances is not a biased observation.

Detailed knowledge about the interaction of the effectors with RAS has allowed us to study how interaction perturbations contribute to the protein abundance of KRAS. Overall, our results suggest that protein degradation plays a major role in the observed differences in mutant protein abundances. Further studies will help to understand if protein synthesis as well as protein localization may contribute to regulate mutant-specific KRAS turnover. Importantly, a better understanding of the factors and mechanisms contributing to the high protein abundances of the oncogenic mutant $KRAS_{G12D}$ KRAS could bring indications for possible therapeutic approaches.

## 3.5 Appendix

### 3.5.1 Expression of KRAS mutants in mouse $APC^{-/-}$ intestinal organoids.

We establish and induce the expression of KRAS mutants in $APC^{-/-}$ intestinal organoids and consistently with previous observations in cell lines, we observe mutant-specific KRAS abundances (Figure 3-20). For instance, E37L is more abundant than E37R as was detected in HEK293. However, I21G is no highly expressed as observed in HEK293. The variations in

abundance between HEK293 and organoids might be due to the effect of the G12D mutation included in the mutants expressed in the cell lines or other physiological effects as for example the cell-type specific expression levels of the effectors. Interestingly, the protein abundance generally agrees with the cellular proliferation of the organoids (Figure 3-21). For example, the number of growing spheres expressing $KRAS_{G12D}$ is more that the double in comparison to WT.

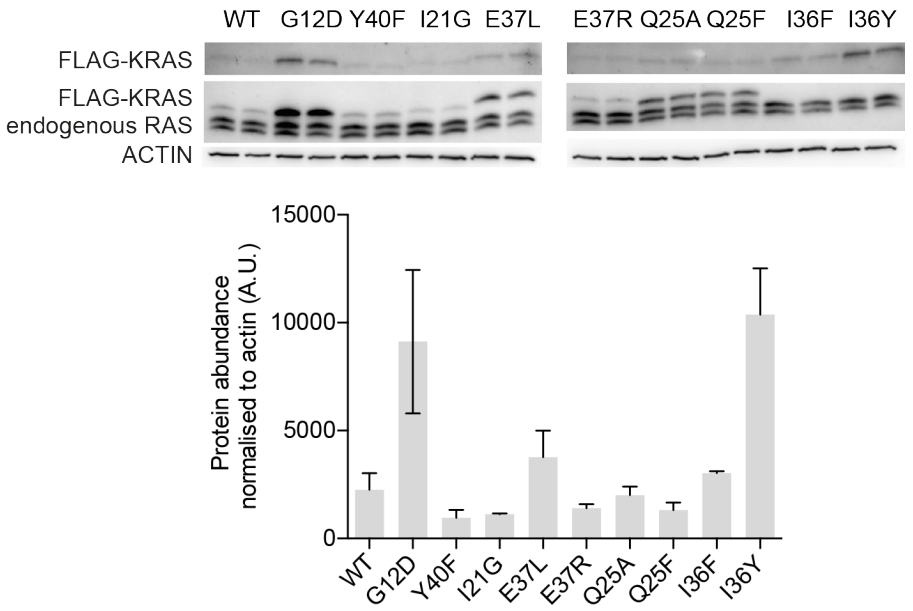



**Figure 3-20 Expression of FLAG-$KRAS_{mutants}$ in $APC^{-/-}$ intestinal orgnanoids.**

*Western blot and protein quantification of $KRAS_{mutants}$ in $APC^{-/-}$ intestinal orgnanoids. Error bars represent the SD of two biological replicates with two technical replicates each.*

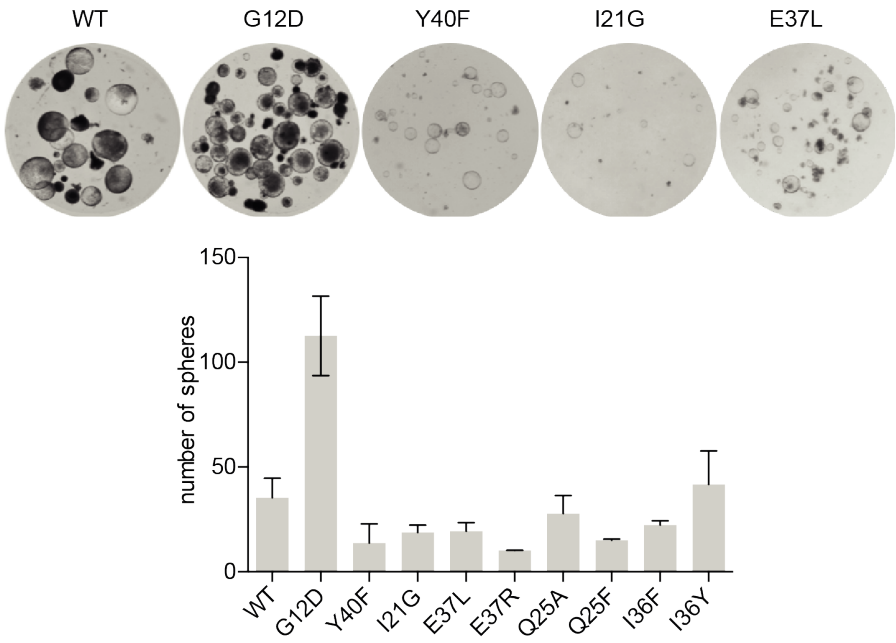



**Figure 3-21 Intestinal organoids sphere formation assay.**

*Example of pictures of intestinal organoids expressing different KRAS mutants taken after 96 hours of growth. The quantification of spheres corresponds as well to intestinal organoids expressing KRAS mutants after 96 hours of growth. Error bars represent two biological replicates with six technical replicates each.*

### 3.5.2 Microscale thermophoresis experiments.

Microscale thermophoresis is an *in vitro* method that allow to measure interactions between purified recombinant proteins. In this case we have measured the interactions between HRAS and the effectors RAF1, RALGDS, RASSF5 and PLCε. In Figure 3-22 we report the binding affinities of HRAS wild type with the effectors and below the relative affinities of HRAS mutants.

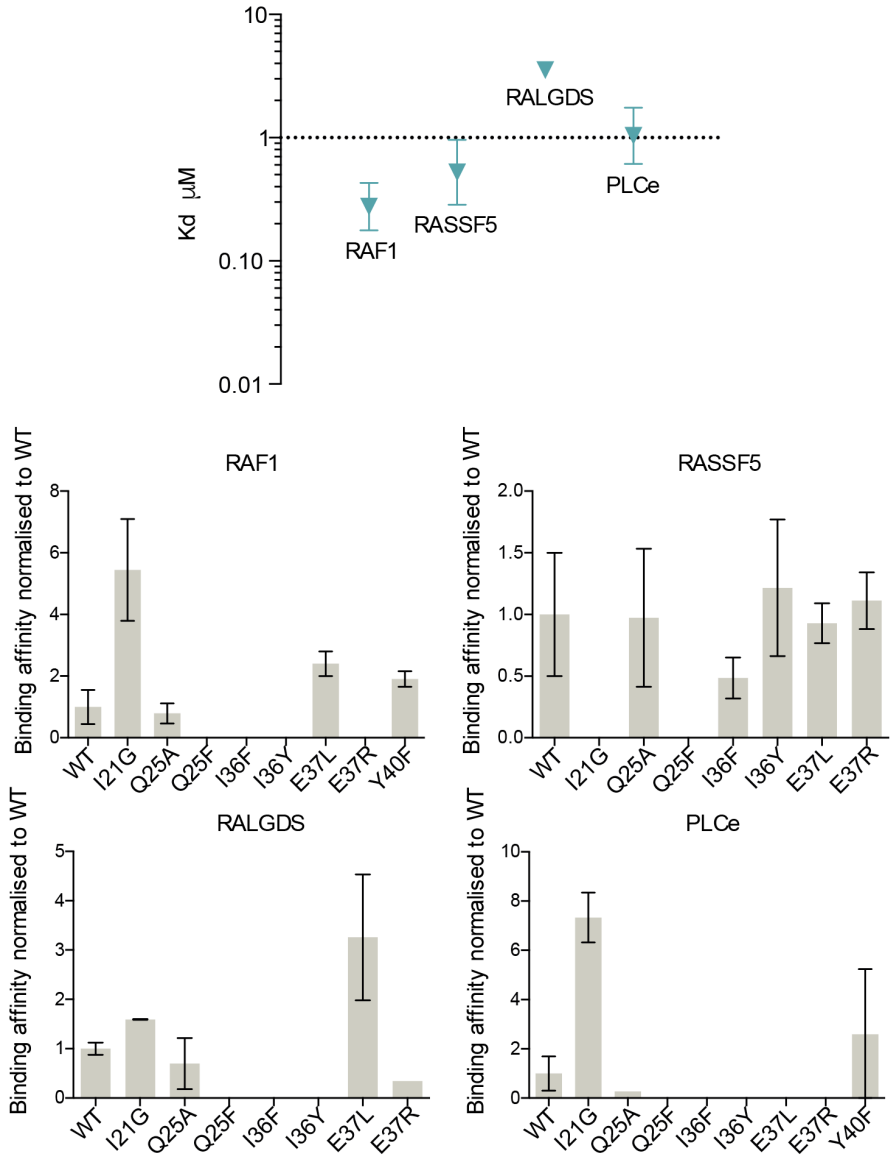



**Figure 3-22 Microscale Thermophoresis measurements between HRAS-GppNHp and four different effectors.**

*The upper panel corresponds to the measurements of affinities between HRAS$_{WT}$ and the RAS binding domain of four different effectors. Below, the relative binding affinities to the WT complex are reported for HRAS mutants. The mutants with no bars are complexes where the measures corresponded to no binding.*

### 3.5.3 Comparison between EF1 and CMV promoters.

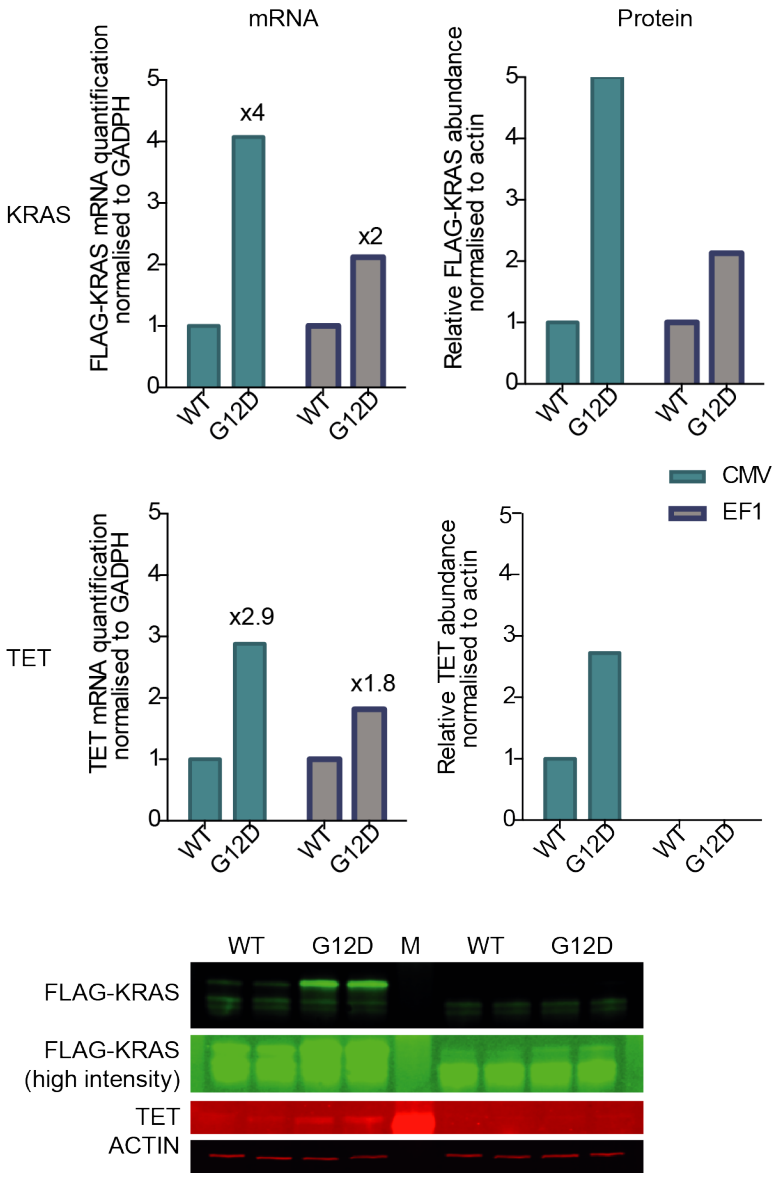



**Figure 3-23 Change from CMV to EF1 promoter**

*mRNA and protein quantification of FLAG-KRAS and TET in cells over-expressing FLAG-KRAS$_{WT}$ and FLAG-KRAS$_{G12D}$. No error bars as no replicates have been performed.*

## 3.6 Methods

### 3.6.1 Structural analysis

**3D Structures**

All the 3D structures of RAS and RAS in complex with different effectors used in this study were crystallographic structures downloaded from the Protein Data Bank.

**FoldX**

FoldX is a sophisticated force field used in this Chapter to predict mutational effects on protein stability (Schymkowitz et al., 2005) Stability change predictions after single mutations were done with structures downloaded from the Protein Data Bank. All structures were repaired before starting the modelling with the command: command=RepairPDB. RepairPDB repairs residues that have bad torsion angles or steric clashes. Mutations were modeled with the command BuildModel, the mutations are specified in an individual and the parameter of number of runs was set to three: file: command=BuildModel; mutant-file=mut_list.txt; numberOfRuns=3. The same commands were used with the graphical interface Yasara (Krieger and Vriend, 2014) including the pluggin of FoldX. FoldX mutates the target residue to itself creating the free energy of the wild type $\Delta G_{WT}$, and then to the indicated residue obtaining the free energy of the mutant $\Delta G_{mutant}$. We report the difference in free energy in kcal.mol$^{-1}$:

$\Delta\Delta G = \Delta G_{mutant} - \Delta G_{WT}$

The standard deviation of the error reported for FoldX (deviation between experimental values and $\Delta\Delta G$ calculated by FoldX) is 0.46 kcal.mol$^{-1}$. The command PSSM was used for the analysis of the effect of mutations between RAS and the effectors. This command uses BuildModel to mutate the residues and AnalyseComplex to calculate the interaction energies. Example: command=Pssm; analyseComplexChains=A,B; pdb=complex.pdb; aminoacids=C,F, positions=YA40a.

**Waters**

In order to compute the water networks involved in the protein folding and affecting the interaction interfaces, we have used the FPocket software (Le Guilloux et al., 2009) originally meant to find “druggable” pockets within a protein structure, in other words regions that are prone to bind drug-like molecules. Briefly, this software computes the alpha spheres within the

protein structure and then assigns a score to a cluster of alpha spheres (the pocket) based on its structural features if they are similar to those pockets already known to bind a drug-like molecule. An alpha sphere is a sphere that contacts four atoms on its boundary and contains no internal atom. By definition the four atoms are at an equal distance (sphere radius) to the alpha sphere center. Each alpha sphere is then assigned as polar, apolar, etc., based on the properties of the atoms involved in its definition. Here, we have modified the input parameters to accept only alpha spheres that are in contact with residues prone to make hydrogen-bonds with waters, alpha spheres with the radius of a water molecule and those that are between them at a distance that can then represent a water-water interaction. Also, we have deleted the restriction that the alpha sphere is not internal in order to model waters trapped within the folded protein. This allowed us to model water molecules and its networks and compute the variation energy with FoldX having these networks into account.

### 3.6.2 Sample preparation and experimental procedures

**Cell lines**

The cell lines included in this study were comprised of HeLa and HEK293. Cells were maintained at 37°C in a humidified atmosphere at 5% $CO_2$ in DMEM 4.5g/L Glucose with UltraGlutamine media supplemented with 10% of Tet-free FBS (Clontech) and 1% penicillin/streptomycin.

**Intestinal organoids**

Intestinal organoids $APC^{-/-}$ and $APC^{-/-}$-$KRAS_{G12D}$ were provided by our collaborators from the laboratory of Prof. Owen Sansom (The Beatson Institute, Glasgow). The organoids were obtained from two different mice that allow for biological replicates ($APC^{-/-}$: from the colony VAR #192309 and #192310; $APC^{-/-}$-$KRAS_{G12D}$: from the colony RBVAPPK #94728 and #100095). Frozen aliquots were defrosted in a 37°C water bath, resuspended in wash media, centrifugated at 600g for 5 minutes, supernatant was aspirated carefully without disturbing the organoids and the pellet was resuspended with 100µL of Matrigel (previously thawed in ice) and 20µL of organoids/Matrigel mix are distribute in a 24-well plate (4-5 wells) pre-heated at 37 °C. The distribution has to be performed very fast in order the avoid the solidification of the Matrigel matrix. The plate was incubated at 37 °C and after 10 minutes the wells were supplemented with 500µL of crypt media. Maintenance and expansion were done in 24-well plates in crypt

media. Passage to 1:5 was done every three days with the following steps: warm up plates in the incubator, disrupt the Matrigel dome with P1000, pipette up/down 10 times with P1000, add the resuspension into a 15mL falcon, supplement with 5mL of wash media, spin at 600g for 5 minutes and aspirate the supernatant, resuspend the pellet with 500µL of wash media, mix up/down 10 times, add 5mL of wash media, centrifuge at 800g for 5 minutes, discard supernatant, resuspend with Matrigel (Matrigel thawed in ice; adapt the volume for a 1:5 passage), distribute 20µL of organoids/Matrigel mix in pre-warmed 24-well plate, after 10 minutes of incubation at 37 °C add 500µL of crypt medium. Alternatively, the commercially available growth media IntestiCult was used instead of in-house prepared crypt medium.
Wash medium: 500mL of Advanced DMEM F12, 5mL of L-Glutamine, 5mL of Pen/Strep, 500µL of N2, 500µL of B27, 4mL of BSA (12.5% BSA).
Crypt medium: Wash media supplemented with EGF and Noggin just before adding medium to well. For each 500µl add: 25µl EGF diluted stock and 10µl Noggin diluted stock.

- EGF stock preparation:

100µg per vial, add 50µl $H_20$ to resuspend powder (do not vortex). Add to 50µl PBS/0.1% BSA. Makes 1µg/µl stock. Dilute 10 µl of the 1µg/µl stock in 10ml of PBS/0.1% BSA. For every 500µl medium add 25µl working stock.

- Noggin stock preparation:

20µg per vial, add 20µl $H_20$ to resuspend powder. Add to 4ml PBS/0.1% BSA. For every 500µl medium add 10µl stock.

| Product | Cat # | Company |
|---|---|---|
| Growth Factor Reduced Matrigel | 356231 | Corning |
| IntestiCult | 06005 | StemCell |
| Advanced DMEM F12 | 12634-010 | ThermoFisher |
| L-Glutamine | 25030-024 | ThermoFisher |
| Pen/Strep | 15140-122 | ThermoFisher |
| EGF | AF-100-15 | Peprotech |
| Noggin | 250-38 | Peprotech |

**Table 3-3 Main components for media preparation for intestinal organoid culture.**

**Expression systems**

The expression of FLAG-KRAS WT and mutants was done using a doxycycline-inducible TetON3G system. The expression vector was

composed of two expression cassettes, one for the expression of the TET transactivator and another containing FLAG-KRAS under the control of a TRE3G promoter (Figure 3-24). The expression in cell lines was induce with 1ng/mL of doxycycline.

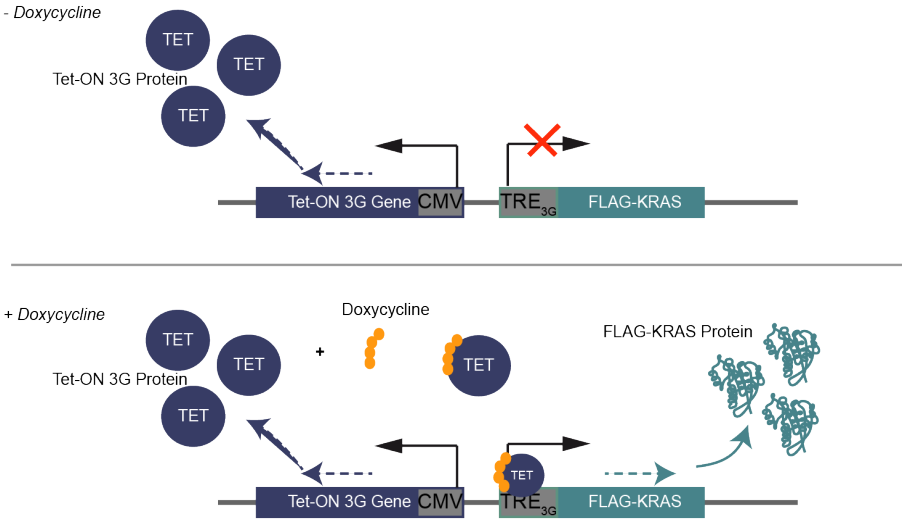



**Figure 3-24 Inducible TetOn3G expression system.**

*The vector used to express FLAG-KRAS contains two expression cassettes. The first one expresses TET under the control of a CMV promoter. The second one expresses FLAG-KRAS when doxycycline has been added to the cells and binds to the transactivator TET.*

The expression system used in intestinal organoids was based on previously published system (Koo et al., 2011). For the conditional overexpression of FLAG-KRAS, a dsRed expression cassette was flanked by loxP inserted upstream of FLAG-KRAS coding sequence (Addgene #32702). Infected organoids were visualized by the red fluorescence, deletion of dsRed by Cre recombinase by the addition of 1µM 4-OHT (Sigma) results in loss of dsRed and expression of FLAG-KRAS.

### Cloning

In order to subclone the different KRAS and effectors coding sequences into the expression vectors, we used the cloning approach known as Gibson assembly (Gibson et al., 2009). This technique combines exonucleases, DNA polymerase, and DNA ligase. The first one hydrolyses the 5'extreme of the DNA and generates single strand 3' ends that have homologies between them.

The polymerase synthesizes the remaining regions after recombination and the ligase binds the extremes to create a complete double-stranded DNA.

**Mutagenesis**

All KRAS mutants were obtained by performing site-directed mutagenesis following the protocol QuickChange site-directed mutagenesis (Agilent).

**Transfection of cell lines**

The transfection of HEK293 and HeLa cell lines for protein expression was done by plating 150000 and 200000 cells per well respectively in a 24-well plate. The cells were transfected after overnight incubation with 500ng of TetON3G-KRAS plasmid (Figure 3-24) following the protocol of Lipofectamine 3000 (ThermoFisher).

**Retroviral infection of intestinal organoids**

To produce the retrovirus containing the expression vector with KRAS gene, the retroviral packaging cell line was used. PLAT-E cells were grown and maintained at 37 °C in a humidified atmosphere at 5% $CO_2$ in DMEM 4.5g/L Glucose with UltraGlutamine media supplemented with 10% of FBS (Gibco) and 1% penicillin/streptomycin. 48 hours before transfection, PLAT-E cells were seeded in a 15cm dish at 60% confluency. Cells were transfected two days after with a mix of 50µg of FLAG-KRAS plasmid, 5mL of DMEM that was pre-incubated during 25 minutes with a mix of 100µL of Lipofectamine 2000 (ThermoFisher) with 5mL of DMEM. After 24 hours of incubation with the mix of DNA and lipofectamine the media was changed, and we could already observe the expression of dsRED if the transfection worked correctly. 48 hours after transfection the media was aspirated with a syringe, filtered with 0.45µm filter and centrifugated during 18 hours at 8000g at 9°C.

Preparing the organoids for retroviral transfection:

Similar to the passage, a total of 6 wells of organoids per mutant were disrupted and washed until obtaining a pellet. The pellet was resuspended with Tryple (ThermoFisher) and incubated during 15 minutes in a 37°C water-bath. Cells were resuspended with 3mL of crypt medium with 5% of FBS and centrifugated 3 minutes at 300g. The supernatant was discarded and the pellet was resuspended in 1mL of infection media. The infection media was prepared with a mix of 70% of crypt medium, 30% PLAT-E cells medium, 10µM Y27632 (Sigma) and 8µg/mL Polybrene (Sigma).

After 18 hours of centrifugation of the virus, the media was carefully discarded (tubes on ice) and 500µL of infection media were added. 250µL were distributed in two 1.5mL Eppendorf tubes and 100µL of disrupted organoids in infection media were added. The mix organoids/virus were transferred to a 24-well plate. The plate was centrifugated during 1 hour at 300g at 32°C and then incubated during 6 hours at 37 °C in a humidified atmosphere at 5% $CO_2$. The cells were collected in 1.5mL Eppendorf tubes and centrifugated at 1000g for 2 minutes, the supernatant discarded and the pellet resuspended in 25µL of Matrigel (thawed in ice) and distributed in two wells of a 24-well plate. After 10 minutes of incubation at 37 °C, 500µL of infection media without polybrene were added. After 48 hours, cells were regaining their characteristic spheroid shape and the media was changed to crypt media. 72 hours after infection, the selection with 2µg/mL puromycin was started.
Expression of FLAG-KRAS was induced with 1µM 4-OHT (Sigma). 48h after induction, 2 wells of organoids were lysed with with 500µL of lysis buffer (M-PER (ThermoFisher), one table of cOmplete-mini (Sigma) and 1x of Halt Phosphatase Inhibitor Cocktail (ThermoFisher) and sonicated for 5 minutes with a cycle of 30 seconds ON and 30 seconds OFF.

Sphere formation assay
Two wells per mutant were induced with 4-OHT. After 48 hours of induction, organoids were collected and centrifuged in order to remove the supernatant and Matrigel. The pellet was resuspended with 200µL of Tryple Express (ThermoFisher) in order to dissociate the cells. Next, cells were counted with the automatic cell counter Countess (Invitrogen). 50000 cells were centrifuged per mutant, resuspended with 150µL and distributed in six wells of a 24-well plate in order to have an average of 8000 cells per well.

**Cycloheximide assay**
Seed 2.5 million cells in a 60mm dish, 24 hours after seeding transfect with a mix of 3000ng of FLAG-KRAS vector, 125µL of Optimem and 5µL of P3000 that was pre-incubated during 15 minutes with a mix of 125µL of Optimem and 5µL of Lipofectamine 3000 (Invitrogen). 6 hours after transfection, cells were trypsinised with 300µL of 0.05% Trypsin-EDTA (Gibco) and counted with the automatic cell counter Countess (Invitrogen). 180.000 cells were seeded per well of a 24-well plate with growth media (see Section Cell lines) supplemented with 0.2ng/mL of Doxycycline (Sigma).

After over-night incubation media was change with fresh media supplemented with 200µg/mL of cycloheximide (Sigma). Cells were lysed with 200µL of lysis buffer (M-PER (ThermoFisher), one table of cOmplete-mini (Sigma) and 1x of Halt Phosphatase Inhibitor Cocktail (ThermoFisher)) at different time points (0, 6, 9 and 12 hours). Then, 50µL of 1x of Laemmli sample buffer (BioRad) with 50µM of DTT were added to each sample. All samples were stored at -20°C and once all samples were collected, they were boiled during 5 minutes at 90°C and stored at -20°C.

**NanoBRET**

Two million of HEK293 cells were seeded in 60mm dish, transfected 24 hours after with a mix of 6000ng of HaloTag-effector vector, 220µL of Optimem and 11.5µL of P3000 that was pre-incubated during 15 minutes with a mix of 220µL of Optimem and 11.5µL of Lipofectamine 3000 (ThermoFisher). 24 hours after transfection cell were trypsinised with 0.05% Trypsin-EDTA (Gibco) and counted with the automatic cell counter Countess (Invitrogen). 400.000 cells were seeded per well of a 24-well plate with growth media (see Section Cell lines) supplemented with 5ng/mL of Doxycycline (Sigma). Six hours after seeding, cells were transfected with a mix of 5ng of NanoLuc-KRAS vector, 11.5µL of Optimem and 1µL of P3000 that was pre-incubated during 15 minutes with a mix of 25µL of Optimem and 1µL of Lipofectamine 3000 (ThermoFisher). After over-night incubation, for each well of the 24-well plate, medium was aspirated and cells washed with 250µL of 1xPBS, then 125µL of 0.05% Trypsin-EDTA (Gibco) were added to each well and incubated at room temperature until cells detach. Cells were resuspended with 500µL per well of medium and the content of each well was transferred to an Eppendorf tube. The tubes were centrifugated at 125g for 5 minutes, the medium was aspirated and the pellet resuspended with 100µL of Optimem supplemented with 4% of FBS. Resuspended cells were counted and diluted to 400.000 cells per milliliter. 25µL of the diluted cells were added to the wells of a 96-well plate (4 wells per sample). The HaloTag NanoBRET 618 ligand (Promega, stock at 0.1mM) was diluted to 100nM by adding 1µL of the ligand to 1000µL of Optimem supplemented with 4% FBS. Next, 25µL of this solution were added to each well to have a total of 10.000 cells per well and 50nM HaloTag NanoBRET 618 ligand (Promega). In parallel, a 0.1% DMSO solution was prepared by adding 1µL of DMSO to 1000µL of Optimem supplemented with 4% FBS and 25µL of this solution were added to each well to have a total of 10.000 cells per well

and 0.05% DMSO. Then, cells were incubated at 37°C in a humidified atmosphere at 5% $CO_2$ for six hours. NanoBRET Nano-Glo substrate (Promega) was prepared to 2.5x in Optimem. This is a 200-fold dilution of the stock reagent. The solution was used within 2 hours of preparation if stored at room temperature and 4 hours if stored at 4°C. 12.5µL of substrate (0.5x) were added per well and the plate was shacked for 30 seconds. The plate was read in a microplate reader (TECAN, Infinite 200) measuring the donor emission (460nm) and the acceptor emission (618nM) within 10 minutes of substrate addition. The corrected NanoBRET ratio was calculated:
NanoBRET (mBU) = Ligand ($\frac{618nm}{460nm}$) – No-ligand DMSO ($\frac{618nm}{460nm}$)

**Microscale Thermophoresis (MST)**

The RAS binding of the effectors (RAF1: aa 51-131, RALGDS: aa 777-872, PLCε: 2130-2240, RASSF5: aa 200-358) and HRAS were purified as described previously by the HTS facility from the CRG (Nakhaei-Rad et al., 2016). The HRAS protein was put through an enzymatic nucleotide exchange to GppNHp. Effectors were fluorescently labeled according to the manufacturer's protocol (amine reactive labeling kit, Nanotemper). HRAS wild type and mutants were serially diluted 1:1 v/v with labeled effector. Measurements were performed in MST buffer. The mixtures were transferred to hydrophilic-treated capillaries. Measurements of the *in vitro* affinities between HRAS and the effectors were done with the instrument Monolith NT.115 from Nanotemper Technologies. Thermophoresis readings were carried out with 40% LED power and 40% laser power. Curve fitting and calculations of binding affinities were done using the Nanotemper analyses software of the Thermophoresis T-jump mode.

**mRNA quantification**

RNA isolation was performed with RNeasy kit (Qiagen). $KRAS_{WT}$ and $KRAS_{HRAS}$ transcript abundances were quantified by RT-qPCR (Power SYBR Green RNA-to-CT 1-Step Kit, ThermoFisher). The Ct values for KRAS or TET were normalized to GADPH, as for example:
$\Delta Ct = (Ct_{KRAS} - Ct_{GADPH})$ and represented as $2^{-\Delta Ct}$.

| Primer | Sequence |
|---|---|
| FLAG-KRAS forward | 5'-CAAGGACGACGATGACAAG-3' |
| FLAG-KRAS reverse | 5'-GGGTCGTATTCGTCCACAA-3' |
| GADPH forward | 5'-GAGTCAACGGATTTGGTCGT-3' |
| GADPH reverse | 5'-TTGATTTTGGAGGGATCTCG-3' |
| TET forward | 5'-ACTCTGCTCTGGAATTACTC-3' |
| TET reverse | 5'-GTTTCGTACTGTTTCTCTGTTG-3' |

**Table 3-4 Primers for RT-qPCR**

**Quantitative fluorescence-based protein blots**

Cells were lysed using M-PER buffer (ThermoFisher) supplemented with anti-proteases. Protein concentration was measured using a BCA Protein Assay Kit (Pierce). Equal amounts of each sample were mixed with 1x Laemmli buffer and boiled for 5 min. Samples were separated using 12% polyacrylamide gels (BioRad). Transfer was performed using the iBlot system (Invitrogen). Membranes were treated with Li-COR Odyssey blocking buffer for 1 hour at RT, then incubated with primary antibody (1:1000) in 0.2% Tween-20/Li-COR odyssey blocking buffer overnight at 4°C. Following three 5 min washes in TBS-T, the membrane was incubated with secondary antibodies (1:10000) in 0.2% Tween-20/Li-COR Odyssey blocking buffer for 45 min at RT. Following three 5 min washes in TBS-T, the membrane was scanned using the Li-COR Odyssey Imaging System. We used the following primary antibodies: anti-pan-RAS (Abcam, ab52939) and anti-β-actin (Sigma, A2228) and were detected using a goat anti-rabbit (Abcam, ab216773) or goat anti-mouse (Abcam, ab216776) IgG antibody conjugated to an IRdye at 800CW and 680CW, respectively. Visualization and quantification were done using ImageJ and Image Studio Lite (LI-COR).

**Quantitative chemiluminescence-based protein blots**

Cells were lysed using M-PER buffer (ThermoFisher) supplemented with anti-proteases. Protein concentration was measured using a BCA Protein Assay Kit (Pierce). Equal amounts of each sample were mixed with 1x Laemmli buffer and boiled for 5 min. Samples were separated using 12% polyacrylamide gels (BioRad). Transfer was performed using the iBlot system (Invitrogen). Membranes were treated with TBS, Tween 0.1%, 5% milk for 1 hour at RT, then incubated with primary antibody (1:1000) in TBS, Tween 0.1%, 0.5% milk overnight at 4°C. Following three 5 min washes in

TBS, Tween 0.1%, the membrane was incubated with secondary antibodies (1:10000) in TBS, Tween 0.1%, 0.5% milk for 45 min at RT. Following three 5 min washes in TBS, Tween 0.1%, the membrane was developed using high sensitivity ECL reagent (ThermoFisher) and visualized using the Fujifilm LAS-3000 developer. Quantification was done using ImageJ.

**Total protein quantification**

Total protein stain of western blot membranes was done using REVERT Total Protein Stain (Licor).

# Chapter 4

# 4 Discussion

## 4.1 Overview

One characteristic common to all cells is the dynamic ability to coordinate their activities with environment changes. This is achieved through a number of pathways that receive and process signals originating from external stimuli. Intracellular signaling usually comprises many components acting sequentially where one component passes the signal to the next component. The major signal transducers are receptors, signaling enzymes and regulatory GTPases like RAS proteins. The malfunction of signaling pathways can contribute to tumorigenesis in several ways. For example, increased RAS signaling enhances cell proliferation. For the oncogene RAS there are two ways of oncogenic activation. On the one hand, the structure of the protein may be affected; this is the case for the oncogenic mutations affecting Glycine 12, which causes a structural perturbation that renders RAS insensitive to the action of GTPase-activating proteins (GAPs). This increases the time in which RAS resides in the active state and leads to prolonged signaling to the downstream effectors. On the other hand, higher RAS signaling might be achieved by increase of RAS abundance. A change in gene expression or higher stability of the oncogene product might lead to a higher cellular concentration of the protein. Due to the increase in RAS concentration downstream signaling might be amplified.

During the past years the studies about RAS biology have focused on the molecular mechanisms of effector activation, their specific biological output, their relative contribution to oncogenic transformation, as well as understanding RAS signal transduction from specialized areas of the plasma membrane and intracellular membranes (Omerovic et al., 2007). Yet, fewer studies can be found about the variation of RAS protein levels and its protein degradation. In this thesis we have focused on the determinants of RAS protein abundance and their possible contribution to RAS oncogenicity.

In this dissertation we find:

- KRAS has a codon usage characteristic of proliferation-related genes. In addition, we find the same pattern in other oncogenes (BRAF, RAC1, RHOA, and COL11A1).
- The proliferative state of the cell influences the protein abundance of RAS genes, with a higher increase of KRAS in comparison to HRAS.
- The tRNA repertoire of HEK293 and HeLa is different and corresponds to the relative protein abundance of KRAS.
- The protein abundance of oncogenic $KRAS_{G12D}$ is higher than KRAS wild type, while the inactive mutant D38A lower.
- Differential perturbation of KRAS protein interactions affect the protein abundance of KRAS.
- Differences in degradation rate could explain part of the differences observed in mutant-specific protein abundance.
- RAF1 plays an important role in regulating the protein abundance of KRAS.

## 4.2 Quantification of oncogene abundances

Mainly, our findings lead to the observation that genes selected to be more mutated in cancer in comparison to their family members correspond to genes with a codon usage characteristic of genes involved in proliferation. We choose to experimentally study this observation is the RAS family. We observe specifically a codon-mediated increase of protein levels of KRAS when cells are proliferating. However, a limitation of our study is the missing experiments with the remaining genes BRAF, RAC1, RHOA, and COL11A1 to consistently confirm our hypothesis. Moreover, our experimental set up lacks a patho-physiological context as our experiments have been performed in standard cell lines. As a future direction it would be interesting to measure RAS protein abundance differences in tumor and healthy samples. We would expect a change in the ratio KRAS/HRAS between both states related to the proliferation rate. Nevertheless, as mentioned in the introduction, the measure of endogenous RAS protein abundances can be challenging for quantitative proteomics due to their high sequence similarity which makes the three RAS proteins difficult to distinguish. This would be the case as well for the other families with a high amino acid identity (e.g. RAC and RHO). A recently published mass spectrometry methodology based on protein standard

absolute quantification (Mageean et al., 2015) could be applied in order to overcome this challenge. This technique employs isotope-labeled full-length protein standards that are spiked into the initial cell lysate. Therefore, this approach would allow to accurately measure the endogenous ratio of KRAS/HRAS and other protein families and to compare the changes of this ratio between different cellular contexts. In addition, antibodies specific for each of the three RAS proteins have been developed, allowing the relatively quantification of RAS proteins by immunoblotting (Waters et al., 2017).
Our comparative study of properties of oncogenes in gene families hints the importance of codon usage. It would therefore be interesting to assess the codon usage of all known oncogenes (and not just those with sequence similar family members). We would hypothesized that the most frequently mutated oncogenes will have a codon usage characteristic for proliferation-related genes, as we have already observed for the family members studied in Chapter 2. Doing so would better our understanding of which factors make an oncogene, might thereby help to identify novel cancer genes and would improve our understanding of the interplay between translational control and proliferation during tumorigenesis.

## 4.3 Codon usage and tRNA

Even though RAS genes have been widely studied for the last 40 years, the surprising observation that they differ in their codon usage was just recently reported (Lampson et al., 2013). Previous studies and our work show that KRAS gene is enriched in so-called rare codons (Lampson et al., 2013; Pershing et al., 2015). The distinction of rare versus common codons in bacteria is typically based on the rarity or commonality of their cognate tRNAs, which is estimated by the copy number of the tRNA genes in the genome, or related to the codon usage of highly expressed genes. Thus, optimal codons are likely under selection in highly expressed genes, in particular in bacteria where optimal expression of those genes may lead to a direct fitness advantage, by increasing biomass production and growth rate. However, in humans, codons have been categorized as rare or common based on how often each codon occurs in mRNA coding sequences at the gene level, without considering mRNA and protein levels or cell type-specific differences. Moreover, the functional relevance of codon usage might be more complex in human cells than in bacteria. Different cell types, development stages, or changing environments, make it impossible to have one global and static set of optimal codons. In addition, the copy number of

tRNA genes is not necessarily correlated with tRNA levels, limiting the usefulness of this type of abundance estimate. It has been shown that codons corresponding to highly abundant tRNAs are translated rapidly (Hanson and Coller, 2017). However, optimal tRNAs seem to vary depending on the cellular context as we and others have reported (Dittmar et al., 2006; Gingold et al., 2014). These observations call into question the definition of consistently rare codons in the context of human genes. Instead, the description of optimal or non-optimal codons depends on the balance between the concentration of tRNAs in the cell (tRNA supply) and the frequency of codon usage in the cellular transcripts (tRNA demand related to RNA levels and codons in those RNAs). This balance might reflect a mechanism that ensures the proper protein expression in the corresponding tissue or cellular status. A challenging goal will be to define which codons are limiting or optimal for translation in distinct cellular contexts. Having access to precise quantification of tRNA pools from multiple cell types, conditions, diseases, developmental stages and organisms will certainly contribute to a better understanding of cell regulation associated to translation.

In recent years, novel methods are allowing to improve the quantification of the dynamic expression and modification patterns of tRNAs (Orioli, 2017), most of them based on next-generation sequencing. Different technologies help to quantify the levels of tRNA regulation, such as changes in transcription, amino acid loading, tRNA cleavage and tRNA modifications. A first challenge is to measure their abundance. Until recently, custom-made microarrays have been typically used to quantify tRNA expression (Dittmar et al., 2006; Gingold et al., 2014). The limitation of this method is the requirement of designing probes based on a known tRNA repertoire, as well as the high sequence similarity might lead to cross-hybridization. Until recently, most high throughput sequencing technologies were hampered by the large secondary structure and abundant modifications of tRNAs that obstruct cDNA synthesis required for the preparation of RNA sequencing libraries. A major recent improvement has been the use of engineered demethylases that allow to remove the most abundant modifications typically causing cDNA synthesis to stop, combined with a highly processive reverse transcriptase that can operate on RNA with stable secondary structure and read-through modifications (Zheng et al., 2015). The method we employ in Chapter 2 to sequence tRNAs overcomes as well some of the difficulties

encountered in tRNA sequencing. The protocol for hydro-tRNAseq includes a size-selection of 60 to 100 nucleotide-long RNA molecules and a partial hydrolyzation of tRNAs which generates smaller fragments, less prone to harbor complex secondary structures, therefore facilitating deep sequencing (Gogakos et al., 2017). Nevertheless, the method is not optimal as fragments containing modifications can interfere with reverse transcriptase and thus, they might be overlooked in the quantification. A complementary but indirect method to study tRNA regulation can be to consider the status of chromatin epigenetics or RNA polymerase III occupancy around tRNA genes. In addition, tRNAs decode codons when they are charged with an amino acid, the variation of charged tRNAs in short timescales (minutes to hours) can thus play an important role in the dynamic regulation of translation. A recent method allows to determine the charged fraction of tRNAs under different cellular conditions (Evans et al., 2017). The combination of complementary omics technologies will bring insight in tRNA biology and help to understand for example, tRNA-related mechanisms involved in diseases.

### 4.3.1 Effect of tRNA pool on protein levels

Recent research has made clear that the coordination of an optimal codon usage for a specific tRNA pool ensures that the same gene reaches proper protein abundances in different cell states, cell types, and developmental stages (Bornelöv et al., 2019; Dittmar et al., 2006; Rudolph et al., 2016) For instance, it has been shown that transcripts involved in proliferation or differentiation have different codon usage and we can hypothesize that this is a regulatory mechanism that allows to control ideal translation in a given cellular state. The ability to translate a transcript that match the current translation program of the cell is increased whereas those that are not needed will probably not match the tRNA pool.

An interesting example of the coordination between tRNA supply and demand is the case of viral protein expression as viruses depend on the translational machinery of the infected cell. The coding sequences of the virus are found in some cases to be adapted to the tRNA availability of a specific cell type or organism that the virus infects (Zhou et al., 1999).

Disturbing the balance between the tRNA supply and demand has been shown to have patho-physiological consequences. For example, the upregulation of two specific tRNAs (tRNAGluUUC and tRNAArgCCG) promotes metastatic progression by enhancing translation of transcripts enriched with the matching codons (Goodarzi et al., 2016). The intriguing

question that remains open is why these two tRNAs are increased in metastatic cells and which is the physiological function of the codon-matching transcripts in healthy cells. For example, we could hypothesize that the upregulation of these two tRNAs is characteristic of a cell migration-related translation program occurring during embryonic development that is reused by the metastatic cancer cells. In this thesis we support these observations by reporting the proliferation-specific codon usage of proto-oncogenes which suggest that cancer cells might be using a proliferation-specific translational control program to their own growth advantage. As discussed in Chapter 2, one feature of these oncogenes, aside from their specific codon usage, is that the absence of expression is embryonically lethal. This highlights the importance of proper protein levels during embryogenesis. We speculate that proliferation-related tRNAs are expressed during embryogenesis in order to produce at necessary levels proteins such as KRAS, BRAF, RAC1, RHOA and COL11A1. Indeed, it has been discussed that cancer mimics and involves a set of canonical cellular processes used during embryo development, such as the regulation of differentiation, proliferation, and cell migration (Aiello and Stanger, 2016; Cofre and Abdelhay, 2017). Therefore, we hypothesize that different programs of translation coordinated during development are appropriated and reactivated by cancer cells to drive tumorigenesis.

Even though KRAS is found mutated in 80% of tumors, there are some cancer types where NRAS or HRAS mutants are predominantly found. For example, there is near-exclusive mutation of KRAS in pancreatic and colorectal cancers, while HRAS mutations are predominant in head and neck squamous cell carcinoma (Hobbs et al., 2016). Due to high codon bias between KRAS and HRAS and based on our findings, it would be interesting to test whether the tRNA pools differ between these cancer types and how they match the respective codon usage of KRAS and HRAS. It will be possible to answer this question thanks to an ongoing effort in our group (Hernandez-Alias et al, submitted) to develop a method that allows to quantify tRNA expression of cancer samples from TCGA using small RNA sequencing data. This will allow us to accurately quantify tRNA abundances in different cancer types and investigate cancer type-specific variation in translational control.

### 4.3.2 Effect of tRNA pool in mRNA stability

In recent years, the protein translation rate has been linked to mRNA stability. Genes using optimal codons give rise to more stable mRNA and higher translation efficiency. This has been shown in yeast as well as in human cells where stable mRNAs appear to be enriched with optimal codons (Presnyak et al., 2015; Wu et al., 2019). In Chapter 2, we have also observed that protein ratios correlate with mRNA ratios, and this effect disappear when mRNA translation is inhibited (removing ATG and the ribosome binding site). The correlation might indicate that slow translation causes mRNA degradation, indeed an mRNA decay mechanism has been described in yeast with the Dhh1p protein targeting slow-translated transcripts for degradation (Radhakrishnan et al., 2016). Therefore, the link between tRNA supply and tRNA demand appears to govern both translation rate and mRNA stability. An interesting example of translation-dependent mRNA stability is the regulation of embryonic mRNA which is critical for development and is highly orchestrated during maternal-to-zygotic transition (Bazzini et al., 2016; Mishima and Tomari, 2016). It has been shown that the stability of maternal mRNAs is tightly controlled. There are pools of mRNAs cleared at different time points during embryogenesis. Codon optimality appears to play a major role in this cellular transition, as non-optimal transcripts are prone to be degraded, whereas transcripts with codons corresponding to abundant tRNAs are stabilized and actively translated (Bazzini et al., 2016).

## 4.4 RAS genes and levels

In vivo studies have shown that the same mutation causes different biological outcomes depending on the RAS gene harboring the mutation (Haigis et al., 2008; Wang et al., 2013). However, there is still little understanding of the factors leading to the RAS gene-specific preferences associated with oncogenic mutations. Previously, it has been discussed that one of the factors shaping the ideal level of RAS signaling is the abundance of oncogenic RAS (Li et al., 2018). In this thesis, we show that codon bias in the RAS family and its effect on protein abundance, might be one of the possible factors contributing to KRAS mutation selection over HRAS and NRAS.

Our results lead to speculate that the expression of KRAS favors cell transformation (i.e. proliferation) that will in turn increase the levels of KRAS in a codon-mediated fashion, suggesting a positive feedback regulation. On the other hand, HRAS expression favors cell transformation but here there will be no positive feedback as HRAS codon usage is less adapted to

translation in a proliferative cell state. Therefore, we could rather consider this a negative feedback loop. Ideally, we could think about an experimental set up where we can quantify cell proliferation over time with the overexpression of $KRAS_{WT}$ and $KRAS_{HRAS}$ to understand to which extend RAS codon bias contributes to a cell proliferative phenotype.
In addition, we could use our experimental design to investigate the role of C-terminal post-translational modifications specific to KRAS or HRAS on protein abundance. One of the most discriminatory features between the two proteins is a distinctive subcellular localization pattern, which is thought to give KRAS an increased capacity to engage downstream effector pathways. Related to our findings in Chapter 3, we could hypothesize that this increased interaction with the effectors will protect KRAS from degradation. For example, we could compare the protein abundance of $KRAS_{HRAS}$ and HRAS, that are only distinct in the C-terminal hypervariable region that harbors post-translational modifications. This would allow to identify the possible role of modifications on RAS protein abundance, independently of the effect of the codon usage.
Overall, more insight on the modulators of RAS abundance and the variations across cancer types could provide help to unravel the complex mutation profiles of RAS.

## 4.5 Interaction with effectors

RAS effectors have a common subdomain in their otherwise unrelated protein body, for their interaction with RAS. Biochemical and structural analysis have revealed a highly similar mode of interaction but with two orders of magnitude differences in the binding affinities (e.g. among all effectors RAF1 shows the highest affinity to RAS on the nanomolar range). Multiple partners compete for binding making RAS a central hub protein. The differences in their protein abundances (Kiel et al., 2013) and subcellular localization contributes to the specificity of the binding and thus to the diversity of RAS signaling. Any process changing the balance and relative stoichiometry of RAS and the effectors proteins can modulate signaling.
Here, we have focused on the effect of these interactions on the degradation of KRAS. Surprisingly, in comparison to the extensive literature on RAS biology, we find few reports about the molecular mechanisms controlling RAS degradation. The two protein degradation pathways, the proteasome and the lysosome, have been linked to RAS degradation. One study reports that the degradation of KRAS occurs in the lysosome (Lu et al., 2009) in contrast

to HRAS which has been shown to undergo proteasomal degradation (Kim et al., 2009). However, more comparative studies between KRAS, HRAS and NRAS are needed to understand which are the main differences related to their degradation. In Chapter 3 we measure KRAS protein degradation overtime, but we have not yet assessed the role of the lysosome and proteasome by using specific inhibitors of each pathway. Interestingly, it has been shown that in addition to the plasma membrane signal output, late endosomes seem to play an important role as KRAS signaling platforms (Lu et al., 2009). Specifically, a sustained RAF1/MAPK signaling activated by KRAS has been detected in this compartment. After signal transduction KRAS is eventually targeted and degraded in the lysosome. In our study we have identified RAF1 as a major player protecting KRAS from degradation. We could imagine KRAS interacting with RAF1 in the endosomes leads to sustained signaling, whereas when KRAS is not able to activate RAF1/MAPK, it would be rapidly targeted for degradation in the lysosome. In addition, KRAS stability has been shown to be controlled by one subunit of the ubiquitin ligase complex, the β-TrCP1 protein (Shukla et al., 2014). Interestingly, β-TrCP1 has been previously shown to be activated by the RAS/RASSF5 signaling pathway (Schmidt et al., 2014), suggesting a possible role of the interaction with the effector RASSF5 on KRAS stability. Our work suggests that the degradation of KRAS modulated by its protein interactions might contribute to the potential complexity of RAS signaling. One interesting aspect of RAS biology are the small biochemical differences between the distinct oncogenic mutations. As presented in the introduction, the spectra of KRAS mutations varies widely in cancer, for instance the main mutation in pancreatic cancer is G12D but G12C in lung cancer (Prior et al., 2012). Recent studies suggest that different KRAS mutants engage distinct sets of effectors (Hunter et al., 2015; Riquelme et al., 2015). For example, *in vitro* binding experiments show that mutant G13D has a higher affinity than mutant G12D for RAF1. In line with these observations we hypothesize that KRAS oncogenic mutants might display distinct protein degradation rates depending on their specific binding to the effectors that in turn might contribute to different levels of activity in cells. Thus, it would be interesting to investigate whether oncogenic mutant-specific KRAS protein levels matters and if it could be partly responsible for the variation of KRAS mutation patterns in human cancers.

Regarding the studies about RAS oncogenicity, the main focus has been in the past on the activation of downstream pathways, and protein levels of mutated RAS are rarely quantified. Here, we have shown that protein levels are distinct among mutants. Therefore, we highlight the importance of considering the quantification of protein abundance and not only the quantification of the signal transduction activity.

## 4.6 KRAS turnover

In relation to our findings in Chapter 2, we have not explored the possibility that the designed KRAS mutations might lead to different levels of cell proliferation which in turn could affect the expression of KRAS in a codon-dependent manner. In this sense, mutation-specific protein synthesis could contribute as well to the observed abundance variation.
In addition, further investigation is required to elucidate the role of mRNA levels on KRAS protein synthesis. We are currently in the process of assessing whether the correlations that we observe between mutant-specific mRNA and protein levels are physiological or if they result from a technical bias of our expression system. Intriguingly, mRNA stability has been shown to be the best predictor of protein half-lives in yeast (Martin-Perez and Villén, 2017), however the molecular mechanism remains elusive. We wonder if this co-regulation between protein turnover and mRNA turnover would be also observed in higher eukaryotes. This observation may be in line with our results (it still needs to be confirmed with a different expression system) where mRNA abundance is higher for KRAS mutants with a lower degradation rate. We could speculate about a feedback mechanism between protein degradation and mRNA turnover.

## 4.7 Concluding remarks

The multitude of factors influencing whether a particular RAS gene with a particular mutation contributes to oncogenicity, makes the understanding of RAS mutation patterns extremely complicated. Here, I have studied two specific aspects of RAS biology leading to differences in RAS protein levels, mRNA translation and protein degradation.
The mechanisms underlying the establishment of RAS protein levels, could be applied to other scenarios concerning different oncogenes. Indeed, in Chapter 2 we have found that a proliferation-related codon usage is not only a feature of KRAS but also of other very well studied cancer genes such as BRAF. In addition, our results in Chapter 3 we find that protein-protein

interactions are a role in mutant-specific protein stability. This observation may be extrapolated to other signaling hubs mutated in disease.
Therefore, the study of modulators of RAS protein levels and other important signaling proteins will need to be continued as the question remains whether isoform- or mutation-specific protein abundance variation are important determinant to understand mutation patterns in cancer.

# Contributions

**Published:**

Systems level expression correlation of Ras GTPase regulators.
Besray Unal E*, Kiel C*, Benisty H, Campbell A, Pickering K, Blüthgen N, Sansom OJ, Serrano L.
Cell Commun Signal. 2018 doi: 10.1186/s12964-018-0256-8.

Interaction Dynamics Determine Signaling and Output Pathway Responses.
Stojanovski K*, Ferrar T*, Benisty H*, Uschner F, Delgado J, Jimenez J, Solé C, de Nadal E, Klipp E, Posas F, Serrano L, Kiel C.
Cell Rep. 2017 doi: 10.1016/j.celrep.2017.03.029.

The yin-yang of kinase activation and unfolding explains the peculiarity of Val600 in the activation segment of BRAF.
Kiel C, Benisty H, Lloréns-Rico V, Serrano L.
Elife. 2016 doi: 10.7554/eLife.12814.

**Under revision:**

Proliferation specific codon usage facilitates oncogene translation.
Benisty H, Weber M, Hernandez-Alias X, Schaefer M.H, Serrano L.
bioRxiv 2019 doi: 10.1101/695957
Under revision in Nature Communications

**Submitted:**

Translational efficiency across healthy and tumor tissues is proliferation-related.
Hernandez-Alias X, Benisty H, Schaefer M.H, Serrano L.

**In preparation:**

Cancer type-specificity of KRAS mutations is explained by tissue-specific brake silencing.
Schaefer MH, Beltran-Sastre V, Benisty H, Head S, Serrano L.

# References

Agris, P.F., Vendeix, F.A.P., and Graham, W.D. (2007). tRNA's Wobble Decoding of the Genome: 40 Years of Modification. J. Mol. Biol. *366*, 1–13.

Ahearn, I., Zhou, M., and Philips, M.R. (2018). Posttranslational Modifications of RAS Proteins. Cold Spring Harb. Perspect. Med. *8*, 1–22.

Ahearn, I.M., Tsai, F.D., Court, H., Zhou, M., Jennings, B.C., Ahmed, M., Fehrenbacher, N., Linder, M.E., and Philips, M.R. (2011). FKBP12 binds to acylated H-ras and promotes depalmitoylation. Mol. Cell *41*, 173–185.

Ahmadian, M.R., Stege, P., Scheffzek, K., and Wittinghofer, A. (1997). Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. Nat. Struct. Biol. *4*, 686–689.

Aiello, N.M., and Stanger, B.Z. (2016). Echoes of the embryo: using the developmental biology toolkit to study cancer. Dis. Model. Mech. *9*, 105–114.

Alvarez-Moya, B., López-Alcalá, C., Drosten, M., Bachs, O., and Agell, N. (2010). K-Ras4B phosphorylation at Ser181 is inhibited by calmodulin and modulates K-Ras activity and function. Oncogene *29*, 5911.

Apolloni, A., Prior, I.A., Lindsay, M., Parton, R.G., and Hancock, J.F. (2002). H-ras but Not K-ras Traffics to the Plasma Membrane through the Exocytic Pathway. Mol. Cell. Biol. *20*, 2475–2487.

Barbieri, C.E., Baca, S.C., Lawrence, M.S., Demichelis, F., Blattner, M., Theurillat, J.P., White, T.A., Stojanov, P., Van Allen, E., Stransky, N., et al. (2012). Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat. Genet. *44*, 685–689.

Bazzini, A.A., Viso, F., Moreno-Mateos, M.A., Johnstone, T.G., Vejnar, C.E., Qin, Y., Yao, J., Khokha, M.K., and Giraldez, A.J. (2016). Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition . EMBO J. *35*, 2087–2103.

Beltran-Sastre, V., Benisty, H., Burnier, J., Berger, I., Serrano, L., and Kiel, C. (2015). Tuneable endogenous mammalian target complementation via multiplexed plasmid-based recombineering. Sci. Rep. *5*, 17432.

Bodemann, B.O., and White, M.A. (2013). Ras GTPases: codon bias holds KRas down but not out. Curr. Biol. *23*, R17--20.

Boël, G., Letso, R., Neely, H., Price, W.N., Wong, K.H., Su, M., Luff, J.D., Valecha, M., Everett, J.K., Acton, T.B., et al. (2016). Codon influence on protein expression in E. coli correlates with mRNA levels. Nature *529*, 358–363.

Bornelöv, S., Selmi, T., Flad, S., Dietmann, S., and Frye, M. (2019). Codon usage optimization in pluripotent embryonic stem cells. Genome Biol. *20*, 1–16.

Bunney, T.D., and Katan, M. (2006). Phospholipase C epsilon: linking second messengers and small GTPases. Trends Cell Biol. *16*, 640–648.

Burd, C.E., Liu, W., Huynh, M. V., Waqas, M. a., Gillahan, J.E., Clark, K.S., Fu, K., Martin, B.L., Jeck, W.R., Souroullas, G.P., et al. (2014). Mutation-Specific RAS Oncogenicity Explains NRAS Codon 61 Selection in Melanoma. Cancer Discov. *4*, 1418–1429.

Cambridge, S.B., Gnad, F., Nguyen, C., Bermejo, J.L., Krüger, M., and Mann, M. (2011). Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover. J. Proteome Res. *10*, 5275–5284.

Castellano, E., and Downward, J. (2011). RAS Interaction with PI3K: More Than Just Another Effector Pathway. Genes Cancer *2*, 261–274.

Chan, P.P., and Lowe, T.M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res. *44*, D184–D189.

Chiu, V.K., Bivona, T., Hach, A., Sajous, J.B., Silletti, J., Wiener, H., Johnson, R.L., Cox, A.D., and Philips, M.R. (2002). Ras signalling on the endoplasmic reticulum and the Golgi. Nat. Cell Biol. *4*, 343–350.

Cofre, J., and Abdelhay, E. (2017). Cancer Is to Embryology as Mutation Is to Genetics: Hypothesis of the Cancer as Embryological Phenomenon. Sci. World J. *2017*.

Consortium, Gte. (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

Corbetta, S., Gualdoni, S., Albertinazzi, C., Paris, S., Croci, L., Consalez, G.G., and de Curtis, I. (2005). Generation and Characterization of Rac3 Knockout Mice. Mol. Cell. Biol. *25*, 5763–5776.

Cox, A.D., Der, C.J., Philips, M.R., and Carolina, N. (2015). Targeting RAS Membrane Association : Back to the Future for. Clin. Cancer Res. *21*, 1–17.

Crick, F. (1970). Central Dogma of Molecular Biology. Nature *227*, 561–563.

Desideri, E., Cavallo, A.L., and Baccarini, M. (2015). Alike but Different: RAF Paralogs and Their Signaling Outputs. Cell *161*, 967–970.

Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. PLoS Genet. *2*, 2107–2115.

Drosten, M., Simón-Carrasco, L., Hernández-Porras, I., Lechuga, C.G., Blasco, M.T., Jacob, H.K., Fabbiano, S., Potenza, N., Bustelo, X.R., Guerra, C., et al. (2017). H-Ras and K-Ras oncoproteins induce different tumor spectra when driven by the same regulatory sequences. Cancer Res. 707–718.

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M.D., Zimonjic, D.B., Donaher, J.L., Popescu, N.C., Hahn, W.C., and Weinberg, R.A. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev. *15*, 50–65.

Emerson, S.D., Madison, V.S., Palermo, R.E., Waugh, D.S., Scheffler, J.E., Tsao, K.-L., Kiefer, S.E., Liu, S.P., and Fry, D.C. (1995). Solution Structure of the Ras-Binding Domain of c-Raf-1 and Identification of Its Ras Interaction Surface. Biochemistry *34*, 6911–6918.

Esteban, L.M., Vicario-Abejon, C., Fernandez-Salguero, P., Fernandez-Medarde, A., Swaminathan, N., Yienger, K., Lopez, E., Malumbres, M., McKay, R., Ward, J.M., et al. (2001). Targeted Genomic Disruption of H-ras and N-ras, Individually or in Combination, Reveals the Dispensability of Both Loci for Mouse Growth and Development. Mol. Cell. Biol. *21*, 1444–1452.

Evans, M.E., Clark, W.C., Zheng, G., and Pan, T. (2017). Determination of tRNA aminoacylation levels by high-throughput sequencing. Nucleic Acids Res. *45*, e133.

Fearon, E.R., and Volgelstein, B. (1990). A genetic model for colorectal cancer. Cell *61*, 759–767.

Feig, L.A., and Buchsbaum, R.J. (2002). Cell signaling: Life or death decisions of Ras proteins. Curr. Biol. *12*, 259–261.

Ferbeyre, G. (2007). Barriers to Ras transformation. Nat. Cell Biol. *9*, 483–485.

Fersht, A.R., and Serrano, L. (1993). Principles of protein stability derived from protein engineering experiments. Curr. Opin. Struct. Biol. *3*, 75–83.

Fiorucci, G., and Hall, A. (1988). All three human ras genes are expressed in a wide range of tissues. BBA - Gene Struct. Expr. *950*, 81–83.

Fu, J., Dang, Y., Counter, C., and Liu, Y. (2018). Codon usage regulates human KRAS expression at both transcriptional and translational levels. J. Biol. Chem. jbc.RA118.004908.

Ganesan, A.K., Vincent, T.S., Olson, J.C., and Barbieri, J.T. (1999). Pseudomonas aeruginosa exoenzyme S disrupts Ras-mediated signal transduction by inhibiting guanine nucleotide exchange factor-catalyzed nucleotide exchange. J. Biol. Chem. *274*, 21823–21829.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison III, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343.

Gingold, H., Dahan, O., and Pilpel, Y. (2012). Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. Nucleic Acids Res. *40*, 10053–10063.

Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A Dual Program for Translation Regulation in Cellular Proliferation and Differentiation. Cell *158*, 1281–1292.

Gogakos, T., Brown, M., Garzia, A., Meyer, C., and Hafner, M. (2017). Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro- Resource Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. CellReports *20*, 1463–1475.

Goodarzi, H., Nguyen, H.C.B., Zhang, S., Dill, B.D., Molina, H., and Tavazoie, S.F. (2016). Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression Article Modulated Expression of Specific tRNAs Drives Gene Expression and Cancer Progression. Cell *165*, 1416–1427.

Gremer, L., Merbitz-Zahradnik, T., Dvorsky, R., Cirstea, I.C., Kratz, C.P., Zenker, M., Wittinghofer, A., and Ahmadian, M.R. (2011). Germline KRAS mutations cause aberrant biochemical and physical properties leading to developmental disorders. Hum. Mutat. *32*, 33–43.

Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009). Fpocket: An open source platform for ligand pocket detection. BMC Bioinformatics *10*, 168.

Haigis, K.M., Kendall, K.R., Wang, Y., Cheung, A., Haigis, M.C., Glickman, J.N., Niwa-Kawakita, M., Sweet-Cordero, A., Sebolt-Leopold, J., Shannon, K.M., et al. (2008). Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. Nat. Genet. *40*, 600–608.

Hakem, A., You-Ten, A., Duncan, G., Wakeham, A., Mak, T.W., Sanchez-Sweatman, O., and Khokha, R. (2005). RhoC is dispensable for embryogenesis and tumor initiation but essential for metastasis. Genes Dev. *19*, 1974–1979.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation. Cell *144*, 646–674.

Hanahan, D., Weinberg, R.A., and Francisco, S. (2000). The Hallmarks of Cancer. *100*, 57–70.

Hancock, J.F., Paterson, H., and Marshall, C.J. (1990). A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21ras to the plasma membrane. Cell *63*, 133–139.

Hanson, G., and Coller, J. (2017). Codon optimality, bias and usage in translation and mRNA decay. Nat. Publ. Gr.

Herrero, A., Matallanas, D., and Kolch, W. (2016). The spatiotemporal regulation of RAS signalling. Biochem. Soc. Trans. *44*, 1517–1522.

Herrmann, C., Martin, G.A., and Wittinghofer, A. (1995). Quantitative Analysis of the Complex p21ras and the Ras-binding Domain of the Human Raf-1 Protein Kinase. J. Biol. Chem. *270*, 2901–2905.

Hobbs, G.A., Der, C.J., and Rossman, K.L. (2016). RAS isoforms and mutations in cancer at a glance. J. Cell Sci. 1287–1292.

Hunter, J.C., Manandhar, A., Carrasco, M. a, Gurbani, D., Gondi, S., and Westover, K.D. (2015). Biochemical and Structural Analysis of

Common Cancer-Associated KRAS Mutations. Mol. Cancer Res. 1325–1336.

Ihle, N.T., Byers, L. a., Kim, E.S., Saintigny, P., Lee, J.J., Blumenschein, G.R., Tsao, A., Liu, S., Larsen, J.E., Wang, J., et al. (2012). Effect of KRAS oncogene substitutions on protein behavior: Implications for signaling and clinical outcome. J. Natl. Cancer Inst. *104*, 228–239.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. *2*, 13–34.

Jeong, W., Yoon, J., Park, J., Lee, S., Lee, S., Kaduwal, S., Kim, H., Yoon, J., and Choi, K. (2012). Ras Stabilization Through Aberrant Activation of Wnt / b -Catenin Signaling Promotes Intestinal Tumorigenesis. *5*, 1–14.

Jindal, G. a, Goyal, Y., Burdine, R.D., Rauen, K. a, and Shvartsman, S.Y. (2015). RASopathies: unraveling mechanisms with animal models. Dis. Model. Mech. *8*, 769–782.

Johnson, L., Greenbaum, D., Cichowski, K., Mercer, K., Murphy, E., Schmitt, E., Bronson, R.T., Umanoff, H., Edelmann, W., Kucherlapati, R., et al. (1997). K-ras is an essential gene in the mouse with partial functional overlap with N-ras. Genes Dev. *11*, 2468–2481.

Jura, N., Scotto-Lavino, E., Sobczyk, A., and Bar-Sagi, D. (2006). Differential modification of Ras proteins by ubiquitination. Mol. Cell *21*, 679–687.

Khosravi-far, R., White, M.A., Westwick, J.K., Solski, P.A., Chrzanowska-wodnicka, M., Aelst, L.V.A.N., Wigler, M.H., and Der, C.J. (1996). Oncogenic Ras Activation of Raf / Mitogen-Activated Protein Kinase-Independent Pathways Is Sufficient To Cause Tumorigenic Transformation. *16*, 3923–3933.

Kiel, C., Serrano, L., and Herrmann, C. (2004). A detailed thermodynamic analysis of Ras/effector complex interfaces. J. Mol. Biol. *340*, 1039–1058.

Kiel, C., Wohlgemuth, S., Rousseau, F., Schymkowitz, J., Ferkinghoff-Borg, J., Wittinghofer, F., and Serrano, L. (2005). Recognizing and defining true ras binding domains II: In Silico prediction based on homology modelling and energy calculations. J. Mol. Biol. *348*, 759–775.

Kiel, C., Verschueren, E., Yang, J.-S., and Serrano, L. (2013). Integration of Protein Abundance and Structure Data Reveals Competition in the ErbB Signaling Network. Sci. Signal. *6*, ra109–ra109.

Kim, S.E., Yoon, J.Y., Jeong, W.J., Jeon, S.H., Park, Y., Yoon, J.B., Park, Y.N., Kim, H., and Choi, K.Y. (2009). H-Ras is degraded by Wnt/β-catenin signaling via β-TrCP-mediated polyubiquitylation. J. Cell Sci. *122*, 842–848.

Koera, K., Nakamura, K., Nakao, K., Miyoshi, J., Toyoshima, K., Hatta, T., Otani, H., Aiba, A., and Katsuki, M. (1997). K-Ras is essential for the development of the mouse embryo. Oncogene *15*, 1151–1159.

Koo, B.-K., Stange, D.E., Sato, T., Karthaus, W., Farin, H.F., Huch, M., van Es, J.H., and Clevers, H. (2011). Controlled gene expression in primary Lgr5 organoid cultures. Nat. Methods *9*, 81–83.

Krengel, U., Scherer, A., Schumann, R., Frech, M., John, J., Kabsch, W., Pai, E.F., and Wittinghofer, A. (1990). Three-Dimensional Structures of H-ras p21 Mutants : Molecular Basis for Their Inability to Function As Signal Switch Molecules. Cell *62*, 539–548.

Krieger, E., and Vriend, G. (2014). YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics *30*, 2981–2982.

Lage, K., Hansena, N.T., Karlberg, E.O., Eklund, A.C., Roque, F.S., Donahoe, P.K., Szallasi, Z., Jensen, T.S., and Brunak, S. (2008). A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. Proc. Natl. Acad. Sci. U. S. A. *105*, 20870–20875.

Lahtvee, P.J., Sánchez, B.J., Smialowska, A., Kasvandik, S., Elsemman, I.E., Gatto, F., and Nielsen, J. (2017). Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. Cell Syst. *4*, 495-504.e5.

Lampson, B.L., Pershing, N.L.K., Prinz, J.A., Lacsina, J.R., Marzluff, W.F., Nicchitta, C. V, MacAlpine, D.M., and Counter, C.M. (2013). Rare codons regulate KRas oncogenesis. Curr. Biol. *23*, 70–75.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501.

Ledford, H. (2015). Cancer: The Ras renaissance. Nature *520*, 278–280.

Li, S., Balmain, A., and Counter, C.M. (2018). A model for RAS mutation patterns in cancers: finding the sweet spot. Nat. Rev. Cancer *18*, 767–777.

Li, Y., Lacerda, D.A., Warman, M.L., Beier, D.R., Yoshioka, H., Ninomiya, Y., Oxford, J.T., Morris, N.P., Andrikopoulos, K., Ramirez, F., et al. (1995). A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis. Cell *80*, 423–430.

Liu, A.-X., Rane, N., Liu, J.-P., and Prendergast, G.C. (2001). RhoB Is Dispensable for Mouse Development, but It Modifies Susceptibility to Tumor Formation as Well as Cell Adhesion and Growth Factor Signaling in Transformed Cells. Mol. Cell. Biol. *21*, 6906–6912.

Lu, A., Tebar, F., Alvarez-Moya, B., López-Alcalaacute, C., Calvo, M., Enrich, C., Agell, N., Nakamura, T., Matsuda, M., and Bachs, O. (2009). A clathrin-dependent pathway leads to KRas signaling on late endosomes en route to lysosomes. J. Cell Biol. *184*, 863–879.

Machleidt, T., Woodroofe, C.C., Schwinn, M.K., Méndez, J., Robers, M.B., Zimmerman, K., Otto, P., Daniels, D.L., Kirkland, T.A., and Wood, K. V. (2015). NanoBRET-A Novel BRET Platform for the Analysis of Protein-Protein Interactions. ACS Chem. Biol. *10*, 1797–1804.

Mageean, C.J., Griffiths, J.R., Smith, D.L., Clague, M.J., and Prior, I.A. (2015). Absolute Quantification of Endogenous Ras Isoform Abundance. PLoS One *10*, e0142674.

Malumbres, M., and Barbacid, M. (2003). RAS oncogenes: the first 30 years. Nat. Rev. Cancer *3*, 459–465.

Martin-Perez, M., and Villén, J. (2017). Determinants and Regulation of Protein Turnover in Yeast. Cell Syst. *5*, 283-294.e5.

Mishima, Y., and Tomari, Y. (2016). Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. Mol. Cell *61*, 874–885.

Nakhaei-Rad, S., Nakhaeizadeh, H., Götze, S., Kordes, C., Sawitza, I., Hoffmann, M.J., Franke, M., Schulz, W.A., Scheller, J., Piekorz, R.P., et al. (2016). The Role of Embryonic Stem Cell-expressed RAS (ERAS) in the Maintenance of Quiescent Hepatic Stellate Cells. J. Biol. Chem. *291*, 8399–8413.

Nakhaeizadeh, H., Amin, E., Nakhaei-Rad, S., Dvorsky, R., and Ahmadian, M.R. (2016). The RAS-Effector Interface: Isoform-Specific Differences in the Effector Binding Regions. PLoS One *11*, e0167145.

Neel, N.F., Martin, T.D., Stratford, J.K., Zand, T.P., Reiner, D.J., and Der, C.J. (2011). The RalGEF-Ral Effector Signaling Network: The Road Less Traveled for Anti-Ras Drug Discovery. Genes Cancer *2*, 275–287.

Newlaczyl, A.U., Coulson, J.M., Prior, I.A., Stephen, A.G., Esposito, D., Bagni, R.K., McCormick, F., Aoki, Y., Niihori, T., Inoue, S., et al. (2017). Quantification of spatiotemporal patterns of Ras isoform expression during development. Sci. Rep. *7*, 41297.

Newman, Z.R., Young, J.M., Ingolia, N.T., and Barton, G.M. (2016). Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. Proc. Natl. Acad. Sci. U. S. A. *113*, E1362-71.

Novoa, E.M., and Ribas de Pouplana, L. (2012). Speeding with control: Codon usage, tRNAs, and ribosomes. Trends Genet. *28*, 574–581.

Nowell, P.C. (1976). The Clonal Evolution of Tumor Cell Populations. Science (80-. ). *194*, 23–28.

Omerovic, J., Laude, A.J., and Prior, I.A. (2007). Ras proteins: Paradigms for compartmentalised and isoform-specific signalling. Cell. Mol. Life Sci. *64*, 2575–2589.

Omerovic, J., Hammond, D.E., Clague, M.J., and Prior, I.A. (2008). Ras isoform abundance and signalling in human cancer cell lines. Oncogene *27*, 2754–2762.

Orioli, A. (2017). tRNA biology in the omics era: Stress signalling dynamics and cancer progression. BioEssays *39*, 1–11.

Pacold, M.E., Suire, S., Perisic, O., Lara-Gonzalez, S., Davis, C.T., Walker, E.H., Hawkins, P.T., Stephens, L., Eccleston, J.F., and Williams, R.L. (2004). Crystal Structure and Functional Analysis of Ras Binding to Its Effector Phosphoinositide 3-Kinase γ. Cell *103*, 931–944.

Pavon-Eternod, M., Gomes, S., Geslain, R., Dai, Q., Rosner, M.R., and Pan, T. (2009). tRNA over-expression in breast cancer and functional consequences. Nucleic Acids Res. *37*, 7268–7280.

Pershing, N.L.K., Lampson, B.L., Belsky, J. a, Kaltenbrun, E., Macalpine, D.M., and Counter, C.M. (2015). Rare codons capacitate Kras -driven de novo tumorigenesis. *125*.

Potenza, N., Vecchione, C., Notte, A., De Rienzo, A., Rosica, A., Bauer, L., Affuso, A., De Felice, M., Russo, T., Poulet, R., et al. (2005). Replacement of K-Ras with H-Ras supports normal embryonic development despite inducing cardiovascular pathology in adult mice. EMBO Rep. *6*, 432–437.

Presnyak, V., Alhusaini, N., Chen, Y.H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon optimality is a major determinant of mRNA stability. Cell *160*, 1111–1124.

Prior, I.A., Muncke, C., Parton, R.G., and Hancock, J.F. (2003). Direct visualization of ras proteins in spatially distinct cell surface microdomains. J. Cell Biol. *160*, 165–170.

Prior, I.A., Lewis, P.D., and Mattos, C. (2012). A comprehensive survey of ras mutations in cancer. Cancer Res. *72*, 2457–2467.

Quinlan, M.P., and Settleman, J. (2009). Isoform-specific ras functions in development and cancer. Futur. Oncol. *5*, 105–116.

Radhakrishnan, A., Chen, Y.H., Martin, S., Alhusaini, N., Green, R., and Coller, J. (2016). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. Cell *167*, 122-132.e9.

Rak, R., Dahan, O., and Pilpel, Y. (2018). Repertoires of tRNAs: The Couplers of Genomics and Proteomics. Annu. Rev. Cell Dev. Biol. *34*, 239–264.

Ramocki, M.B., White, M.A., Konieczny, S.F., and Taparowsky, E.J. (1998). A Role for RalGDS and a Novel Ras Effector in the Ras-mediated Inhibition of Skeletal Myogenesis *. *273*, 17696–17701.

Reinstein, E., and Ciechanover, A. (2006). Narrative Review: Protein Degradation and Human Diseases: The Ubiquitin ConnectionProtein Degradation and Human Diseases: the Ubiquitin Connection. Ann. Intern. Med. *145*, 676–684.

dos Reis, M., Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: A test for translational selection. Nucleic Acids Res. *32*, 5036–5044.

Riquelme, E., Behrens, C., Lin, H.Y., Simon, G.R., Papadimitrakopoulou, V.A., Izzo, J., Moran, C., Kalhor, N., Lee, J.J., Minna, J.D., et al. (2015). Modulation of EZH2 expression by MEK-ERK or PI3K-AKT signaling in lung cancer is dictated by different KRAS oncogene mutations. Cancer Res. 1–12.

Roberts, A.W., Chaekyun, K., Zhen, L., Lowe, J.B., Kapur, R., Petryniak, B., Spaetti, A., Pollock, J.D., Borneo, J.B., Bradford, G.B., et al. (1999). Deficiency of the hematopoietic cell-specific Rho family GTPase Rac2 is characterized by abnormalities in neutrophil function and host defense. Immunity *10*, 183–196.

Rodova, M., Jayini, R., Singasani, R., Chipps, E., and Islam, M.R. (2013). CMV promoter is repressed by p53 and activated by JNK pathway. Plasmid *69*, 223–230.

Rodriguez-Viciana, P., Warne, P.H., Khwaja, A., Marte, B.M., Pappin, D., Das, P., Waterfield, M.D., Ridley, A., and Downward, J. (1997). Role of phosphoinositide 3-OH kinase in cell transformation and control of the actin cytoskeleton by Ras. Cell *89*, 457–467.

Roux, P.P., and Topisirovic, I. (2012). Regulation of mRNA translation by signaling pathways. Cold Spring Harb. Perspect. Biol. *4*, 1–25.

Rudolph, K.L.M., Schmitt, B.M., Villar, D., White, R.J., Marioni, J.C., Kutter, C., and Odom, D.T. (2016). Codon-Driven Translational Efficiency Is Stable across Diverse Mammalian Cell States. PLoS Genet. *12*, 1–23.

Santos, E., and Nebreda, R. (1989). Properties of Ras Proteins. Faseb 2151–2163.

Sarkisian, C.J., Keister, B.A., Stairs, D.B., Boxer, R.B., Moody, S.E., and Chodosh, L.A. (2007). Dose-dependent oncogene-induced senescence in vivo and its evasion during mammary tumorigenesis. Nat Cell Biol *9*, 493–505.

Sasaki, A.T., Carracedo, A., Locasale, J.W., Anastasiou, D., Takeuchi, K., Kahoud, E.R., Haviv, S., Asara, J.M., Pandolfi, P.P., and Cantley, L.C. (2011). Ubiquitination of K-Ras enhances activation and facilitates binding to select downstream effectors. Sci. Signal. *4*, ra13.

Schmick, M., Vartak, N., Papke, B., Kovacevic, M., Truxius, D.C., Rossmannek, L., and Bastiaens, P.I.H. (2014). KRas localizes to the

plasma membrane by spatial cycles of solubilization, trapping and vesicular transport. Cell *157*, 459–471.

Schmidt, M.L., Donninger, H., and Clark, G.J. (2014). Ras regulates SCFb-TrCP protein activity and specificity via its effector protein NORE1A. J. Biol. Chem. *289*, 31102–31110.

Schmitt, B.M., Rudolph, K.L.M., Karagianni, P., Schmitt, B.M., Rudolph, K.L.M., Karagianni, P., Fonseca, N.A., White, R.J., Talianidis, I., Odom, D.T., et al. (2014). High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA − tRNA interface High-resolution mapping of transcriptional dynamics across tissue development reveals a stable mRNA – tRNA interface. 1797–1807.

Schoenheimer, R. (1942). The Dynamic State of Body Constituents. Harvard Univ. Press 78.

Schroeder, M.P., Rubio-Perez, C., Tamborero, D., Gonzalez-Perez, A., and Lopez-Bigas, N. (2014). OncodriveROLE classifies cancer driver genes in loss of function and activating mode of action. Bioinformatics *30*, 549–555.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: An online force field. Nucleic Acids Res. *33*, 382–388.

Seeburg, P.H., Colby, W.W., Capon, D.J., Goeddel, D. V, and Levinson, a D. (1984). Biological properties of human c-Ha-ras1 genes mutated at codon 12. Nature *312*, 71–75.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16(INK4a). Cell *88*, 593–602.

Shukla, S., Allam, U.S., Ahsan, A., Chen, G., Krishnamurthy, P.M., Marsh, K., Rumschlag, M., Shankar, S., Whitehead, C., Schipper, M., et al. (2014). KRAS Protein Stability Is Regulated through SMURF2: UBCH5 Complex-Mediated β-TrCP1 Degradation. Neoplasia *16*, 115-IN5.

Simanshu, D.K., Nissley, D. V, and Mccormick, F. (2017). RAS Proteins and Their Regulators in Human Disease. Cell *170*, 17–33.

Smith, M.J., Neel, B.G., and Ikura, M. (2013). NMR-based functional profiling of RASopathies and oncogenic RAS mutations. Proc. Natl. Acad. Sci. *110*, 4574–4579.

Sobhani, N., Ianza, A., D'Angelo, A., Roviello, G., Giudici, F., Bortul, M., Zanconati, F., Bottin, C., and Generali, D. (2018). Current Status of Fibroblast Growth Factor Receptor-Targeted Therapies in Breast Cancer. Cells *7*, 76.

Sriskanthadevan-Pirahas, S., Deshpande, R., Lee, B., and Grewal, S.S. (2018). Ras/ERK-signalling promotes tRNA synthesis and growth via the RNA polymerase III repressor Maf1 in Drosophila. PLOS Genet. *14*, e1007202.

Stephens, R.M., Yi, M., Kessing, B., Nissley, D. V, and McCormick, F. (2017). Tumor RAS Gene Expression Levels Are Influenced by the Mutational Status of RAS Genes and Both Upstream and Downstream RAS Pathway Genes. Cancer Inform. *16*, 1176935117711944–1176935117711944.

Stieglitz, B., Bee, C., Schwarz, D., Yildiz, Ö., Moshnikova, A., Khokhlatchev, A., and Herrmann, C. (2008). Novel type of Ras effector interaction established between tumour suppressor NORE1A and Ras switch II. EMBO J. *27*, 1995–2005.

Sugihara, K., Nakamura, K., Nakao, K., Aiba, A., Katsuki, M., Nozawa, S., Nakatsuji, N., Hashimoto, R., Otani, H., Sakagami, H., et al. (1998). Rac1 is required for the formation of three germ layers during gastrulation. Oncogene *17*, 3427–3433.

Supek, F. (2016). The Code of Silence: Widespread Associations Between Synonymous Codon Biases and Gene Function. J. Mol. Evol. *82*, 65–73.

Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T., and Lehner, B. (2014). Synonymous mutations frequently act as driver mutations in human cancers. Cell *156*, 1324–1335.

Torrent, M., Chalancon, G., Groot, N.S. De, Wuster, A., and Babu, M.M. (2018). Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. Sci. Signal. *11*, eaat6409.

Torres, A.G., Batlle, E., and Ribas de Pouplana, L. (2014). Role of tRNA modifications in human diseases. Trends Mol. Med. *20*, 306–314.

Tsai, F.D., Lopes, M.S., Zhou, M., Court, H., Ponce, O., Fiordalisi, J.J., Gierut, J.J., Cox, A.D., Haigis, K.M., and Philips, M.R. (2015). K-Ras4A splice variant is widely expressed in cancer and uses a hybrid membrane-targeting motif. Proc. Natl. Acad. Sci. *112*, 779–784.

Umanoff, H., Edelmann, W., Pellicert, A., and Kucherlapati, R. (1995). The murine N-ras gene is not essential for growth and development (gene targeting/embryonic stem cells/oncogenes). Genetics *92*, 1709–1713.

Vigil, D., Cherfils, J., Rossman, K.L., and Der, C.J. (2010). Ras superfamily GEFs and GAPs: validated and tractable targets for cancer therapy? Nat. Rev. Cancer *10*, 842–857.

de Vos, A.M., Tong, L., Milburn, M. V, Matias, P.M., Jancarik, J., Noguchi, S., Nishimura, S., Miura, K., Ohtsuka, E., and Kim, S.H. (1988). Three-dimensional structure of an oncogene protein: catalytic domain of human c-H-ras p21. Science (80-. ). *239*, 888 LP – 893.

Wang, Y., Velho, S., Vakiani, E., Peng, S., Bass, A.J., Chu, G.C., Gierut, J., Bugni, J.M., Der, C.J., Philips, M., et al. (2013). Mutant N-RAS protects colorectal cancer cells from stress-induced apoptosis and contributes to cancer development and progression. Cancer Discov. *3*, 294–307.

Waters, A.M., Ozkan-dagliyan, I., Vaseva, A. V, Fer, N., Strathern, L.A., Hobbs, G.A., Tessier-cloutier, B., Gillette, W.K., Bagni, R., Whiteley, G.R., et al. (2017). Evaluation of the selectivity and sensitivity of isoform- and mutation-specific RAS antibodies. *3332*.

Weinberg, R. (2006). The Biology of Cancer (Garland Science).

Weinberg, R.A. (2014). Coming full circle - From endless complexity to simplicity and back again. Cell *157*, 267–271.

Wennerberg, K. (2005). The Ras superfamily at a glance. J. Cell Sci. *118*, 843–846.

Wenstrup, R.J., Florer, J.B., Brunskill, E.W., Bell, S.M., Chervoneva, I., and Birk, D.E. (2004). Type V collagen controls the initiation of collagen fibril assembly. J. Biol. Chem. *279*, 53331–53337.

White, R.J. (2004). RNA polymerase III transcription and cancer. Oncogene *23*, 3208–3216.

White, M. a, Nicolette, C., Minden, a, Polverino, a, Van Aelst, L., Karin, M., and Wigler, M.H. (1995). Multiple Ras functions can contribute to mammalian cell transformation. Cell *80*, 533–541.

Wienken, C.J., Baaske, P., Rothbauer, U., Braun, D., and Duhr, S. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. Nat. Commun. *1*, 100.

Wittinghofer, A., and Herrmann, C. (1995). Ras-effector interactions, the problem of specificity. FEBS Lett. *369*, 52–56.

Wittinghofer, A., and Vetter, I.R. (2011). Structure-Function Relationships of the G Domain, a Canonical Switch Motif. Annu. Rev. Biochem. *80*, 943–971.

Wojnowski, L., Zimmer, A.M., Beck, T.W., Hahn, H., Bernal, R., Rapp, U.R., and Zimmer, A. (1997). Endothelial apoptosis in Braf-deficient mice. Nat. Genet. *16*, 293–297.

Wojnowski, L., Stancato, L.F., Zimmer, A.M., Hahn, H., Beck, T.W., Larner, A.C., R. Rapp, U., and Zimmer, A. (1998). Craf-1 protein kinase is essential for mouse development. Mech. Dev. *76*, 141–149.

Wu, Q., Medina, S.G., Kushawah, G., DeVore, M.L., Castellano, L.A., Hand, J.M., Wright, M., and Bazzini, A.A. (2019). Translation affects mRNA stability in a codon-dependent manner in human cells. Elife *8*, 1–22.

Xu, H., and Ren, D. (2015). Lysosomal Physiology. Annu. Rev. Physiol. *77*, 57–80.

Yan, J., Roy, S., Apolloni, A., Lane, A., and Hancock, J.F. (1998). Ras Isoforms Vary in Their Ability to Activate Raf-1 and Phosphoinositide 3-Kinase. J. Biol. Chem. *273*, 24052–24056.

Yang, M.H., Nickerson, S., Kim, E.T., Liot, C., Laurent, G., Spang, R., Philips, M.R., Shan, Y., Shaw, D.E., Bar-Sagi, D., et al. (2012). Regulation of RAS oncogenicity by acetylation. Proc. Natl. Acad. Sci. *109*, 10843 LP – 10848.

Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M., and Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. Nat. Methods *12*, 835–837.

Zhou, J., Liu, W.J., Peng, S.W., Sun, X.Y., and Frazer, I. (1999). Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. J. Virol. *73*, 4972–4982.