Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

Torsten Schmenger, M.Sc.

born in: Lampertheim

Oral examination: 08.12.2022

Exploration of Natural Variants in Human Health and Disease

Referees:

Prof. Dr. Robert Bruce Russell

Prof. Dr. Britta Brügger

Acknowledgements

First and foremost, I want to express my deepest gratitude to my thesis supervisor Prof. Dr. Robert Russell for his comprehensive support and encouragement to explore science freely, for his generosity and superb ideas and in particular for his guidance.

I want to sincerely thank Prof. Dr. Britta Brügger for examination and evaluation of my thesis as well as for being a member of my PhD Thesis Advisor Committee. I am very grateful to Prof. Dr. Anne-Claude Gavin for also being a member of my Thesis Advisor Committee.

My thanks go to Juan Carlos González Sánchez for his scientific ideas and support, and for sharing his knowledge about population genetics and pop culture. Furthermore, I would like to thank Dr. Gurdeep Singh for his interest and suggestions. I am grateful for our collaboration partners at Eberhardt-Karls-University of Tübingen, especially Prof. Dr. Marius Ueffing, Dr. Karsten Boldt and Tobias Leonhardt. I also like to thank the team of the Nikon Imaging Center for their assistance during live cell microscopy. Last but not least I would like to express my gratitude to Prof. Dr. Rocio Sotillo and the DKFZ for assisting me with Microarray analysis.

Many thanks also go to all former and present lab members of *AG Protein Evolution* and colleagues from other groups or institutes for their helpfulness and their contributions. My gratitude goes especially to Nina Bremec for her assistance in the laboratory. Thanks also go to Dr. Samantha Ebersoll, Jonas Weidenhausen and Natalie Dirdjaja for scientific as well as non-scientific discussions during lunch breaks. Furthermore, I am grateful for Yvonne Lara-Martinez' very relaxed way of keeping things running in the background, and in this regard, I want to express my gratitude also to Christiane Graffy and the whole BioQuant staff, dealing professionally and quickly with many technical aspects of daily lab work.

Last but not least I am very grateful to my parents and sisters, as well as to Charlotte, who always supported me in all matters.

Summary

Data generation is rapidly progressing and data interpretation is hardly keeping up. New tools and approaches are needed to assess, filter and decide upon the impact of variants on biological systems. This work first presents a novel method of interpreting variants found to be exclusively heterozygous in public datasets (1000 Genomes Project and gnomAD). Variant pathogenicity is scored by a benchmarked, Bayesian integration of a diversity of gene, protein, sequence and structural features.

A new 3D clustering method was developed to identify and understand pathogenic variant mechanism. This method uses known functional knowledge from a protein of interest and is homologs, and intramolecular distances from known/predicted structures to group and map functional information. This puts candidate variants into context by according to previously known functions of residues within the same group.

Experimental validation of putative pathogenic variants is important for understanding variant mechanism. Several variants of RHOA, a small GTPase relevant to many crucial signalling pathways, were subjected to a battery of laboratory experiments. Variants affecting sites involved in interactions with other proteins are known to typically cause a loss-of-function phenotypes in cancer. However, it became evident that variants showed unique mechanisms in terms of observed phenotypes. Overall it can be argued that there is no single mechanism related to RHOA dysregulation in cancer.

Lastly, a new method was developed with the aim of assisting diagnostics via clinical genetics. The method is able to produce a biological triplicate of laboratory results within four to five weeks, which would be a clinically useful timeframe in which to act on decisions related to dysregulated proteins. Modified HEK cells are transfected with tetracycline-controlled plasmids, containing the wild-type and mutated genes of interest. After selection and tetracycline induction the cells are harvested and total cell RNA is used to study gene expression via microarrays. Gene expression patterns can then be used to assess whether a given protein mutation is changing the enzyme activity, which can additionally be tested by rapid follow-up experiments such as phosphoantibodies. This hope is that the results can then be used to adjust treatment regimens, for instance, by highlighting putative oncogenes.

Zusammenfassung

Die Generierung von Daten schreitet mit unglaublicher Geschwindigkeit voran, die Interpretation der Daten jedoch kann dabei kaum mithalten. Neue Werkzeuge und Herangehensweisen sind nötig um die Daten zu analysieren und zu filtern und zur Entdeckung neuer pathogener Mutationen. Diese Arbeit stellt eine neuartige Methode vor, um öffentliche Datensätzen (wie das 1000 Genom Projekt und gnomAD). Neu zu interpretieren. Die Bestimmung der Pathogenität einzelner Varianten fußt hierbei auf eine Sammlung verschiedener Datenpunkte (Gene & Protein Sequenz, Strukturelle Daten), welche über bayesianische Integration letztlich zu einer Bewertung führt. Diese Bewertung kann dann dabei helfen krankmachende Mutationen zu identifizieren, da eine höhere Bewertung indikativ ist für eine größere Wahrscheinlichkeit eine Krankheit auszulösen.

Um neue krankmachende Mutationen nicht nur zu identifizieren sondern auch zu verstehen, wurde eine neue 3D Gruppierungsmethode entwickelt. Diese Methode nutzt bekanntes funktionelles Wissen über das Protein von Interesse (und homologer Proteine) zuzüglich intramolekularer Distanzen, um funktionelle Informationen zu gruppieren und auf die Proteinstruktur zu übertragen. Dies ist hilfreich um Kandidatenmutationen in einen funktionellen Kontext zu setzen, indem man sich auf bekannte Funktionen umliegender Positionen bezieht.

Experimentelle Validierung von zuvor identifizierten pathogenen Varianten ist ein wichtiger Schritt moderner Wissenschaft. Mehrere Mutationen in RHOA, eine kleine GTPase relevant für viele wichtige Signalwege, wurden im Labor untersucht. RHOA Mutationen an Positionen die typischerweise Interaktionen zwischen RHOA und anderen Proteinen vermitteln sind bekannt dafür, dass diese einen Funktionsverlust-Phenotyp in Tumoren produzieren. Für die untersuchten Mutationen konnte jedoch gezeigt werden, dass diese auf verschiedene Weise einen Funktionsverlust auslösen, und dass nicht ein einzelner Signalweg zur Progression einer Krebserkrankung beiträgt.

Darüber hinaus wurde eine Methode entwickelt um medizinische Forschung besser assistieren zu können. Diese neue Methode ist in der Lage ein biologisches Triplikat in vier bis fünf Wochen zu produzieren, was ein klinisch nützlicher Zeitrahmen darstellt. Zuerst werden modifizierte HEK Zellen mit Tetrazyklin-kontrollierbaren Plasmiden transfiziert, welche das Gen von Interesse beinhalten. Nach Selektion und

Induktion mit Tetrazyklin werden die Zellen geerntet. Zelluläre RNA wird dann genutzt um die Genexpression mit Microarrays zu untersuchen. Genexpressionsmuster können anschließend genutzt werden um zu bestimmen ob bestimmte Mutationen die Aktivität des Proteins verändern, und welche mit Nachfolgeexperimenten weiter getestet werden könnte. Hoffentlich können diese Ergebnisse genutzt werden um Behandlungsmethoden zu verbessern, indem mögliche Onkogene aufgezeigt werden.

❖ List of Publications

1) Schmenger, Torsten; Diwan, Gaurav; Singh, Gurdeep; Apic, Gordana; Russell, Robert Bruce. “Never-homozygous genetic variants in healthy populations are potential recessive disease candidates.” npj Genomic Med, September 2022, https://doi.org/10.1038/s41525-022-00322-z
2) Andre, Timon; van Berkel, Annemiek A; Singh, Gurdeep; Abualrous, Esam T.; Diwan, Gaurav D, Schmenger, Torsten; Braun, Lara; Toonen, Ruude F; Freund, Christian; Russell, Robert B; Verhage, Matthijs; Söllner, Thomas H. Improved monogenic disease prediction: a multi-dimensional approach using protein-specific information, validated in STXBP1 syndrome, Genome Medicine (submitted), September 2022.

❖ Presented Posters

BZH Meeting 2018, 15. - 17. July 2018, *How Natural Variants Affect Human Health and Disease*

BZH Meeting 2019, 25. – 27. July 2019, *How Natural Variants Affect Human Health and Disease*

❖ Presented Talks

BioQuant Internal Seminar Series, 13. December 2018, *Exploring the Effect of Natural Variants on Human Health and Disease.*

BZH Department Seminar, 23. April 2019, *Exploring the Effect of Natural Variants on Human Health and Disease.*

BZH Meeting 2021, 22. July 2021, *Keyboard Science: Computational Approaches to Interrogate Molecular Mechanism.*

❖ Awards

Best Talk, BZH Meeting 2021, *Keyboard Science: Computational Approaches to Interrogate Molecular Mechanism.*

❖ Contributions

Chapter II of this thesis was submitted for publishing under the title “Never-homozygous genetic variants in healthy populations are potential recessive disease candidates” by Torsten Schmenger, Gaurav D. Diwan, Gurdeep Singh, Gordana Apic and Robert B. Russell. Data, text and figures are based on Chapter II, which were originally analysed, written and drawn by me unless specified otherwise.

Two experiments were performed by collaborators in Tübingen (Chapter IV, 4.3.11.) or collaborators from DKFZ (Chapter V, 5.3.4.) which are mentioned in the respective Methods sections. Data analysis was afterwards conducted by me.

Chapter I: Introduction

1.1 Keyboard science and its challenges

The year 1972 saw a publication of 65 closely related sequences in Dayhoffs Protein Atlas[1]. This was a decade after what was possibly the first distributed (in FORTRAN via punch cards) protein sequence analysis program COMPROTEIN (to assemble Edman protein sequencing reads[2]). This foreshadowed decades of computational tool developments. In 2022 the number of bioinformatics tools exceeded 22.000[3] and the number of R packages released on CRAN climbed to over 18.000 **(Fig. 1A, A)**.

Yet this jump in available bioinformatics software is of course dwarfed by the number of sequences deposited monthly into public databases, with more than 1500 public databases currently available[4]. For example, between June 2021 and August 2021 more than 20 million sequences were stored in the NCBI GenBank and Whole Genome Shotgun (WGS) database, **Fig. 1A, B**). The difficulties for researches worldwide are obvious: which sequence to look at, what to look for, and how to look, and finally, how to make sense of it. The benefits present itself, as these data can ultimately help researchers to better understand monogenic diseases such as Huntington's[5] or Cystic Fibrosis[6] as well as complex diseases including breast cancer[7,8] or diabetes mellitus[9–11].

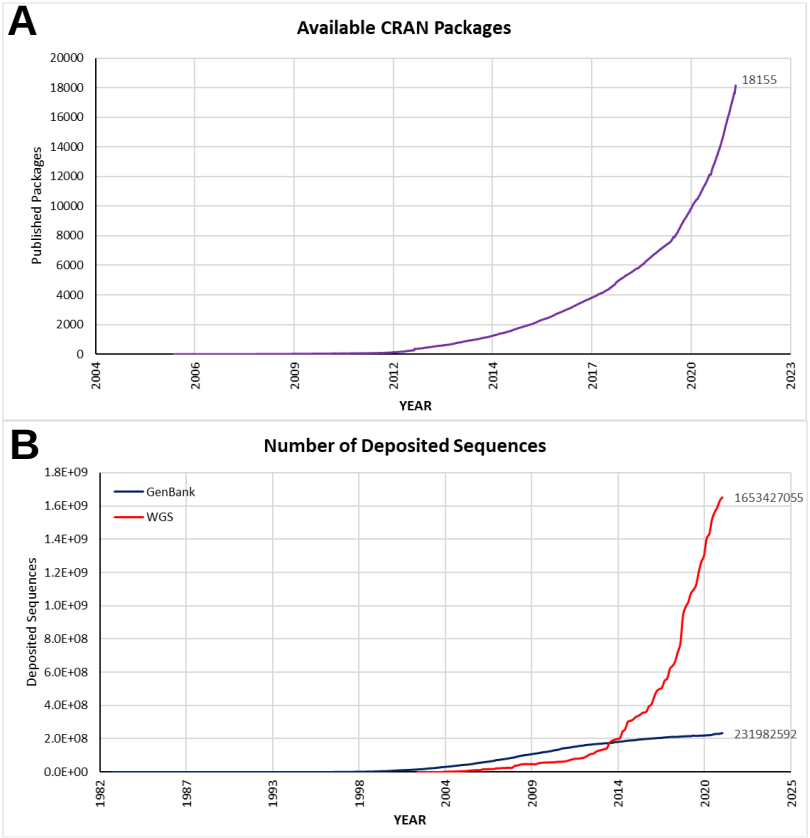


Figure 1A. Rapid Development of Bioinformatic Tools and Data sources. **A** The number of CRAN packages released every year started to significantly increase in the year 2014. Data retrieved from https://cran.r-project.org/web/packages/available_packages_by_date.html. **B** Sequences deposited in public databases such as GenBank. Sequences deposited through WGS dramatically increased and outnumbered the sequences in GenBank by 2014.

1.2 Comparing apples to oranges – data sources

In 1977 the first publicly available genome was that of bacteriophage ϕX174 (5386 base pairs)[12] and almost three decades later in 2001 The Human Genome Project released their version of the human genome (over 3 billion base pairs) and a cost of approximately 500 million USD. The drastic reduction in cost and increase in speed now means scientists know tens of thousands of genomes. Of special interest are datasets of healthy individuals such as The 1000 Genomes Project (1kG)[13] or the Genome Aggregation Database (gnomAD)[14]. Clinical researchers for example can use this information to compare their patient data to healthy genomes. Databases can also be specific for a large spectrum of clinically relevant diseases (ClinVar)[15] or for a group of diseases such as cancer (Catalogue of Somatic Mutations in Cancer, COSMIC)[16]. There are a number of databases dedicated to specific diseases, for instance ADHDgene, a database collecting variants causative of Attention Deficit Hyperactivity Disorder[17].

While all these curated datasets are similar, as they present their information on the genomic level, the user typically encounters several challenges. Some annotated variants refer exclusively to germline, inherited variants for example reported in 1kG, or gnomAD. Cancer databases, such as COSMIC, often focus on somatic variants that accumulate in tumours, while other databases (i.e. ClinVar) contain both variant types. Furthermore, users usually have to filter datasets according to variant type. Comparison across datasets is not only difficult because of differing contexts, for example healthy versus diseases genomes – how was healthy defined? Which age groups were included? Which populations were considered? These questions often have to be answered individually for each dataset, further making it difficult to perform comparisons across them. Allele frequency (AF) relative to the dataset size is another parameter to consider, rendering it quite difficult to evaluate (and compare) absolute counts. Interestingly, some variants are presented with AF > 0.5, suggesting that the observed "variant" is the dominating variant in the surveyed population (i.e. WDR25 Trp88Arg, seen

251161/251244 times). These are clearly the result of a selected few individuals being sequenced in the “reference” genome, which has implications for how the reference is defined. This raises questions of which reference genome to use, how changes are being tracked and why different databases even use different reference genomes (e.g. COSMIC uses GRCh38 and gnomAD defaults to GRCh37). These discrepancies also question the usefulness of legacy data, and the debate in the field still ongoing[18–20].

Moreover, many databases do not fully report genotypes. For example, whether a missense variant affecting a known post-translational modification (information found in proteomic databases) was found to be either homozygous or heterozygous is important to determine if, and how, a variant is causative of disease. Dosage sensitivity in genes is captured in yet another database (ClinGen, https://dosage.clinicalgenome.org/). Understanding variant impact invariable requires multiple cross-database comparisons.

1.3 Strings of change – the many types of variants

Biomolecular scientists studying genetic variants will see themselves eventually confronted with a string of letters indicating a change in a gene. Unfortunately, gene identifiers often differ for each database, i.e. identifiers for the tumour suppressor gene TP53 might range from a simple ‘P53’ or ‘TP53’ to ‘P04637’ (UniProt), ‘191170’ (OMIM), ‘NM_000546’ (NCBI RefSeq), ‘ENSG00000141510’ (ENSEMBL) or ‘NM_000546.6(TP53)’ (ClinVar) and many more, unnecessarily creating an additional challenge in translating gene or protein identifiers across databases. Not speaking the same scientific language makes it easier to report results more vaguely, but is also setting up the scientific community to repeat mistakes, overall slowing scientific progress.[21,22]

However, once the gene ID has been identified, one needs to consider the change itself, as genetic changes can come in different flavours. The majority of variants lie outside of any coding region, and even if they do affect a gene, many variants will still cause no observable phenotype on the organism (neutral variants)[23]. In the context of somatic cancer neutral variants are often called passenger mutations, while variants considered to be causative of the disease are called driver mutations. Genetic variants are normally classified by a five-point scale, where number 1 and 2 contain variants with little to no clinical significance, 3 holds variants with uncertain significance and the higher numbers 4 and 5 are assigned to variants that are likely pathogenic/pathogenic[24]. If a variant is found during a sequencing project without any link to an obvious phenotype it

is often called a ‘Variant of Unknown Significance’ (VUS). Because mutations are physical changes to the DNA not all VUS are equal. Point mutations change only a single letter, causing a single nucleotide polymorphism (SNP) that either changes the resulting amino acid (non-synonymous) or has no effect on the translated product (synonymous)[25]. Non-synonymous mutations are often deleterious, but synonymous variants can lead to disease by more subtle mechanisms (i.e. translation speed, tRNA selection, elongation rate)[26,27]. Insertions or deletions (indel) in various sizes are more likely to change the protein structure. These indels can either insert/delete trios of nucleotides, where the protein sequence remains (but for the inserted amino acids) largely unchanged (in-frame indels), or insert/delete any other number of nucleotides (not divisible by 3) to change the translational reading frame, resulting in a changed – and often destroyed – protein (frameshift, fs). These changes can also introduce or create a premature stop-codon (nonsense), where protein translation stops too early. Another type of variant is splice variants at the boundary of introns and exons. Such variants can lead to a loss of exons or a false inclusion of introns into the sequence to be translated[28]. On one hand, the consequences of VUS variant types, such as indels, splice or synonymous variants are not as easy to predict. On the other hand, fs or nonsense mutations are often causative of disease phenotypes. For example a fs mutation in COL5A1 is causative of an atypical form of Ehlers–Dalnos syndrome[29], hearing loss is associated with fs mutations in GRXCR2[30] and nonsense mutations are frequently found in Duchenne muscular dystrophy patients[31].

There are also additional structural variants, often affecting large parts of genomic DNA by either inverse and hence change the orientation of DNA or chromosome mutations, often visible under the microscope, resulting in chromosomal loss or duplication, with severe effects on the individual[25].

A widespread approach to decide on the impact of a given amino acid change is to consult substitution matrices, such as PAM[32] or BLOSUM[33], where amino acid variants are judged by comparing them to observed changes throughout protein evolution. This requires an alignment of the protein sequence of interest with similar sequences in order to determine whether the variant affects a conserved (i.e. largely unchanged across the aligned sequences) or variable residue. The level of conservation does not provide direct clues into function, but is a general measurement of its importance in terms of protein structure or function.

Variants can illicit effects on function in many different ways. For example many disease causing amino acid changes are buried within the protein structure, suggesting that they might affect protein stability[34,35]. However, surface/exposed variants can also be of special interest, as they might perturb protein-protein interactions (PPIs) or protein-small molecule interactions, with the latter point being particularly important to understand the proteins response to small molecule inhibitors[36]. Structural context is also important and often means that the same residues can play very different roles. For instance, a histidine or serine might be part of a catalytic triad in one protein (**Fig. 1B, A**), while another histidine might coordinate a zinc (Zn) atom (**Fig. 1B, B**) and another serine can be a phosphorylation site in yet another protein (**Fig. 1B, C**). Amino acids should not merely be reduced to their level of sequence conservation alone but they must also be examined relative to this biological context. This is even more important as it has been shown multiple times that sequence conservation is detached from structure conservation: protein folds can be similar even when there is little sequence similarity, for example the typical helix-turn-helix motif of some histones[37] or WD repeat-containing proteins (**Fig. 1B, D**). Modern day biologists therefore need a refined battery of tools in order to elucidate the function of an individual VUS.

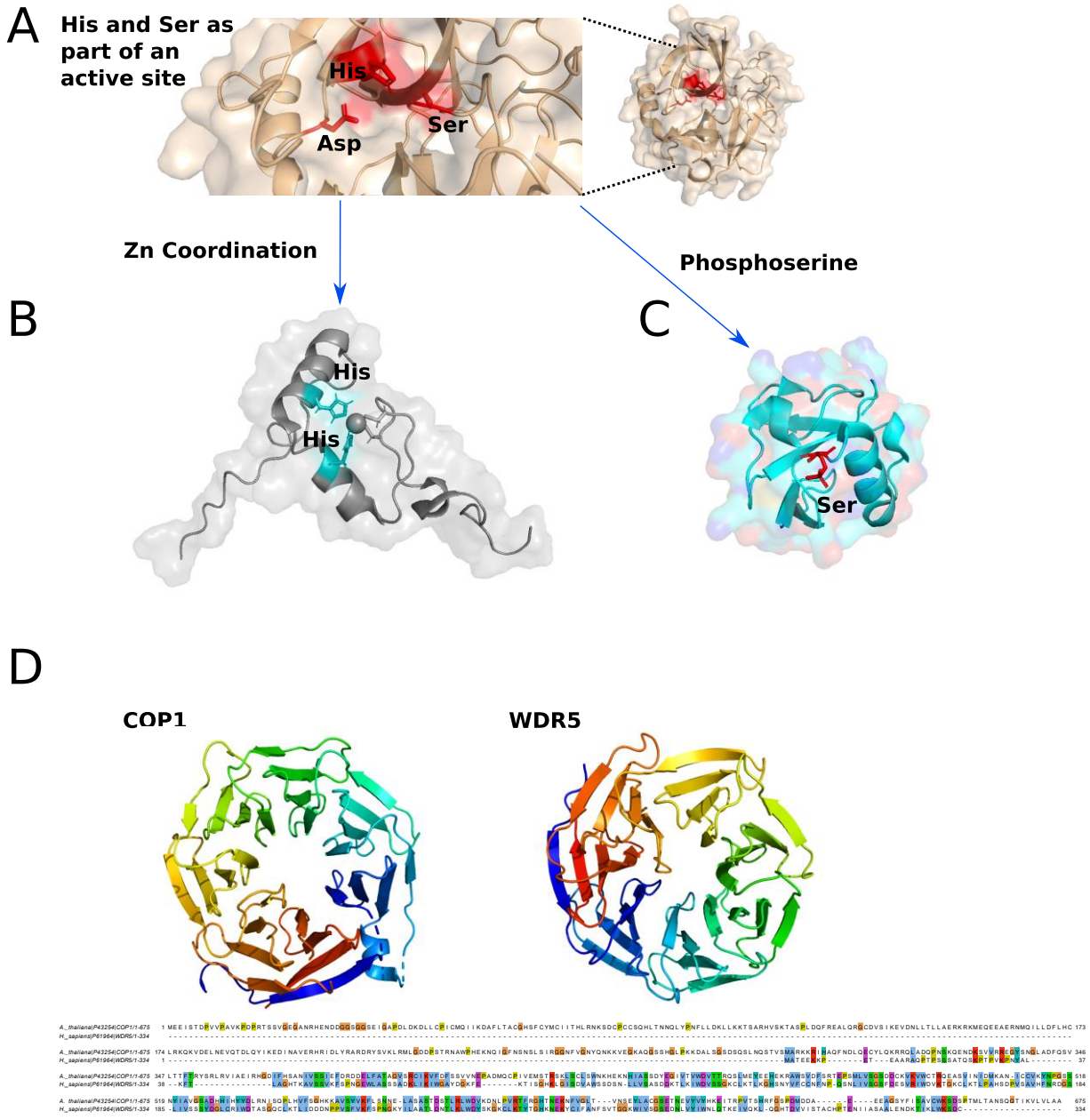


Figure 1B. Amino acid Function Depends on Context **A.** Catalytic triad of the human serine protease Chymase 1 (CMA1, PDB: 1KLT) **B.** Two histidine contribute to Zn coordination in Zinc Finger Protein 593 (ZNF593, PDB: 1ZR9). **C.** A serine is phosphorylated in Polyubiquitin-B (UBB, PDB: 5K9P). **D**. A. thaliana COP1 (top left, PDB: 6QTT) has a conserved structure compare to H. sapiens WDR5 (top right, PDB: 6OFZ) despite only sharing 22.57 % sequence identity.

1.4 Obstacles in linking variants to mechanisms

Understanding the effects of a SNP and a resulting single amino acid change is still difficult. Protein changes are usually described as having one of three outcomes: neutral (i.e. nor perceivable effect), loss-of-function (LoF) or gain of function (GoF). LoF variants

are typically thought to lessen or ablate the normal function of the protein. GoF is harder to clearly define, though it can often mean a hyper- or constitutive activation of a protein. In rare cases, GoF can mean a genuine new function, i.e. Glu545Lys of PIK3CA gains the ability of PIK3CA to associate with insulin receptor substrate 1, rewiring several oncogenic signalling pathways[38]. As might be predicted, many variants don't fit neatly into this tripartite classification: there are grey areas in between, for example, where only certain protein functions are affected while others remain normal. Often, moreover, there are conflicting lines of evidence making it harder to understand the effect.

A common explanation for missense LoF variants, especially when they can be scattered haphazardly around a protein, is that they cause protein misfolding. Misfolded proteins are the cause of diseases such as Cystic Fibrosis or Alzheimer's[39–41]. Surprisingly fs variants might cause the same Cystic Fibrosis disease phenotype as some more subtle amino acid changes[42,43], and there are many examples where LoF variants are seemingly equivalent to missense variants, most often found in mono-genetic diseases[44].

Despite this quest for a simplistic classification of variants, the above examples show that things are subtler and more complicated. For example, the widely held view that variants scattered throughout the protein implies LoF due to effects on protein folding/stability (e.g. seen in tumour suppressors such as VHL and Rb1[16,45], **Fig. 1C, A**) while variants concentrated around specific sites suggests GoF via hyperactivation (e.g. seen in oncogenes such as EGFR, RAC1, KRAS or BRCA1[16,45]) does not always hold. For instance, mutations affecting SCNN1B, a gene involved in blood pressure regulation, can either cause Bronchiectasis, through LoF mutations, or Liddle Syndrome by disrupting certain PPIs[46]. Proteins are often multi-functional and possess more than one conserved domain, and certain variants affecting a specific domain can lead to a perturbation of only some interactions. This phenomenon where only some connections of a network are affected is called edgetics[47,48] (**Fig. 1C, B**).

The advancement of sequencing technologies happened in parallel with advancements in other fields of biology[49]. This lead to an ongoing large-scale look into protein networks and whether perturbation of certain edges are characteristic for diseases[50–54], and several tools and database were created to analyse and store such findings[50,55,56]. More specific examples included a detailed study of perturbed network edges to investigate how defects in upper motor neurons are either causative of

hereditary spastic paraplegia or primary lateral sclerosis[57]. Another study could detect perturbed interactions in different breast cancer samples, helping to reclassify breast cancer[58]. An extreme case is the near complete rewiring of protein interaction networks, that can not only be observed in cancer[59] but also in diseases such as schizophrenia[60] or heart failure[61]. Network rewiring suggests that some proteins must act as hub proteins (being in the centre of many PPIs). A form of mutual exclusivity can be also observed in cancer, where putative hub proteins are the target of pathogenic variants but not their interaction partners[62], perhaps due to a lower perturbation tolerance of the latter. Differential gene expression is complicating things further, as the expression for members of a network might change in a tissue-dependent context. This information is available[63] but not often consulted.

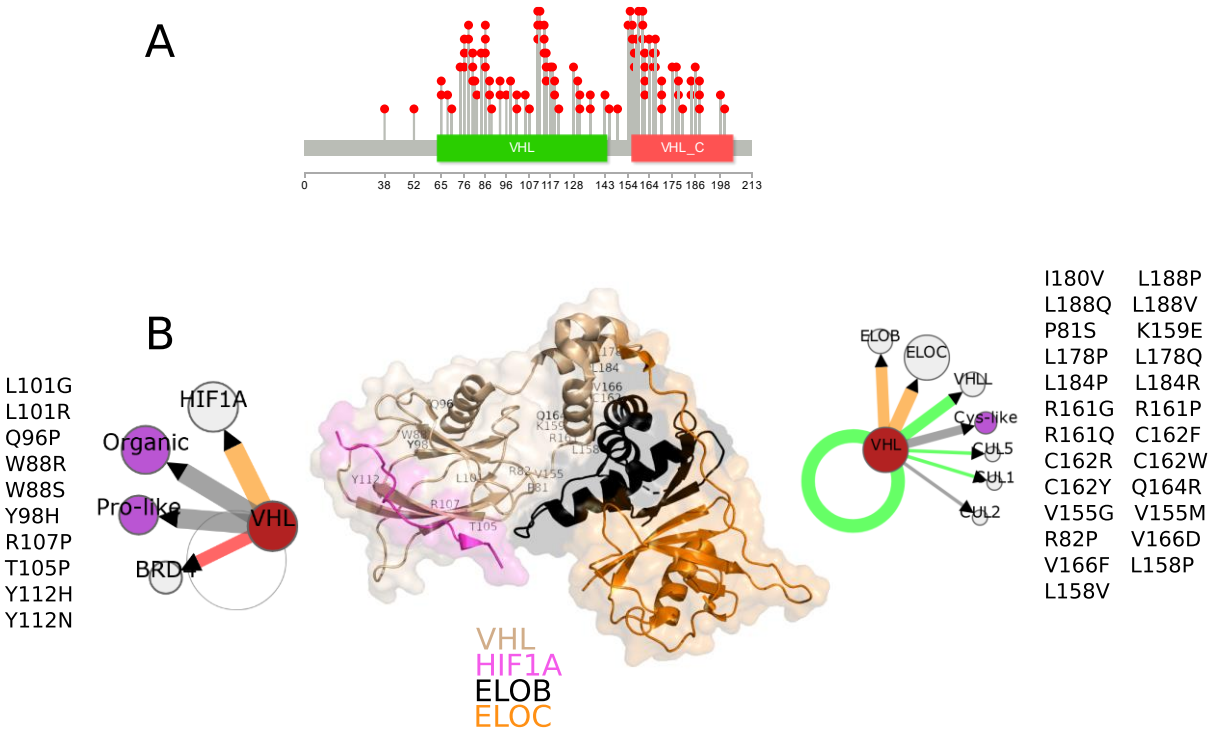


Figure 1C. Hereditary VHL Mutations **A.** Domain and variant representation of VHL[64]. VHL variants causative of Von-Hippel-Lindau syndrome are scattered throughout the protein, suggesting protein misfolding and destabilization as the main LoF mechanism. **B.** Annotated VHL variants can either perturb the VHL-HIF1A or the VHL-ELOB/C interfaces. Interface perturbations were predicted with Mechismo[55]. PDB: 1LM8.

1.5 Conclusion

Overall, the field is rapidly moving forward even if not every discipline of *in silico* research is holding pace with the flood of data that is being produced. The data generated are certainly promising, but clearly new tools and methods are needed to bring together every aspect of research. No single discipline is able to reveal the full story, meaning that generalized views are increasingly found to not hold up to reality, and an immense amount of work is required to understand mechanistic details. A first, albeit very small step, could be a more direct and logical naming approach, which would help in unifying the existing data, for example simply by being easier to compare. However, there are excellent possibilities to better predict variant impact and putative function by an integration of existing data and disciplines. Certain tools[46] already do parts of this, though a greater integration of biological data and indeed disciplines is ultimately required for this to become a reality.

However, until such a more holistic approach will be achieved, one must be aware that many analyses, while not being wrong, will only be part of a bigger, and in many cases perhaps hidden, picture. The increased productivity of the bioinformatics community observed since the outbreak of COVID19 pandemic is uplifting[65,66] and it will be exciting to see which tools and methods future generation scientists will have in stock.

1.6 Aim of the PhD Project

The aim of this thesis was to apply established and newly developed methods to existing datasets to predict mechanistic insights on genetic variants in addition to testing predictions with experimental methods.

In Chapter II, the publicly available data of the 1000 Genomes Project (i.e. variants with ostensibly healthy people) was explored. Specifically, I uncovered a curiously high number of exclusively heterozygous variants: those variants seem in comparatively high counts as heterozygous, but never/rarely homozygous. I explored where such variants, often with allele frequencies > 1 %, could be indicative of a disease in the homozygous state. The identification of novel disease-associated variants could help in understanding certain protein functions and their potential importance in human health and disease. Results based on this chapter were later published[67].

Some variants affect genes that have not been well studied. This limits the *in silico* methods scientists can use to further our understanding of the genes function. I worked on a new approach to functionally cluster known variants. Variants known from homologous genes based on 3D-structure distances (Chapter III) further help to collect information about less understood genes. The availability of predicted structures (Alphafold) for nearly all human proteins enables this novel approach.

In silico approaches should ideally be confirmed with experimental data. In Chapter IV, I study several previously observed disease variants of RHOA using Yeast-Two-Hybrid assays, cell culture experiments and protein affinity purification to determine aspects of their phenotypes. The results surprisingly reveal a diversity of outcomes for the set of loss-of-function variants which raises questions regarding how one should define variants in the future.

Lastly, I developed an experimental workflow using HEK cells as a tool for rapid testing of variant function (Chapter V). HEK cells were transfected and induced with tetracycline to overexpress genes in a controlled manner. Afterwards, gene expression profiles are studied in order to determine activating or inactivating effects of protein variants compared to their wildtype form. Tests with known variants show considerable promise for this workflow to assess activating variants rapidly such that they might ultimately be useful in clinical diagnostics.

Chapter II: Never-homozygous, genetic variants in healthy populations as potential recessive disease candidates

2.1 Introduction

Somatic mutations can be the cause of cancer and other non-malignant diseases by accumulating in every cell of a human's body during life. Mutations often stem from errors during DNA replication or repair, and the number and therefore consequences increases with age[68,69]. Besides several forms of spontaneous cancer one other well-known example is Neurofibromatosis 1, a disorder with increased susceptibility to the formation of benign and malignant tumours, with somatic mutations in the gene NF1 being a major cause (i.e. up to 50% of all cases are somatic)[70,71]. Notably, somatic mutations are not passed on to following generations.

Predisposition for both cancer and Neurofibromatosis[72] generally stems from germline variants. Not only do humans inherit half of the genome from each parent, there is also a chance to inherit putatively damaging variants. The risk of inheriting disease-causing mutations was shown to increase with parental age, and diseases typically associated with inheritance can be cancer, as well as cystic fibrosis or some cases of deafness (**Fig. 2A**). The parental carriers can be unaffected by the recessive allele and phenotypes might develop in their children if both parents pass on the recessive allele, resulting in the child carrying 2 damaged copies of the respective gene. However, in some cases a disease can develop in the presence of only 1 damaged allele (dominant).

Another significant difference between germline and somatic variants is their onset. Germline variants increase the predisposition to give rise to a disease such as cancer, reducing the individual's fitness more strongly by more often developing at an earlier age. A well-studied example is the formation of colorectal cancer. Here, patients with a family history of colorectal cancer develop tumours on average approximately 10 years before patients without a family history of colorectal cancer[73], with important implications for diagnosis and preventive measures.

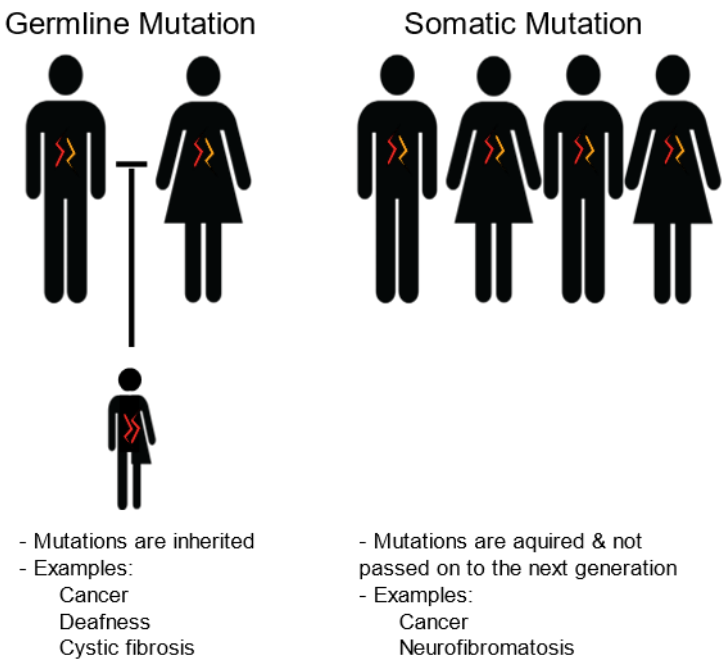


Figure 2A. Different types of mutations. Left) Germline mutations, inherited from each parent. The mutation might be silent in the parental generation (1 intact & 1 damaged gene copy) with a risk of unknowingly passing on 2 damaged copies to the next generation, causing a disease with early-onset. **Right)** Somatic mutations, acquired during life. These variants are not passed on to the next generation, and diseases associated with somatic mutations often show a late-onset.

The final phase of the 1kG[13] contains millions of germline variants from 2504 ostensibly healthy individuals, sampled from 26 populations, grouped into 5 superpopulations: American (AMR), African (AFR), East Asian (EAS), European (EUR) and South Asian ancestry (SAS)(**Fig. 2B**). The 1000 Genome Project (1kG) marked a next milestone after the completion of the Human Genome Project[74] in 2001, as the 1kG contributed (or confirmed) 80 million variants in the public dbSNP database, which at the time attributed to 80 % of dbSNPs[13].

Amongst the 1kG findings is the description of a 'healthy' genome. A typical genome differs on 4.1 – 5 million bases from the reference genome[13]. These changes include 2100 – 2500 structural variants, of which ~ 1000 are deletions, ~ 160 are copy number variants and the remaining alterations are either insertions or rearrangements, affecting a total of 20 million bases[13]. Furthermore, the 1kG charted 149 to 182 protein truncating variants per 'healthy' genome as well as 10.000 to 12.000 missense variants[13]. Additionally, half a million variants lie in regulatory regions[13]. A typical and seemingly

healthy genome also carries up to 2000 variants associated with complex traits, as well as 24 – 30 variants implicated in rare diseases[13].

Notably, the AFR population shows the highest diversity, which is in support of the 'Out of Africa Theory'[75,76], describing the migration of modern human out of Africa. However, the 1kG also noticed that, while individuals of the African ancestry show the highest genetic diversity, it is individuals of European ancestry that show the most variation related to genetic diseases. This is, however, not explainable with demographics or population genetics but rather demonstrates that there is a research bias, where medical research is focusing mostly on diseases that are prevalent in western societies[13].

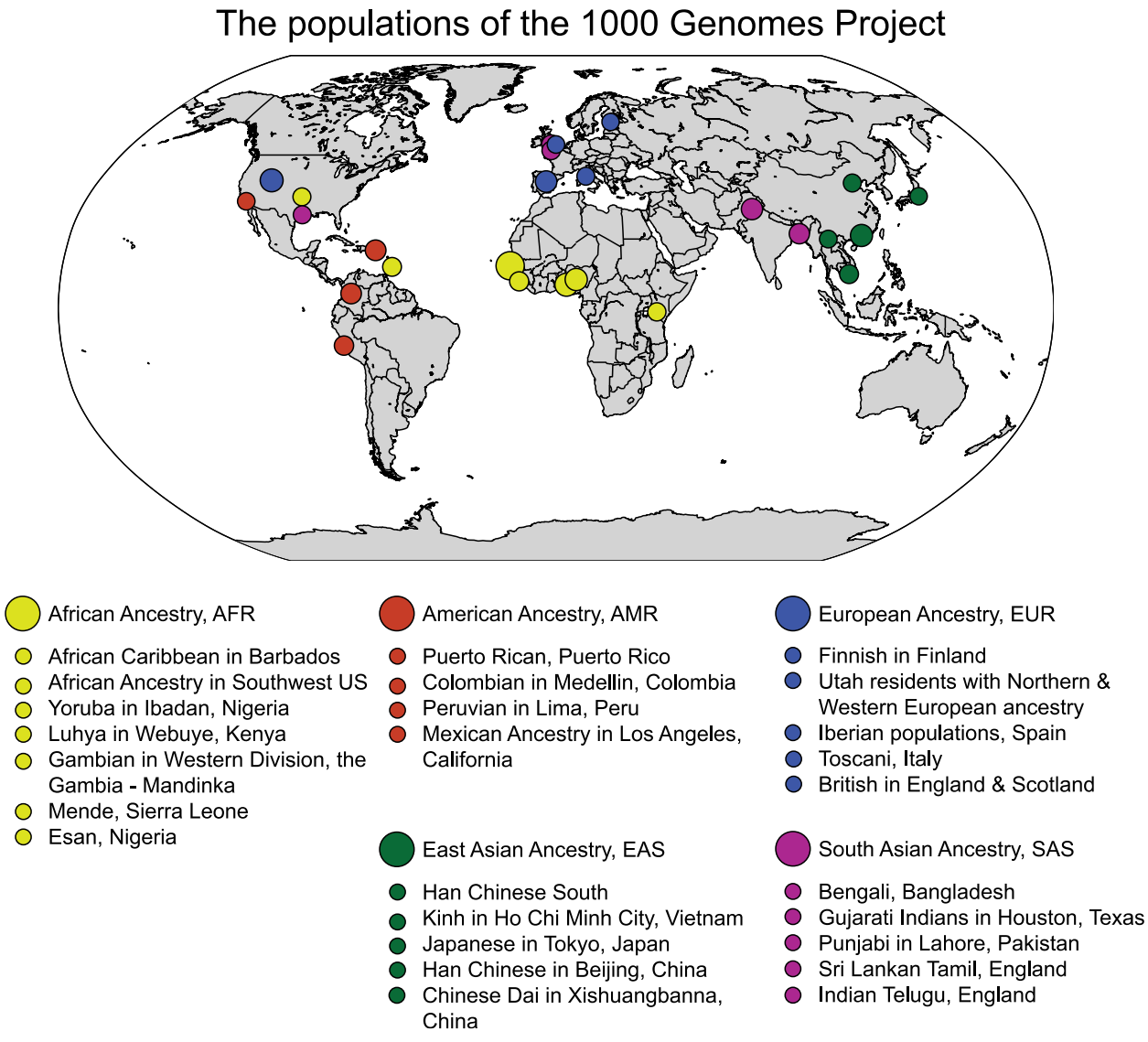


Figure 2B. The populations of the 1000 Genomes Project. Sampling locations are indicated in the respective Superpopulation colour on the world map. Circle diameter corresponds to the number of individuals sampled on-site. African Ancestry (AFR) = yellow; American Ancestry (AMR) = red; East Asian Ancestry (EAS) = green; European Ancestry (EUR) = blue; South Asian Ancestry (SAS) = purple.

The aim of this project was to explore the data presented by the 1kG. Is it possible to use the 1kG data for more than simply as a background model of human genetic variation? Having noticed frequent exclusively heterozygous variants, I set out to look at them as potential disease candidates using a variety of different prediction tools, public databases and custom algorithms.

2.2 Material and Methods

2.2.1 1000 Genomes Project data processing

Missense single nucleotide variants (SNVs) from variant call files (VCFs) of the 1kG (Phase 3, 2,504 individuals) were extracted and gnomAD (V2.1.1, 125,748 exomes) data was handled in similar fashion. Only those 1kG variants where data were available in gnomAD and where the difference in minor allele frequency (MAF) was ≤10% were considered. Some variants displayed a MAF > 50%, these instances were also converted by inverting them (100-MAF)[67]. Variants other than missense were also considered (synonymous, in-frame indels, frameshifts, stop-gains, UTR, splicing, intronic and non-coding). 11.000 variants were randomly selected and heterozygous vs. homozygous counts were compared.

2.2.2 Defining exclusively heterozygous variants

The parental genotype in 1kG data is unknown. If it can be assumed that homozygous variant carriers are viable and able to reproduce, the chance of homozygous offspring is approximately 0.29 after combination of all possible G0 constellations (AA x AA, Aa x Aa, AA x aa, aa x aa, Aa x AA and Aa x aa with A = wildtype (WT) trait and a = mutant trait). However, when the assumption is that homozygous variant carriers are most likely unable to reproduce - and perhaps not even viable - the chance of homozygous offspring shrinks to 0.083, after combination of all remaining G0 constellations (Aa x Aa, AA x AA, Aa x AA). I used the Mendelian laws of inheritance-based likelihood for homozygous offspring

to subject each 1kG variant to binomial testing, defining the minimum requirement of ≥ 41 heterozygotes with 0 homozygotes for variants to be further considered[67] (**Fig. 2C**). Gene enrichment was performed using g:Profiler[77] with default settings, unless otherwise stated.

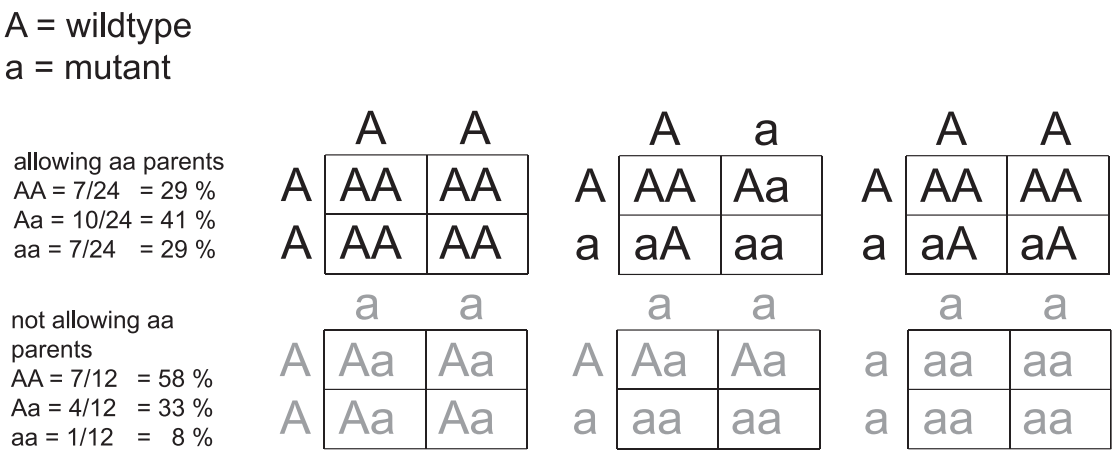


Figure 2C. Punnett squares showing the likelihood of homozygous offspring.

2.2.3 Genotype shuffling

A random decimal between 0 and 1 was chosen based on a uniform pseudo-random number generating algorithm[78] and compared to the observed allele frequency for a given variant. If the resulting number was smaller or equal to the observed allele frequency then this instance would be considered to simulate a mutated allele. Simulated individual genotypes would consist of two shuffled alleles. A total of 2504 individuals (5008 alleles) were subjected to this approach and each of the 1kG variants would undergo 100 cycles. Lastly, a mean average simulated genotype would be calculated from the heterozygous and homozygous counts for each variant of each cycle[67].

2.2.4 Sequence conservation analysis

Orthologs for all proteins in the Uniprot proteome of Human (Proteome ID - UP000005640; retrieved April 2021) were computed by Dr. Gaurav D. Diwan using the Orthofinder program[79]. In short, the canonical proteomes of Human and several hundred other organisms from the tree of life were used to calculate the orthologs. We used the option of computing multiple sequence alignments (MSA) to build gene trees which comes with an in-house species tree[67]. For every protein in the Human proteome we collected all orthologs across species. Orthofinder also calculated the MSA

for each group or homologous group that contains orthologs and paralogs. The alignments were calculated using the MAFFT L-INS-i method when there were <500 sequences in a group and the native MAFFT method[80] for larger groups. The alignments for orthologs were obtained by subsetting the Orthogroup alignments for each Human protein and its respective orthologs. Naturally, positions that contained all gaps were removed[67]. I created images of protein structures using PyMol[81].

2.2.5 Creating a bayesian score to evaluate functional impact

Alignments of orthologs and homologs were used to calculate HMMer profiles[82] which provided scores for each amino acid and each position. The score for variants was taken as the difference between the mutated value in these profiles and the wild-type. Additionally, scores from the BLOSUM62 matrix for each variant were used.

Furthermore, structures for all human proteins predicted by Alphafold[83] were used to define a variety of structural parameters, including: secondary structure, main-chain dihedral (psi/phi) angles, and accessibility using DSSP[84,85]. Moreover, the degree of burial, which is defined as the accessibility of a Gly-X-Gly tripeptide minus the DSSP accessibility, was computed. Amino acids were studied in representatives (fourth level of the hierarchy) of the ECOD database[86] (v281) to first define divisions into zones: secondary structure: helical (characters H,G) strand (E,B), or coil (others); dihedral angles: a 12x12 grid with phi and psi (-180 – 180) in increments of 30. accessibility: low (0-15), medium (16-59), (≥60); burial: low (0-114), medium (115-164), high (≥165). Then, log-odds scores of observed counts versus expected (based on the abundance of amino acids and the totals in each zone) were calculated. For every variant the score for each commodity was defined as log-odds mutant – log-odds wild-type) with negative values indicating a poorer fit for the mutant and vice versa. For structural parameters using Alphafold data the confidence scores and quality was omitted, not considering how this will affect wild-type and mutants. Lastly, the impact score from Mechismo[55] for each variant was calculated and devised an equivalent score using residue pair-potentials for intramolecular (in contrast to intermolecular) contacts across the ECOD dataset.

Information about approved drug targets was retrieved from the U.S. Food & Drug Administration[87] (FDA). I then scored genes where medications were already approved for and when the gene was listed as disease causing in the Online Mendelian Inheritance in Man database[88] (OMIM) with a value of 1.

For each gene I retrieved annotations on haplotype insufficiency from ClinGen[89,90]. Genes that were associated with an autosomal recessive phenotype received a (haplotype) score of 1 and decreased to 0.75 when sufficient information was available or to 0.5 and 0.25 when some or only minimal information was available. Absence of information or unlikeliness for dosage sensitivity scored 0.

To complete the analysis I collected information about post-translational modifications, active centres and known variants from UniProt[91].

All were combined into a functional impact score using Bayesian integration[92,93] (Formula 1).

$$log_2(O_{prior}) + \sum_{i=1}^{N} log_2(\frac{(D_i|P_{true}}{D_i|P_{false}}) \qquad (1)$$

Where $D_i|P_{true}$ and $D_i|P_{false}$ correspond to the true and false positive rates (TPR and FPR), which were obtained from ROC curves considering 26767 known disease causing variants from ClinVar as positives and a 4103 as negatives. $O_{prior} = 1$ was set arbitrarily[67].

2.3 Results and Discussion

2.3.1 1kG missense variants frequently display a lack of homozygosity

The non-synonymous variants in the 1kG dataset show a general increasing proportion of homozygotes (**Fig. 2D, A**), agreeing with mendelian inheritance. However, a notable number of missense variants within the 1kG dataset remained exclusively heterozygous even when MAF ≥ 10 %. The same trend could be seen for non-synonymous variants in gnomAD exomes (**Fig. 2D, B**), while the proportion of homozygotes increases with variant frequency, a significant number of variants remain exclusively heterozygous.

The Mendelian laws of inheritance suggests that homozygosity in the offspring generation should lie within 8-29%. Subjecting the 1kG variants to this assumption suggests that positions exclusively observed to be heterozygous in ≥41 genomes (in 1kG) would be unlikely to have zero homozygous (hom) counts. In allele shuffling simulations very few exclusively homozygous positions above this value and none at all above 259 (**Fig. 2D, C**) can be seen.

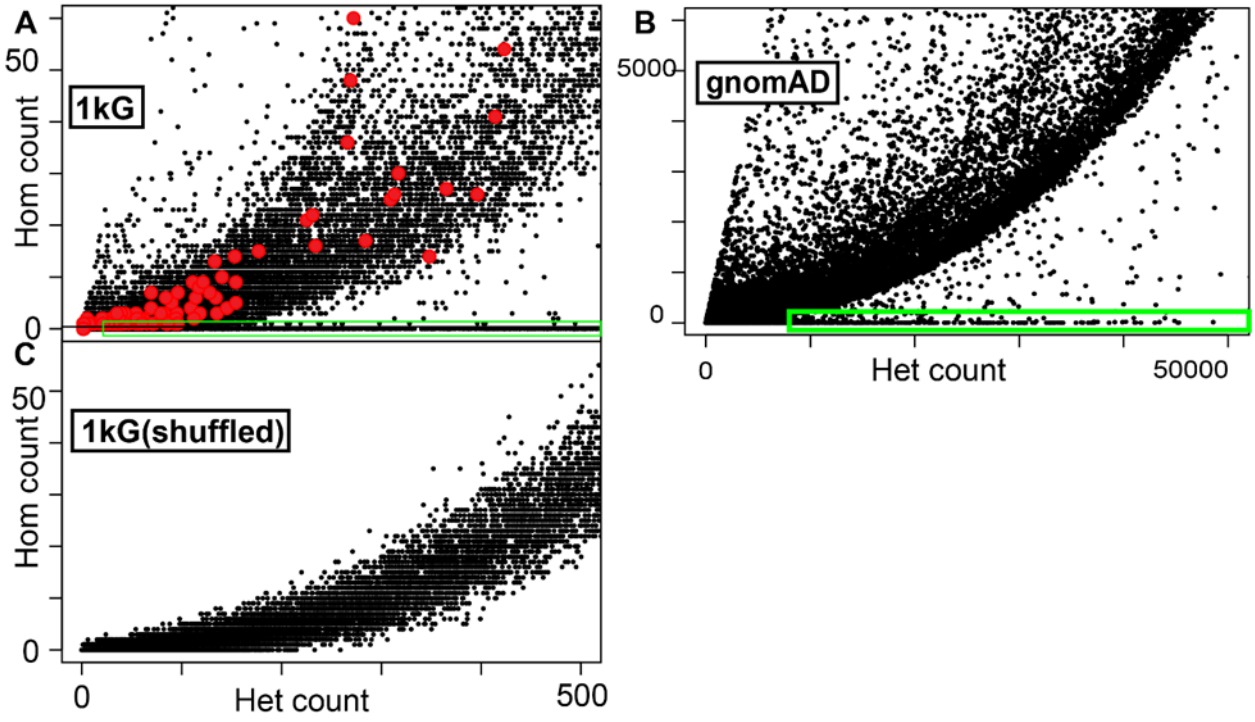


Figure 2D. Heterozygous vs. homozygous counts of different datasets. A. Plots of homozygous vs heterozygous counts for the 1kG dataset. The preponderance of values on the X axis (i.e. zero homozygous counts) are indicated, red dots indicate known disease variants. **B.** As in A. but with gnomAD data. **C.** As in A. but with allele shuffled 1kG data. Adapted from Schmenger et al. 2022.

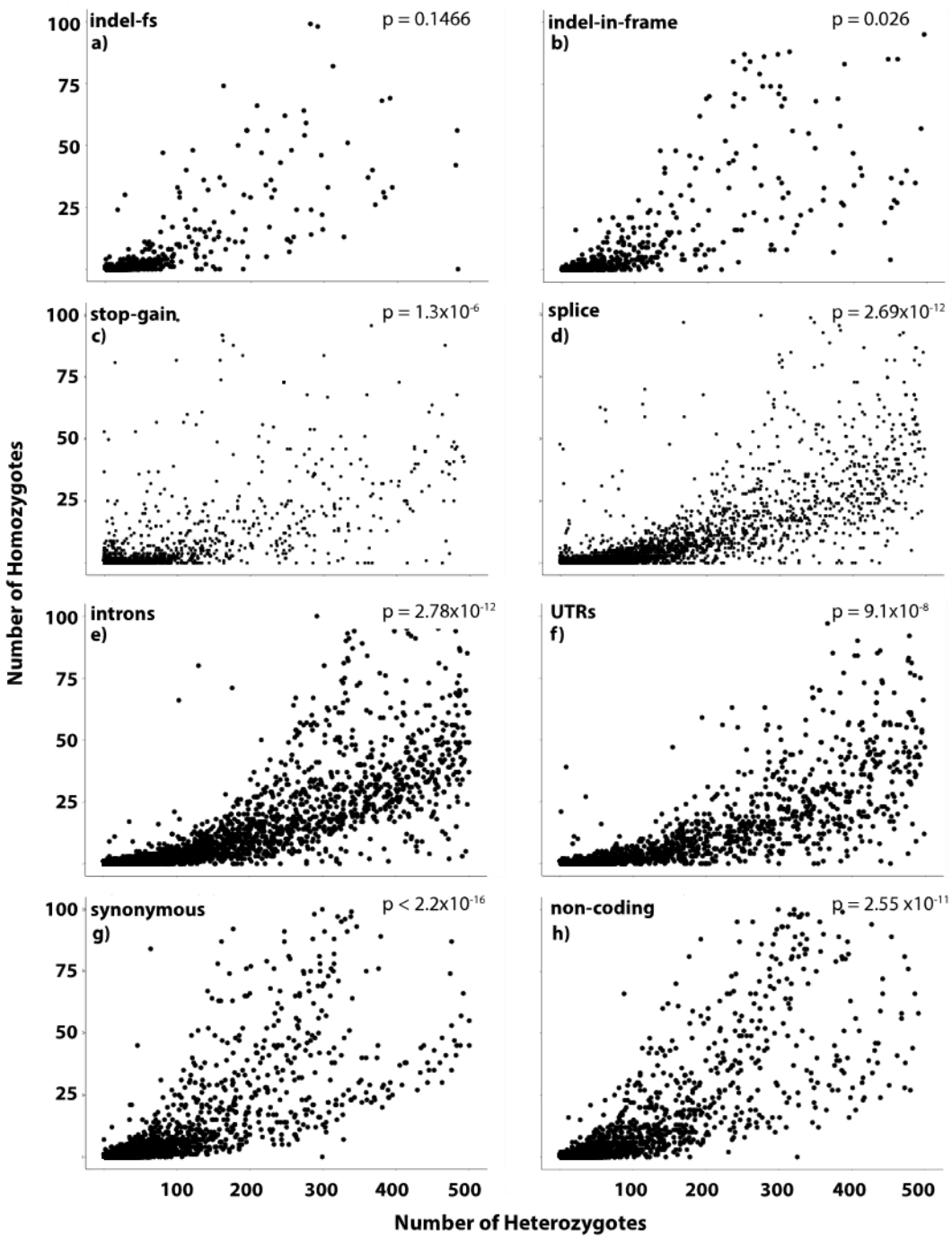


Figure 2E. Plots of homozygous vs heterozygous counts for a) 1k indel-fs variants; b) 1k indel-in-frame variants; c) 11k randomly selected stop-gain variants; d) 11k randomly selected splice variants; e) 11k randomly selected intronic variants. f) 11k randomly selected UTR variants (3'-UTRs and 5'-UTRs were equally considered); g) 11k randomly selected synonymous variants; h) 11k randomly selected non-coding variants. P-values are given for each variant type comparing (Wilcoxon rank sum test) exclusively heterozygous (≥ 41) counts to those of missense variants.

Notably, 2k frame-shift variants, in-frame insertions/deletions and stop-gains (**Fig. 2E, a-c**) have a number of exclusively heterozygous variants as might be expected as they are likely to alter/abate protein function[94]. Reassuringly however, significantly fewer exclusively heterozygous splice (Wilcoxon rank sum test, $p = 2.69 \times 10^{-12}$), intronic ($p = 2.78 \times 10^{-12}$), UTR ($p = 9.1 \times 10^{-8}$), synonymous ($p < 2.2 \times 10^{-16}$) or non-coding ($p = 2.55 \times 10^{-11}$, **Fig 2E, d-h**) variants were seen compared to missense. Overall, these observations support the notion that many of the observed exclusively heterozygous missense variants could be functional[67].

Subjecting exclusively heterozygous variants to enrichment analysis yielded several enriched gene ontology (GO) terms including many Biological Process terms related to detecting olfactory stimuli as well as cell-cell adhesion. These receptors are commonly found to be mutated in human populations[95]. Cellular Component terms related to the plasma membrane and cytoskeleton were also enriched, as were several terms related to human phenotypes, including those related to male fertility and cardiovascular disease (**Fig. 2F**).

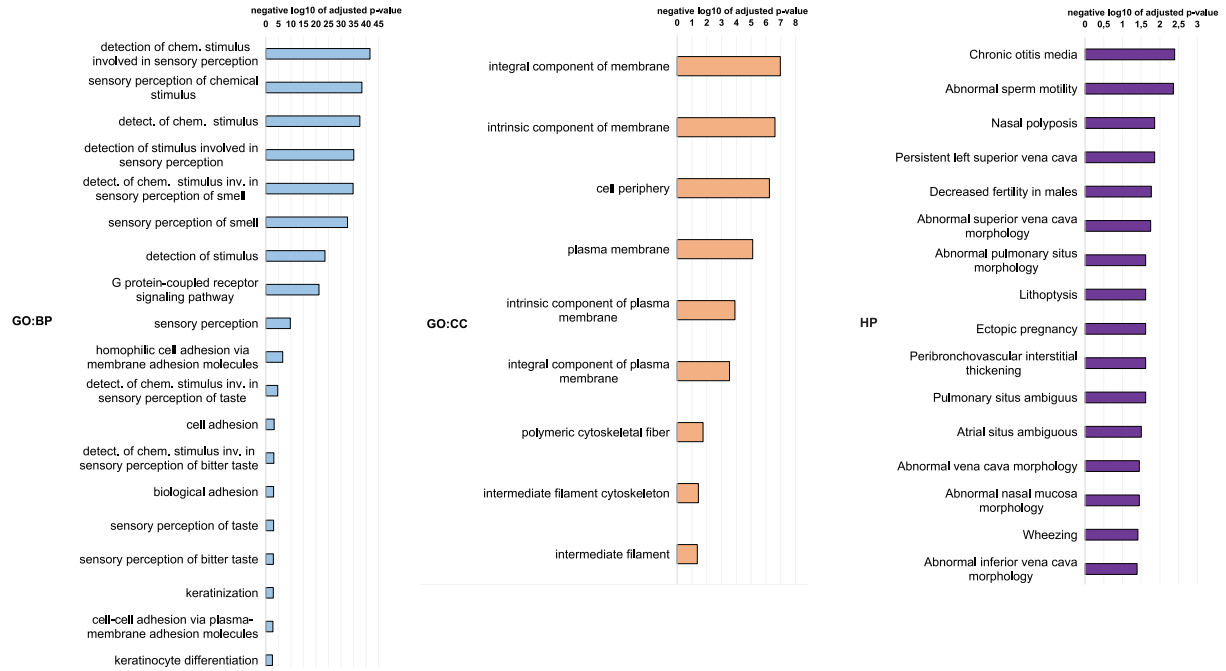


Figure 2F. Gene enrichment analysis of exclusively heterozygous variants. Gene enrichment analysis was performed using the g:Profiler webserver. BP: biological pathways; CC: cellular compartment; HP: human phenotype

2.3.2 Identifying likely functional variants

Inspection showed numerous issues with the initial set of exclusively heterozygous variants, so I adopted multiple strategies to filter out likely artefacts and to highlight those likely to be functional.

First, the original set of 1kG variants was filtered to remove variants where 1kG and gnomAD disagreed for MAF, yielding 167k variants of which 1943 were exclusively heterozygous in at least 41 individuals. gnomAD data has a much larger total count compared to 1kG, increasing the chance for sequencing errors. By tolerating a homozygous count of < 5 in gnomAD data yielded 353 variants. A final set of 313 variants was defined by also removing repeat-prone genes, or genes unusually subject to mutations (including Filaggrin, Mucins and Olfactory receptors). Interestingly, a small number of 33 exclusively heterozygous variants lie in 32 known disease genes[88], but are currently not known to be causative. However, the majority of variants (280) are in genes not currently associated with disease[67]. Several metrics were used to evaluate both the structural and functional impact of these variants on protein function and their likely association with human disease (see Methods 2.2.5). The combination of different metrics gave a performance benchmark similar to predictors such as SIFT[96] or PMUT[97]. As several of the negatives used to train the Bayesian score were also used to train these predictors, their scores were not integrated into my approach.

Exploration of the 1kG variants displays an enrichment of functional impact for exclusively heterozygous variants. In addition, allowing for homozygous counts reduces the enrichment of functional impact (**Fig. 2G**). These observations suggest that a) exclusively heterozygous variants are enriched for functionally disruptive variants and b) these variants are more likely to modify or even disrupt protein function. Of the 313 exclusively heterozygous variants, 108 (33.4%) have a functional impact score ≥11 with a false-discovery rate < 1% and a false positive rate < 5%, also suggesting a substantial enrichment of functionally relevant changes[67]. These 108 genes show little coherence in terms of function, though certain groups stand out. For instance, 12 genes (e.g. FCGR2B, LILRB4, TRADD) are broadly associated with autoimmune diseases, 13 with obesity (e.g. ADRB1, ALPI) and four are associated with ciliary function/ciliopathies (e.g. CROCC, GLIS2). Notably, these are all conditions that might lead to symptoms in single tissues later in life (e.g. eyes or kidneys in many ciliopathies) or might only manifest under certain circumstances (autoimmunity or obesity)[67].

2.3.3 PANK3 Ile301Phe perturbs coenzyme A biosynthesis

The non-synonymous Ile301Phe variant of Pantothenate kinase 3 (PANK3) is seen exclusively heterozygous in 528 1kG (21 %) and 1820 gnomAD (0.7 %) individuals. PANK3 is one of three kinases essential in coenzyme A biosynthesis[98] and is largely preserved across all three kingdoms of life (**Fig. 2H**). Loss-of-function variants in other organisms, for example *S. cerevisiae* or *D. melanogaster,* are not viable[99]. Residue 301 lies in the fumble domain[100] and is buried in the core of the protein (**Fig. 2H**) just under the active site of the enzyme. This position is most often isoleucine, in some cases valine, in homologs of PANK1-3, suggesting that even the seemingly conservative change to phenylalanine would not be tolerated (**Fig. 2H**).

Mutations in the PANK3 paralog PANK2 negatively affect coenzyme A biosynthesis and are associated with Neurodegeneration with Brain Iron Accumulation (NBIA) or Hallervorden-Spatz syndrome. The latter is a recessive neurological disorder. Notably, known variants include PANK2 Ile501Thr which is in the equivalent and conserved position in PANK3 Ile301[101]. Introducing human wild-type PANK3 to PANK2 equivalent knockout ($fbl^{-/-}$) Drosophila can partly rescue the WT phenotype[102] suggesting some equivalency of these close paralogs. The absence of homozygous individuals despite so many heterozygous carriers makes it tempting to suggest that this PANK3 mutation could cause a similar recessive condition[67].

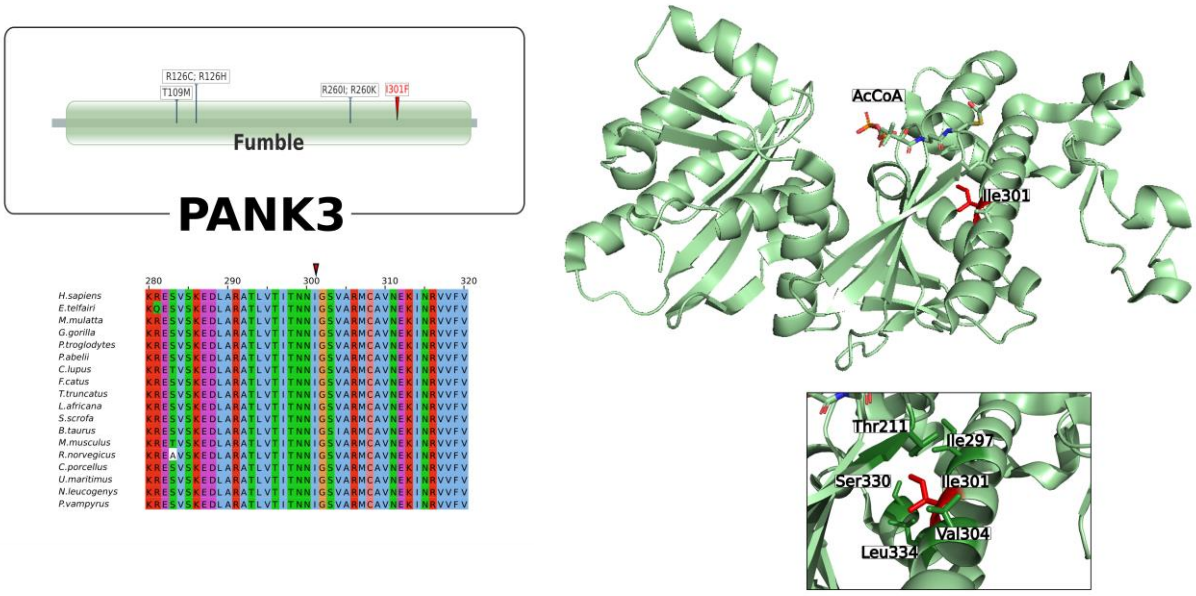


Figure 2H. PANK3 Ile301Phe. Top left: Domain overview of PANK3. Known disease variants are highlighted in black, the position of Ile301Phe in red. Bottom left: Jalview[103] alignment of selected model organisms showing the residues around Ile301. Conserved residues are shown in ClustalX colours. Top right: Zoomed out view of PANK3 (PDB: 5KPR). Bottom right: Zoomed in view of Ile301 and neighbouring residues.

2.3.4 CCDC8 Gln200Leu and its relevance for 3M syndrome

The variant Gln200Leu in coiled-coil domain-containing protein 8 (CCDC8) potentially affects ciliary processes. The variant is seen exclusively heterozygous in 44 (1.8 %) 1kG and in 694 (0.26 %) of gnomAD individuals (though with some homozygous instances). CCDC8 is part of the 3M complex – together with CUL7 and OBSL1 – which is regulating microtubule dynamics as well as maintaining genome integrity[104].

Reports state that mutually exclusive, homozygous or compound heterozygous mutations in these three genes are causative of 3M syndrome[104], an autosomal recessive growth disorder with prenatal growth restriction and the failure of postnatal catch-up, resulting in short stature and skeletal abnormalities[67,105]. This diseases is likely a ciliopathy[106], a group of heterogenous rare diseases, affecting cilia[107]. The difficulty in spotting the phenotype of this and other ciliopathies might also explain why there are some homozygous carriers seen in gnomAD that might suffer from mild symptoms but remain undiagnosed. Residue Gln at position 200 is largely conserved in vertebrates and lies within a short ordered segment[108] (**Fig. 2I**). CCDC8-null mice showed defects in trophoblast motility known to result in complications during pregnancy such as placentation failures or even fetal death[109]. Other known CCDC8 3M syndrome mutations are stop-gains or frameshifts.

Pathogenic variants in CCDC8 are reported to disrupt the binding of ANKRA2, a protein known to recognize a C-terminal motif in CCDC8 (**Fig. 2I**)[110]. Gnl200 lies within a putative WW domain and high-throughput studies suggest phosphorylation events nearby at Tyr197 and Ser202[111] that are thought to mediate interactions with other 3M proteins[109]. It is possible that Gln200Leu might disrupt structure and/or interactions involving this region of CCDC8[67].

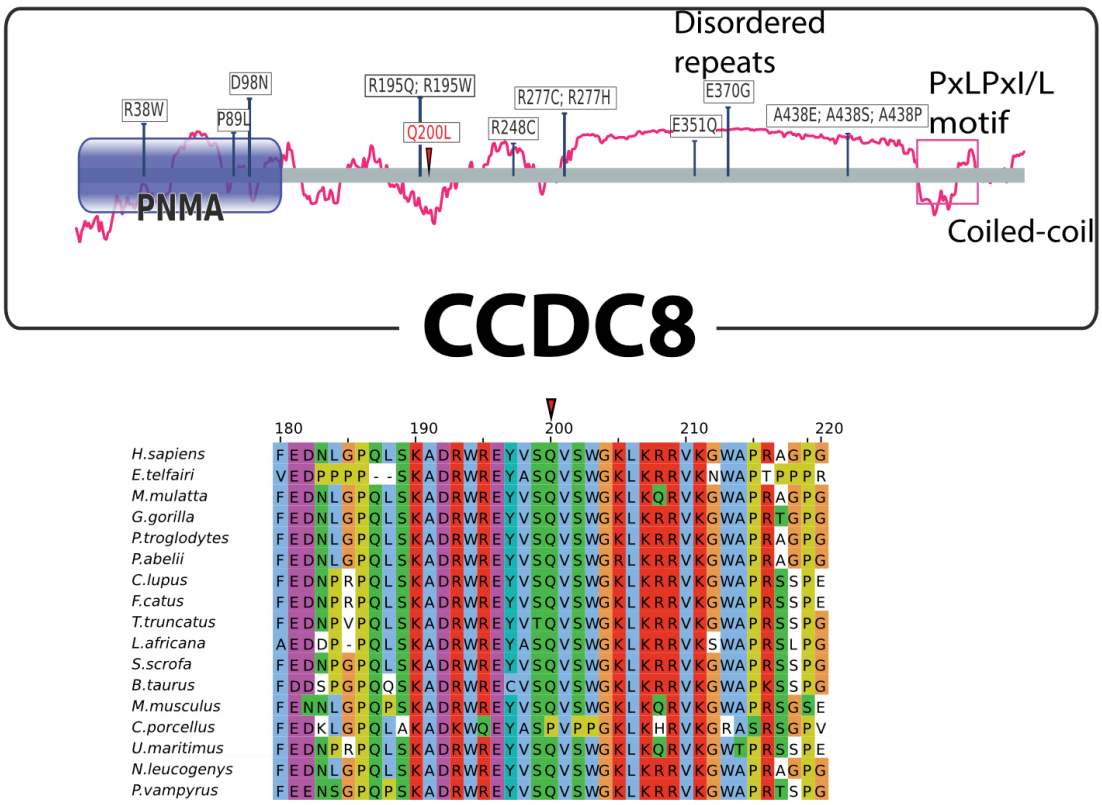


Figure 2I. CCDC8 Gln200Leu. Top: Domain overview of CCDC8 superimposed on top of a IUPred plot of protein disorder[108] (pink). Known disease variants are highlighted in black, the position of Gln200Leu in red. Bottom: Jalview alignment of selected model organisms showing the residues around Gln200. Conserved residues are shown in ClustalX colours.

2.3.5 NLRP12 Asn394Lys – causative of FCAS2?

Another interesting variant is in Asn394Lys in NLRP12, a protein involved in the immune system. This protein is expressed in dendritic cells as well as macrophages[112] and the Asn394Lys variant is seen exclusively heterozygous in 72 (2.9 %) 1kG and 70 (0.03 %) gnomAD individuals. Asn394Lys lies in the NACHT domain of NLRP12 and is highly conserved in homologous proteins (**Fig. 2J**). The protein acts as a negative regulator of several inflammatory pathways[113]. Deletions or frame-shift variants have been detected in NLRP12 and are typically associated with Familial cold autoinflammatory syndrome 2[114], a disease triggered by exposure to cold with symptoms associated to inflammation (i.e. fever, rashes, myalgia and headaches). While there is currently no resolved 3D structure for NLRP12 available, the Alphafold predicted model suggests that Asn394 is in close contact to several neighbouring largely polar sidechains (**Fig. 2J**). One of these

is Arg429 that would not favour contacts with another positively charged Lysine in the variant.

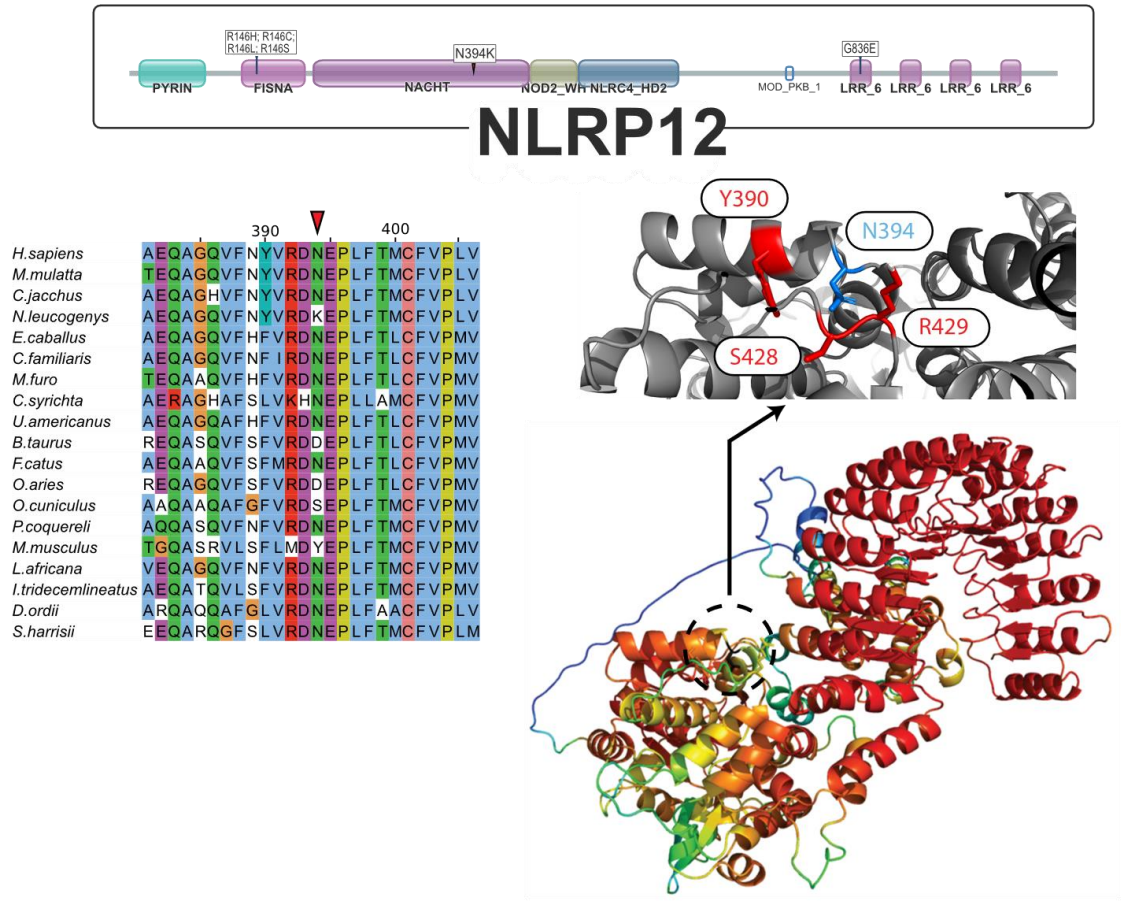


Figure 2J. NLRP12 Asn394Lys. Top: Domain overview of NLRP12. Known disease variants are highlighted in black, the position of Asn394Lys in red. Bottom Left: Jalview alignment of selected model organisms showing the residues around Asn394. Conserved residues are shown in ClustalX colours. Bottom right: Zoomed out view of the Alphafold model for NLRP12 and zoomed in on the putative location of Asn394. Warmer colours indicate higher confidence. Modified from Schmenger et al. 2022.

2.3.6 RHD Tyr311Ser in haemolytic disease

The Tyr311Ser variant in blood group Rh(D) polypeptide (RHD) is exclusively heterozygous in 623 1kG and 1080 gnomAD individuals. RHD is a non-transporting homolog of other transporters, such as RHCG. RHD forms heterotrimers with these other transporters and together they are involved in ammonium transport between erythrocytes and the kidneys or liver[115]. The crystal structure of RHCG[116] shows that the corresponding residue to Tyr311 (Tyr323) lies at the protein-membrane interface, with several

intramolecular hydrophobic contacts to other protein residues (**Fig. 2K**). Sequence conservation analysis demonstrates that Tyr311 is nearly always hydrophobic in wider homologs, and is never a Serine. Interestingly, whether the position is either a Tyr or a Cys seems to depend on whether the protein is RHD or the close paralog RHCE[67].

Serine is disfavoured at membrane interfaces[117] and replacing Tyr311 with Ser could alter the membrane position or the trimer structures[116]. At least 14 variants in RHD have been previously associated with the "weak D antigen"[118] of which 8 are concentrated in a region around position 311 (residues 270-339). It is plausible that two copies of this mildly or fully dysfunctional RHD subunit could lead to a disease phenotype[67], similar to haemolytic disease of fetus and newborn, known to be induced by a mismatch of maternal and fetal RHD antigens[119].

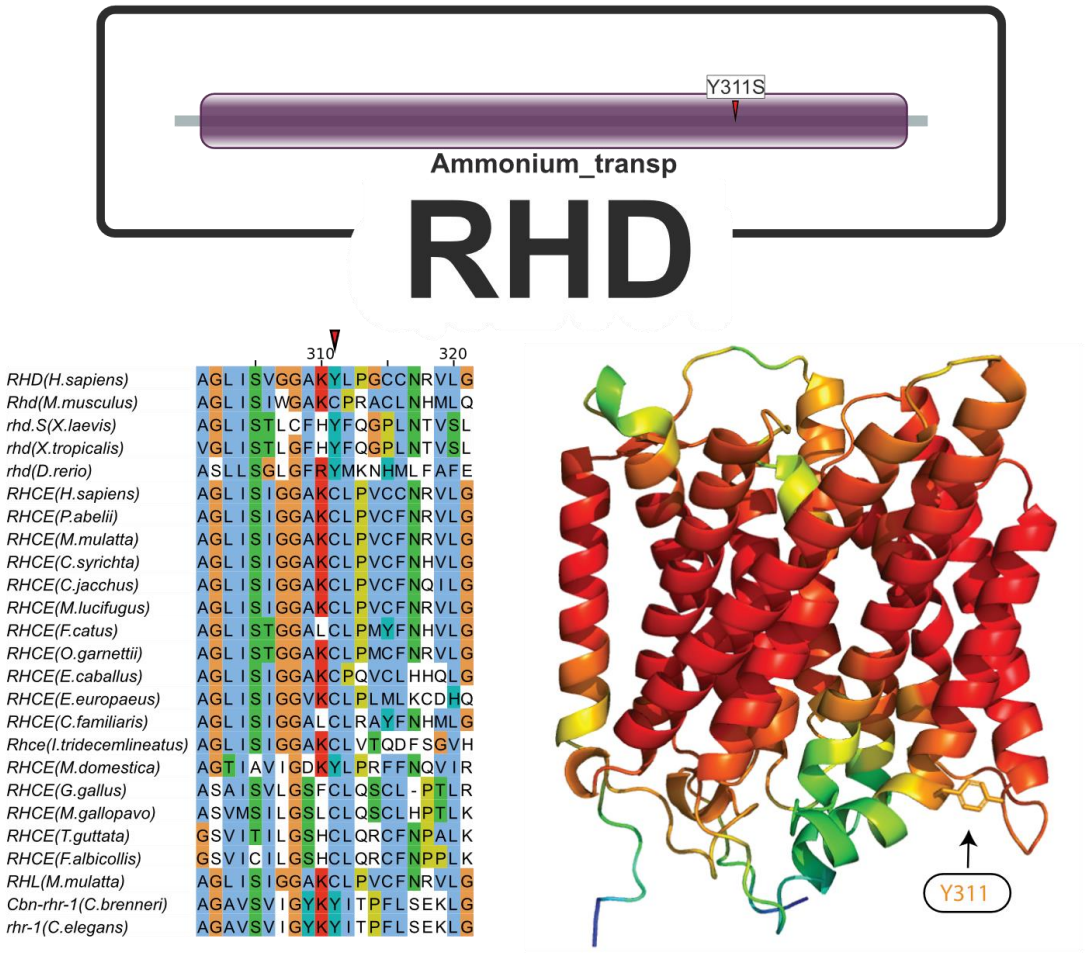


Figure 2K. RHD Tyr311Ser. Top: Domain overview of RHD. Bottom Left: Jalview alignment of genes from selected model organisms showing the residues around Tyr311. Conserved residues are shown in ClustalX colours. Bottom right: View of the Alphafold model for RHD. Warmer colours indicate higher confidence. Modified from Schmenger et al. 2022.

2.3.7 CMA1 His66Arg destroys the active site of a serine proteases

His66Arg from Human mast cell protease 1 (CMA1) is exclusively heterozygous in 42 (1.67%) 1kG individuals. CMA1 is involved in wound healing, inflammation and respiration[120]. The protein converts angiotensin I to angiotensin II, which it does more efficiently than angiotensin-converting-enzyme (ACE)[121], and is a promising drug target for cardiovascular disease[122]. CMA1 is involved in remodelling of the extracellular matrix[123–126], which is crucial during early human development[127–129], including the remodelling of fetal spiral arteries[130]. A preliminary study found that 3.98% (40/1004) of participants treated for atypical eczema and dermatoses were carriers and noted the absence of homozygosity[131]. His66 is part of the catalytic triad in this protease[132] (**Fig. 2L**) thus implying a complete loss of CMA1 activity in homozygous individuals.

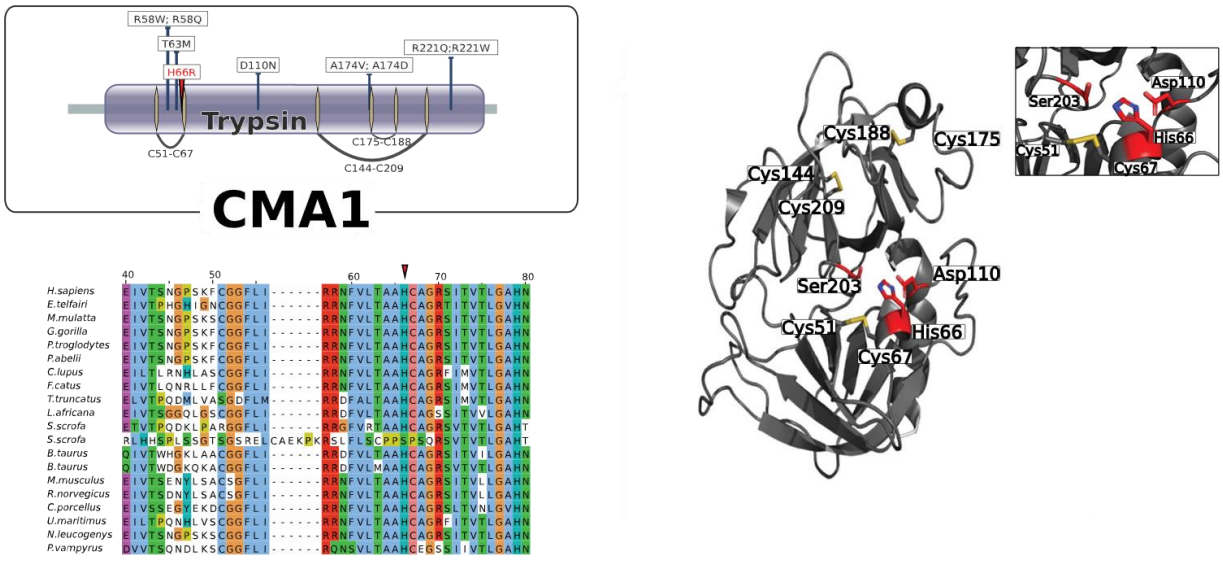


Figure 2L. CMA1 His66Arg. Top left: Domain overview of CMA1. Variants with effects on human health and disease are highlighted in black, the position of His66Arg in red. Bottom left: Jalview alignment of selected model organisms showing the residues around His66. Conserved residues are shown in ClustalX colours. Top right: Zoomed out view of CMA1 (PDB: 2HVX) and zoomed in of the catalytical triad of CMA1.

2.3.8 Conclusion and outlook

The goal of 1kG was to create a background model of human genetic variation. To achieve this, they had to focus on including mostly healthy participants into their program, doing so by visual inspection and using a questionnaire (this also applies to most similar sequencing efforts, summed up in gnomAD). It cannot be ruled out, that homozygous

variants are lacking in the 1kG dataset (and in gnomAD) not only due to homozygous carriers not being viable, but also because homozygous carriers would show clear signs of illness or disabilities.

For certain conditions (e.g. ciliopathies only affecting certain tissues in later life), the disorder might simply not be known or not being detected, which might be true for cases such as CCDC8, perhaps causative of difficult to diagnose ciliopathies, and where there are possibly a small number of homozygous individuals. Equally possible is that the homozygous variants are so severe that embryos are not viable. This possibility is easier to argue for variants such as those in PANK3, CROCC or RHD that have more than 1000 heterozygous counts despite zero homozygous[67].

Population genetics suggests that deleterious and harmful variants will eventually be removed from a population by a process called purifying selection. If homozygous variant carriers are indeed not viable or otherwise affected by a disease, then consequently such carriers would show a lower reproductive fitness. Obviously, the question presents itself on why these variants are still present in modern humans. It has been argued that humans are indeed undergoing purifying selection[133–135] and certain recessive diseases are probably examples[136–138] of those in the process of being removed (e.g. SMA1[139,140], IMD31B[141,142], NSHPT[143,144]). This is clearly not true for the majority of these examples, though several of them (48/353, 12.2%) show enrichment in human sub-populations (Table S1). For instance, NLRP12 Asn394Lys is twice as frequent in American, East-Asian and European populations as African populations. PANK3 Ile301Phe has a frequency of 11-14% in non-African populations that is 2-3 times that of African populations (5%)[67]. These variants were possibly enriched in the original migratory populations.

These results demonstrate the power of exploiting putatively healthy genomes to identify new insights into molecular protein function, and how these effects translate into human health and disease. Previous studies and their findings support the idea that such databases can be used to explore similar questions[145–147]. While datasets similar to 1kG continue to be generated the approach demonstrated here can be used to exploit the underutilized feature of heterozygous vs. homozygous genotypes, aiding our understanding of protein function an assisting in finding new methods for diagnosis and treatment of human disease.

Chapter III: Hereditary Disease Variants and 3D Distance Based Functional Clustering

3.1 Introduction

Between 5 to 10 % of all cancers carry an inherited genetic signature[148–150]. Despite the fact that medical observations of exotic familial phenotypes were known for several centuries[151] geneticists started to consider genetic factors as causes for familial cancer only in the last two generations[152].

According to the two-hit hypothesis[153] many cancers require two hits for oncogenesis, ultimately losing both WT copies of the respective gene. This hypothesis originally suggested a model for retinoblastoma formation, where inherited Retinoblastoma protein (Rb) mutations are seen as the first hit, and acquired somatic mutations in the same gene are seen as the second. This concept is now applied to the inactivation of tumour suppressor genes and activation of oncogenes, with TP53 being a prominent example for the former[154,155], where mutations are causative of Li-Fraumeni syndrome[156]. Early-onset colorectal cancer[73,157] is just one instance showing that inherited cancer mutations show a much higher penetrance and display an earlier onset than their somatic counterparts.

Notably, advances in sequencing technologies have led to better understanding of genetic diseases in general. However, the contextual differences between hereditary and somatic cancer are still poorly understood, and therefore the focus of much current research[158,159]. For example, hereditary breast or ovarian cancer, as well as lynch syndrome, are common maladies[160] involving an inherited component, and differentiation between somatic or germline variants as the origin of a given cancer already has implications for patient treatment[161,162].

Most human gene sequences (and therefore protein sequences) are known[163], and while experimental 3D structures are not yet available for every protein, the thousands of 3D structures being released into PDB annually (**Fig. 3A**)[164] can provide accurate structures by homology. This concept is used in many prediction tools, such as Mechismo, where the protein sequence of interest is mapped to homologous structures, if otherwise no structural information is available. A lack of experimental structural information is also often indicative of a general lack of understanding for a given proteins function (except historically for proteins with difficult structures to determine, such as

membrane proteins[165]), or at least indicates a higher uncertainty of any given assumptions about the protein of interest.

A first step to better understand variation in any gene (and consequently their effect on protein function) involves gathering existing knowledge presented from the literature. This process is often time consuming and artificially challenging due to differing naming conventions (also see Chapter 1) or generally hard-to-access data formats.

The aim of this particular project is to better understand hereditary cancer and to find novel ways to address the pathogenicity of given variants. What are the differences between somatic and germline variants in cancer? Moreover, is there a way to automate some aspects of literature research in meaningful ways?

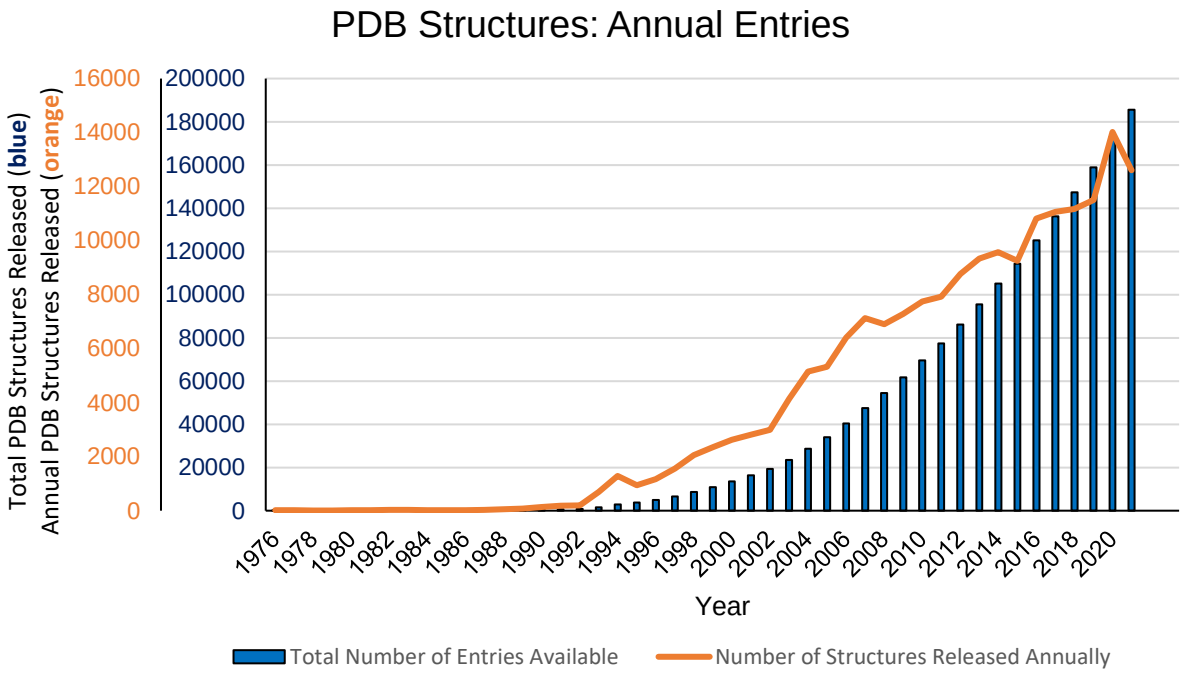


Figure 3A. Annual PDB structures releases. In 2022 53k (of 185k, 29 %) structure entries were of human proteins.

3.2 Material and Methods

3.2.1 Datasets of somatic and hereditary disease variants

I retrieved hereditary disease variants, many of which were variants of familial cancer syndromes, directly from Uniprot, yielding a set of 61 genes and 811 variants.

Somatic cancer variants where retrieved from COSMIC[16]. The PDB[164] was queried for each protein present in the dataset and I kept only proteins where at least 50% of its amino acid sequence was covered by PDB 3D structures, yielding a final set of 243 proteins and 12k variants. I used these structures to calculate residue accessible surface area using FreeSASA[166]. Relative accessible surface area was calculated by dividing these values by those of a G-X-G tripeptide[167]. I defined buried residues as those with relative accessible surface area < 25 %[167].

3.2.2 Analysis of hit/avoided protein domains in somatic and hereditary disease

For all variants, I identified protein domains using Pfam[100] and interaction interfaces determined using Mechismo[55]. Since the Hereditary and Somatic variants are fundamentally different in how they were described in the data sources, I treated them differently in certain statistical and/or filtering steps.

For Hereditary variants, the expected frequency of mutations in these domains was calculated by randomizing 10k variants across the 61 genes present in the dataset. I then compared these expected frequencies to the observed frequencies. I subjected genes deemed interesting to clustering based on intramolecular distances (see below).

For somatic variants, I considered only the 14 most represented cancer types (breast, central nervous system, haematopoietic and lymphoid tissue, kidney, large intestine, liver, lung, oesophagus, pancreas, prostate, skin, small intestine, stomach and thyroid). Within the context of a single protein it was assumed that each residue of the same amino acid (e.g. all positions with alanine) has roughly the same likelihood to be mutated regardless of protein position. For each of the 20 amino acids the total number of variants per amino acid would be summed (e.g. all alanine residues have a total variant count of X). A binomial test was performed (n = total number of variants affecting a specific amino acid, successes x = number of variants for a specific position and probability p = mutation frequency retrieved from gnomAD[14]: A: 0.08; C: 0.03; D: 0.03; E: 0.04; F: 0.02; G: 0.05; H: 0.04; I: 0.06; K: 0.04; L: 0.06; M: 0.04; N: 0.04; P: 0.05; Q: 0.04; R: 0.1; S: 0.08; T:

0.08; V: 0.09; W: 0.01; Y: 0.02). Variants with a p-value < 0.05 would be considered *significantly enriched* if the respective variant count would lie above the average count observed for this amino acid type and within the respective protein, or as *significantly underrepresented* if the respective variant count would lie below the average.

.

3.2.3 Clustering based on intramolecular distances and the detection of functional hotspots (CONNECTOR)

I adopted several strategies to extract or define functional information from various sources. Available positional information for a protein is selected from UniProt[91], Pfam[100], Phosphosite[111], ClinVar[15] or COSMIC[16]. Sets of homologous proteins (for each human protein) are selected from pre-computed alignments generated through OrthoFinder[79]. These alignments are used to define the degree of sequence conservation for each position in the protein of interest.

I defined functional residues as positions in the protein of interest if they were associated with a disease or known protein function in the protein of interest or any homologous protein.

For each protein of interest, I defined one or more structures. I used exact structures from the PDB when available, otherwise I used those predicted by Alphafold[83]. Intramolecular (using C_α for glycine and C_β for all other amino acids) atomic distances between all *functional-residues* are measured and the average is kept. To avoid measuring distances between sidechains on opposing sides of the backbone, that also point away from each other, I compared two distances: residue-1-Cα vs. residue-2-Cβ to the default measurement of residue-1-Cβ vs. residue-2-Cβ (ignoring this comparison for Gly) and kept the shorter distance value. This captures extreme cases where sidechains of a given pair of amino acids are very unlikely to ever interact. Atomic distances ≤ 8 Å are then subjected to a random walk algorithm[168] to define clusters of putatively functional residues, based on intramolecular distances (**Fig. 3B**).

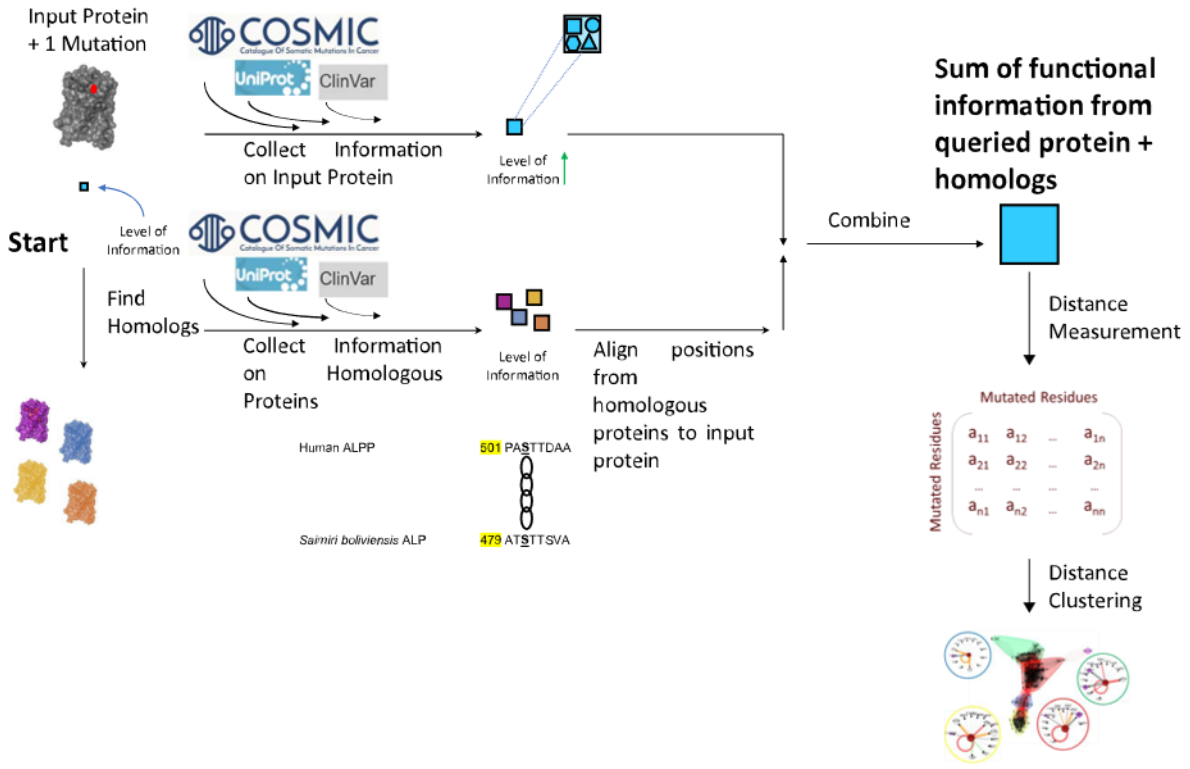


Figure 3B. Functional Clustering Workflow.

3.3 Results and Discussion

3.3.1 Hereditary and somatic cancer variants are often buried

It is often hypothesized that the majority of pathogenic variants are buried within a protein core[34,47], where changes are most likely to perturb the protein structure as a whole. These changes might destabilize the protein, decreasing its half-life and leading overall to less available cellular protein[34,169]. Alternatively, they might change (but not destabilize) the structure by placing residues in environments they disfavour (e.g. charged residues in the protein core or hydrophobic residues on the surface)[170].

The average relative surface accessible area (rel. SASA) for somatic variants from 14 different cancer types is 0.19 which supports the notion that a majority of these variants are buried (rel. SASA < 0.25). Variants sitting at putative interfaces, as defined by Mechismo, generally show higher rel. SASA values with an average median across all cancer types of 0.327 vs. 0.135 for those not at interfaces (**Table 1** and **Fig. 3C**, **A**). This difference is less pronounced when considering hereditary disease variants, with values 0.08 for all, 0.11 for interfaces and 0.05 for non-interface variants (**Table 1** and **Fig. 3C**, **B**), overall suggesting a greater tendency to be buried regardless of whether a residue is close to an interface or not.

Table 1. rel. SASA values for 12k somatic cancer variants and 344 hereditary disease variants

	Interface		Non-Interface	
	rel. SASA (median)	n	rel. SASA (median)	n
breast	0.31	43	0.01	1.637
central nervous system	0.33	14	0.00	623
haematopoietic & lymphoid tissue	0.23	53	0.04	2.077
kidney	0.14	12	0.21	309
large intestine	0.26	62	0.00	1.761
liver	0.90	1	0.43	137
lung	0.34	70	0.00	2.407
oesophagus	0.12	2	0.37	134
pancreas	0.21	9	0.30	234
prostate	0.13	15	0.11	661
skin	0.40	40	0.00	1.138
small intestine	0.42	4	0.28	154
stomach	0.41	6	0.04	265
thyroid	0.37	14	0.10	580
Average	***0.327***		***0.135***	
Hereditary	0.11	119	0.05	225

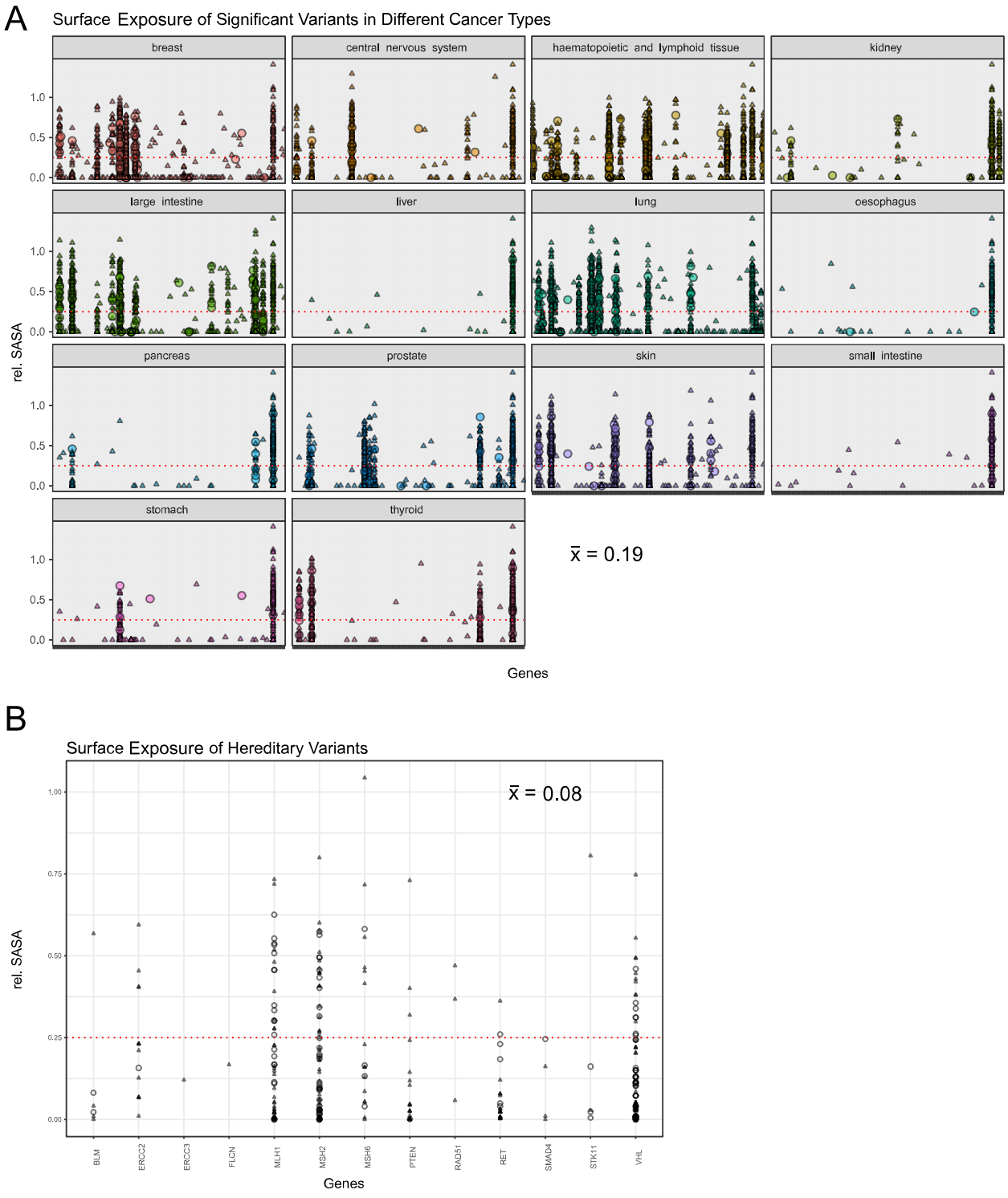


Figure 3C. Surface exposure of somatic and hereditary disease variants. A. Rel. SASA values for 14 different somatic cancer types. Red-dashed line indicates the 0.25 cutoff. Values below are considered to be buried, values above are considered to be exposed. Putative interface variants are plotted as circles. **B.** As described in A showing rel. SASA values for hereditary disease variants.

3.3.2 Somatic cancer variants and their blind spot for the *P53_tetramer* domain

The p53 tumour suppressor protein is amongst the most frequently mutated proteins in cancer and one of the most well understood. Generally, p53 is thought of as a tumour suppressor as missense variants in p53 diminish its activity and contribute to oncogenesis[171], while some rarer instances even attribute oncogenic abilities to p53[172,173].

p53 acts as a transcription factor, affecting hundreds of genes, which explains why the vast majority of somatic cancer variants lie within the DNA-binding domain of the protein[171]. Also important for DNA-binding is the tetramerization domain that facilitates the dimer formation of p53 dimers[174,175]. Interestingly, the p53 tetramerization domain shows a depletion of variants in some cancers, including malignancies of the central nervous system, haematopoietic & lymphoid tissue, large intestine, liver, lung, oesophagus, pancreas, prostate, skin and stomach, but not in tissues such as breast, kidney, small intestine or thyroid (**Fig. 3D**). Tetramerization between WT and mutant p53 has been shown to typically lead to normal p53 WT phenotype[175]. It is then plausible to assume that a wide range of mutations affecting the p53 tetramerization domain might alter or modify the transcription factor capabilities of the p53 oligomer. However, they fail to diminish its activity to an extent that is required for a loss of its tumour suppressive activities, hence why variants in the tetramerization domain are less likely to be seen in cancer patients. The observation that such variants are still present in somatic cancer indicates, of course, that even such alterations can be enough to initiate tumour progression in some, but not all, cases. Putative reasons for this perceived selectivity could be tissue-dependent differences in DNA methylation and consecutively DNA accessibility[176], where slight changes in p35 tetramerization are simply not cancerous enough to translate into enhanced gene transcription.

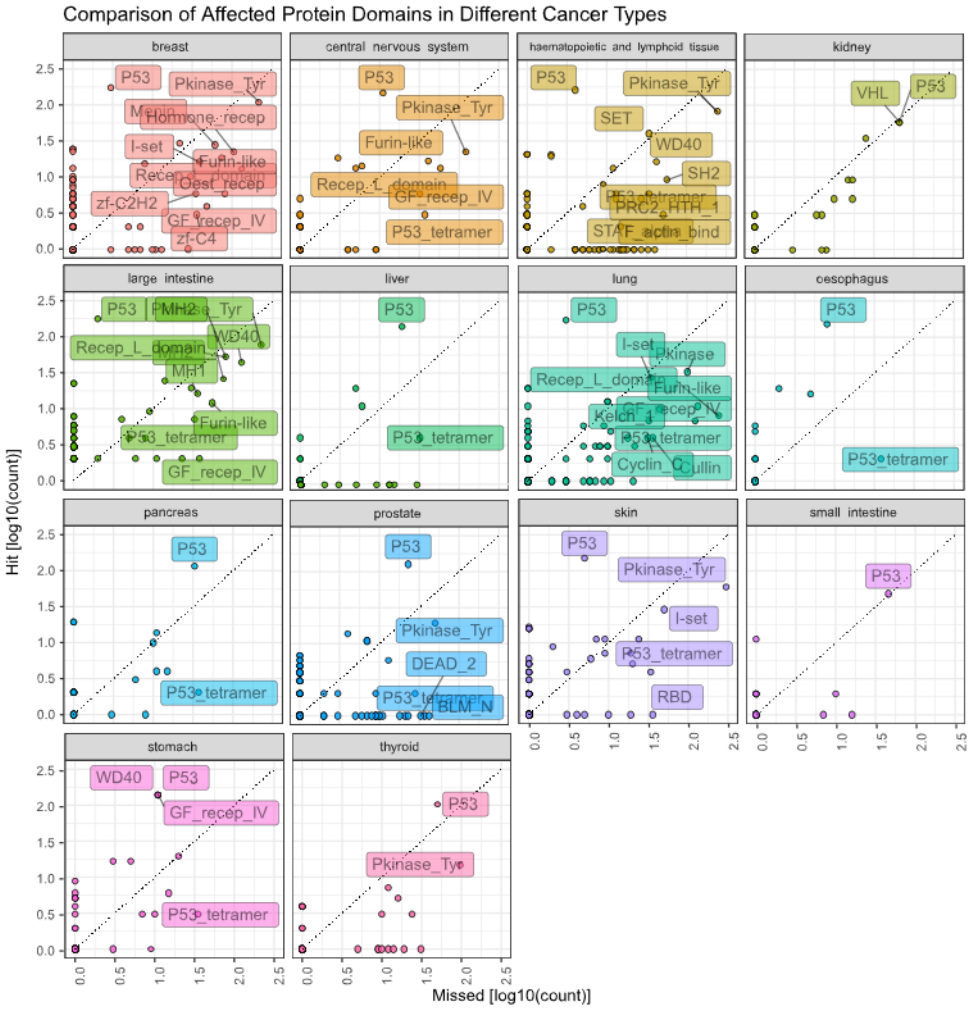


Figure 3D. The P53 tetramerization domain is often left untouched in somatic cancer. The ratio of how many times given protein domains were significantly hit or depleted/missed in 14 different cancer types is shown.

3.3.3 Perturbation of DNA repair pathways as a major contributor to hereditary cancer

A similar analysis as in 3.3.2. was performed for hereditary variants (see Methods 3.2.2). Variants related to DNA repair were most often hit, but not domains in protein kinases (**Fig. 3E**), which are often oncogenic drivers in somatic cancer[177].

Amongst the domains that are significantly more frequently mutated in the hereditary dataset are domains related to mismatch repair (*DNA_mis_repair*, *MutS* domains), as well as VHL domains. These hits are certainly due to an overabundance of variants in either VHL or MSH2 (23.4 % of variants in the dataset), but nonetheless

suggest that interference with DNA repair is an important contributor to hereditary disease, and hereditary cancer in particular. While humans possess a whole battery of DNA repair genes, with some redundancies[178], it is plausible to assume that even slight alterations in such central pathways will be causative of diseases[179], and interference with DNA repair starting at birth would explain the typically early onset of hereditary cancers[157].

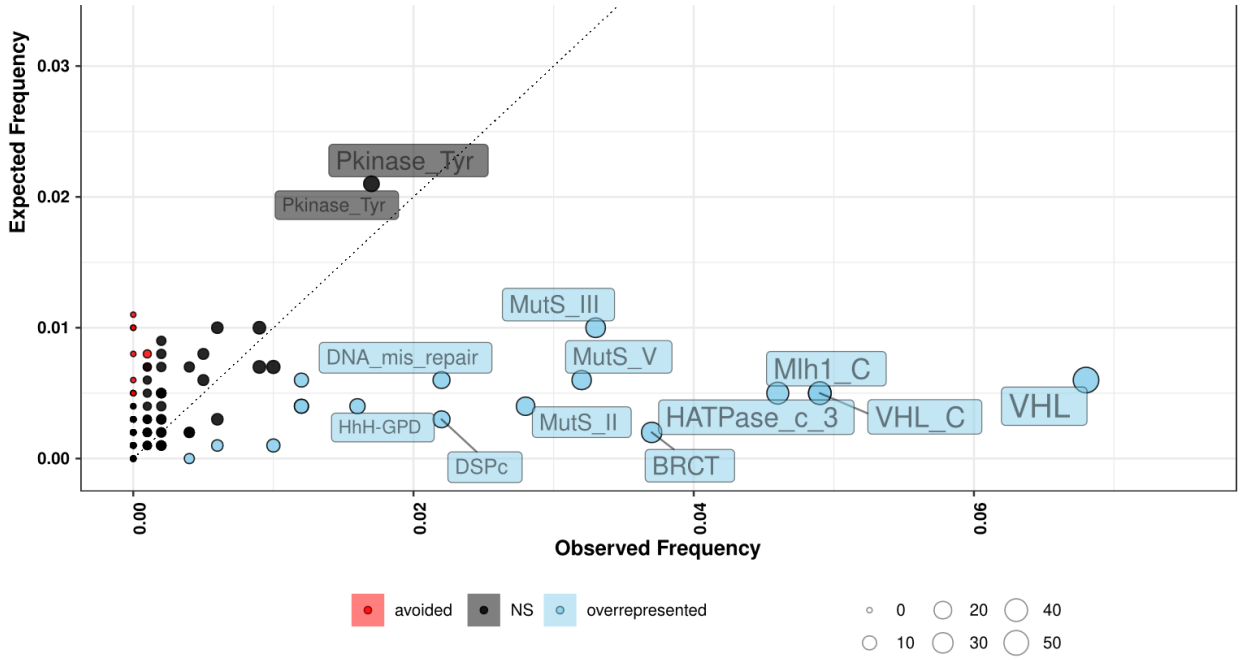


Figure 3E. Perturbed domains in hereditary cancer. The expected frequency of mutation was calculated as described in 3.2.2. Domains more often mutated than expected are highlighted in blue, domains less often mutated are highlighted in red. Dot size indicates the number of variants affecting a given protein domain.

3.3.4 Distance-based clustering reveals functional groups

Perhaps the hereditary disease variants dataset was both too small and too biased towards certain cancer types to derive novel hypotheses, and proteins present in the dataset are involved in diseases that are too heterogeneous (for example breast cancer[180]) to compare them to a subset of COSMIC. It then becomes evident that for many hereditary disease variants their mechanism of action cannot be easily deciphered (also see Chapter I & II), and that literature research is always required to better understand putative consequences of missense variants.

Recently developed tools have approached the need of combining available information to predict putative PPI consequences (Mechismo[55]) or by summarizing available phenotypic and functional knowledge (Mechnetor[46]). However, none of them give a truly holistic view on a given variant of interest by putting the variant in context to what is already known. While a new version of Mechismo (termed *MechismoX*, unpublished) is currently under development to address this issue and to give a more holistic perspective, the here presented clustering based on intramolecular distances can detect functional hotspots (see Methods 3.2.3) and assist in illuminating putative consequences of non-synonymous variation.

Applying this approach to a test dataset of benign and disease-causing mutations (also see 2.2.5 and **Fig. 3F**, **A** and **B**) shows that proteins with disease-causing variants are more likely to produce clustered data, and that variants of interest are more often part of a functional cluster if the mutation is disease-causing ($p < 1x10^{-22}$, **Fig. 3F**, **C** and **D**). While this approach was developed as an exploratory tool it also shows acceptable predictive power with a receiver operating characteristic curve AUC = 0.705 (**Fig. 3F**, **E**).

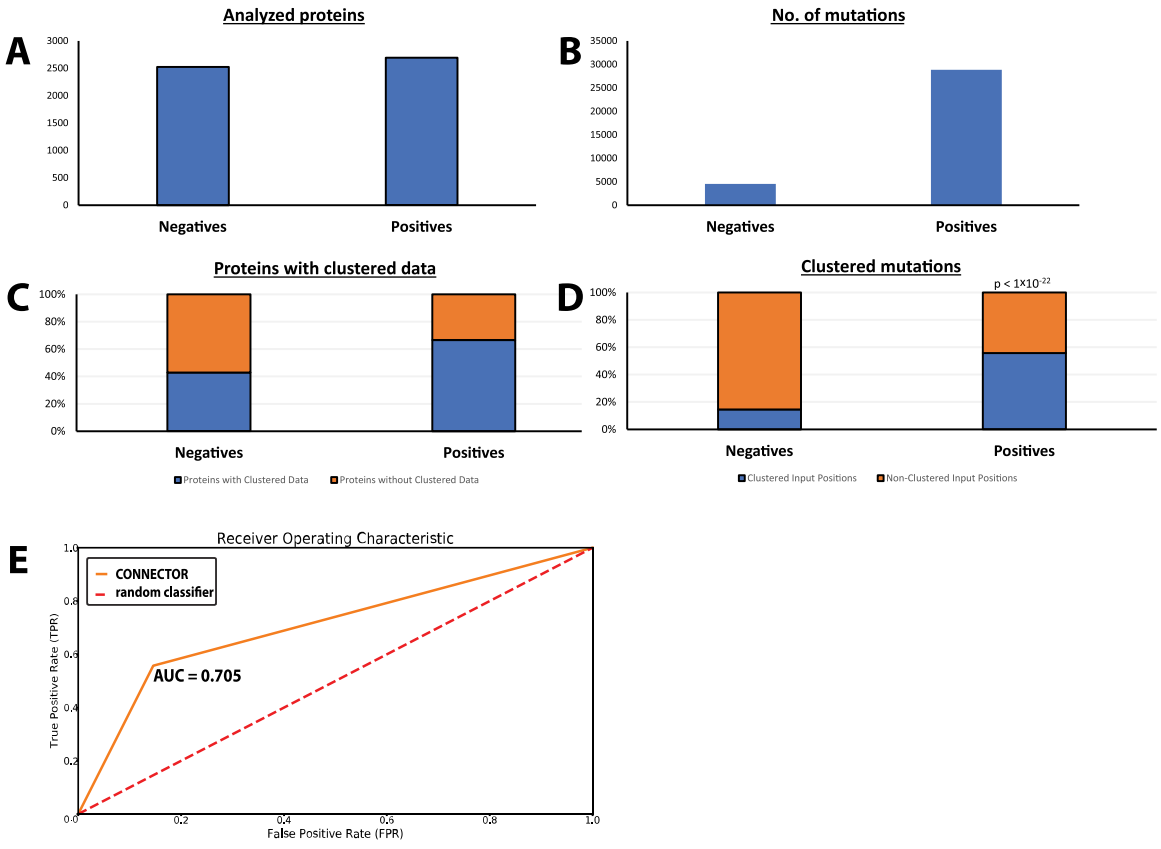


Figure 3F. A. Dataset sizes. Proteins were taken from Humsavar and benign variants were confirmed if an appreciable number of homozygotes were seen in gnomAD (also see chapter II).

B. Mutations in each dataset. Disease causing variants were more frequent. **C. Proteins with clustered data.** The positive dataset yielded more proteins with functionally clustered data. **D. Clustered mutations.** Variants that are disease causing are significantly more frequently part of functional clusters than their benign counterparts. **E. Receiver Operating Characteristic**. The exploratory version of CONNECTOR displays a reasonable predictive power with AUC = 0.705.

To highlight the usefulness of this approach three relevant cases will be discussed.

Case 1: von Hippel-Lindau factor

The first case is the tumour suppressor protein[181–183] von Hippel-Lindau factor (VHL). VHL is part of the VCB complex that consists of VHL, Elongin B, Elongin C and CUL2[184]. The VCB complex is acting as a ubiquitin-ligase E3 and degrades hypoxia-inducible factor (HIF)[184] under normoxic conditions. However, perturbations of VHL or the VCB complex as a whole give rise to von Hippel-Lindau disease (VHLD). Carriers of this disease are susceptible to the formation of primary tumours in many tissues, including the central nervous system, kidney, pancreas or the reproductive systems[185,186].

I subjected VHL mutations to distance-based clustering (see 3.2.4, **Fig. 3G**, **A**). This analysis yielded functional clusters mainly based on known disease variants from Uniprot (**Fig. 3G**, **A**, coloured spheres). Subjecting residues within these clusters to the Mechismo algorithm clearly identified the correct interfaces according to published 3D structures of the VCB complex. For example, the dataset of 811 hereditary disease variants contained several VHL variants affecting Ser111 (Ser111Arg, Ser111Cys, Ser111Asn) which are all found in VHLD. Not only is Ser111 a putative phosphosite[111], it is also neighbouring residues that are part of a functional cluster involved in the interface between VHL and HIFs (**Fig. 3G**, **A,** green coloured spheres). Indeed, phosphorylation of VHL Ser111 has been shown to affect p53 in cell context specific ways through HIF signalling[187–189]. Ser111 likely has very different mechanistic consequences for VHL than a variant such as Glu186Lys, located between residues of a functional cluster participating in the VHL-Elongin C interface (**Fig. 3G**, **A**, yellow coloured spheres), where a perturbation of that interface and hence a perturbation of the VCB complex is a more likely explanation.

VHL perturbations can have a large variety of mechanistic consequences due to the involvement of VHL in the VCB complex. Even though these different mechanistic

paths may lead to similar outcomes, the here presented approach can help in uncovering mechanistic details and differences on how the outcome is reached.

Case 2: Hypophosphatasia and ALPP Ser244Gly

This approach is also applicable on cases from Chapter II. One such example is the variant Ser244Gly in placental alkaline phosphatase (ALPP), which is exclusively heterozygous in 1kG participants as well as in gnomAD (also see Chapter II). ALPP is an extracellular, membrane attached enzyme that hydrolyses phosphate monoesters, and deficiencies cause hypophosphatasia (HOPS), a hereditary metabolic disorder involving seizures and skeletal hypomineralization, often already evident in children[190]. Altered activity and abundance of ALPP has been suggested as a marker for preterm delivery and placental insufficiency[191] and several ALPP polymorphisms have been shown to influence the outcome of *in vitro* fertilizations[192]. Ser244 lies in a loop at the entrance to the active site of the enzyme (**Fig. 3G**, **B** left). The increased backbone flexibility introduced by a Glycine at this loop could be disruptive of the overall enzyme structure or possibly alter substrate recognition. The clustering analysis shows that residues in the vicinity of Ser244 correspond to residues in tissue-nonspecific ALP (**Fig. 3G**, **B** centre) that are known to strongly influence the enzyme activity (**Fig. 3G**, **B** right) and indicate an involvement in the formation of HOPS.

While previously little was known about the Ser244Gly variant in ALPP, this analysis could collect available information from a homologous protein, map the information to ALPP and therefore assists in evaluating potential outcomes of the Ser244Gly variant. This approach is even more powerful when it is mixed with the naïve Bayesian approach from Chapter II, where ALPP Ser244Gly also scores a reasonable 9.31 on its own, but lacks a mechanistic explanation.

Case 3: FGFR3 Val507Met

The missense variant Val507Met in Fibroblast growth factor receptor 3 (FGFR3) was not present in the dataset of hereditary disease variants. Nonetheless it was submitted as a variant of uncertain significance in ClinVar and hence subjected to the functional clustering approach.

FGFR3 is a tyrosine kinase and plays a role in cell proliferation and apoptosis[193]. The protein is well conserved and Val507 is never a Met (with singular instances of Val to Ala or Leu). Perturbations in FGFR3 are found in cancer[193–195] as well as in Thanatophoric Dysplasia Type I (TD1), a severe short-limb dwarfism syndrome that is usually lethal within the first year of life[196,197]. An accepted explanation for this disease phenotype is a constitutively active FGFR3, as several variants causative of TD1 introduce additional cysteines on the extracellular part of FGFR3, putatively assisting homodimerization even in absence of any ligand, leading to the activation of downstream pathways[198,199]. According to the clinical submitter the Val507Met variant was found in a patient suffering from TD1.

While the Val507Met lies within the kinase domain of FGFR3 it became evident immediately that this residue is right next to Lys508 (**Fig. 3G**, **C**), required for ATP binding. The bulkiness of valine (due to it being C_β-branched) might limit the ATP binding capabilities of the neighbouring lysine in FGFR3 WT. However, the introduction of the similar methionine, another hydrophobic amino acid that contains a sulphur atom, might make ATP binding easier, as the conformation restriction due to the bulkiness of valine is removed. While methionine is still a fairly non-reactive amino acid, due to its sulphur atom being connected to a methyl group and not to a hydrogen atom as in cysteine, it is still possible for methionine to assist in binding to metals[200]. Such metal binding could further assist the ATP binding of FGFR3, as soluble ATP is typically found as ATP-Mg[201], hence lowering the threshold for FGFR3 activation. It is thus plausible that also the Val507Met variant could cause TD1.

While the ATP binding site Lys508 might be discovered quickly by experienced analysts this method could very rapidly deliver and contextualize this knowledge to researchers lacking specific knowledge or time.

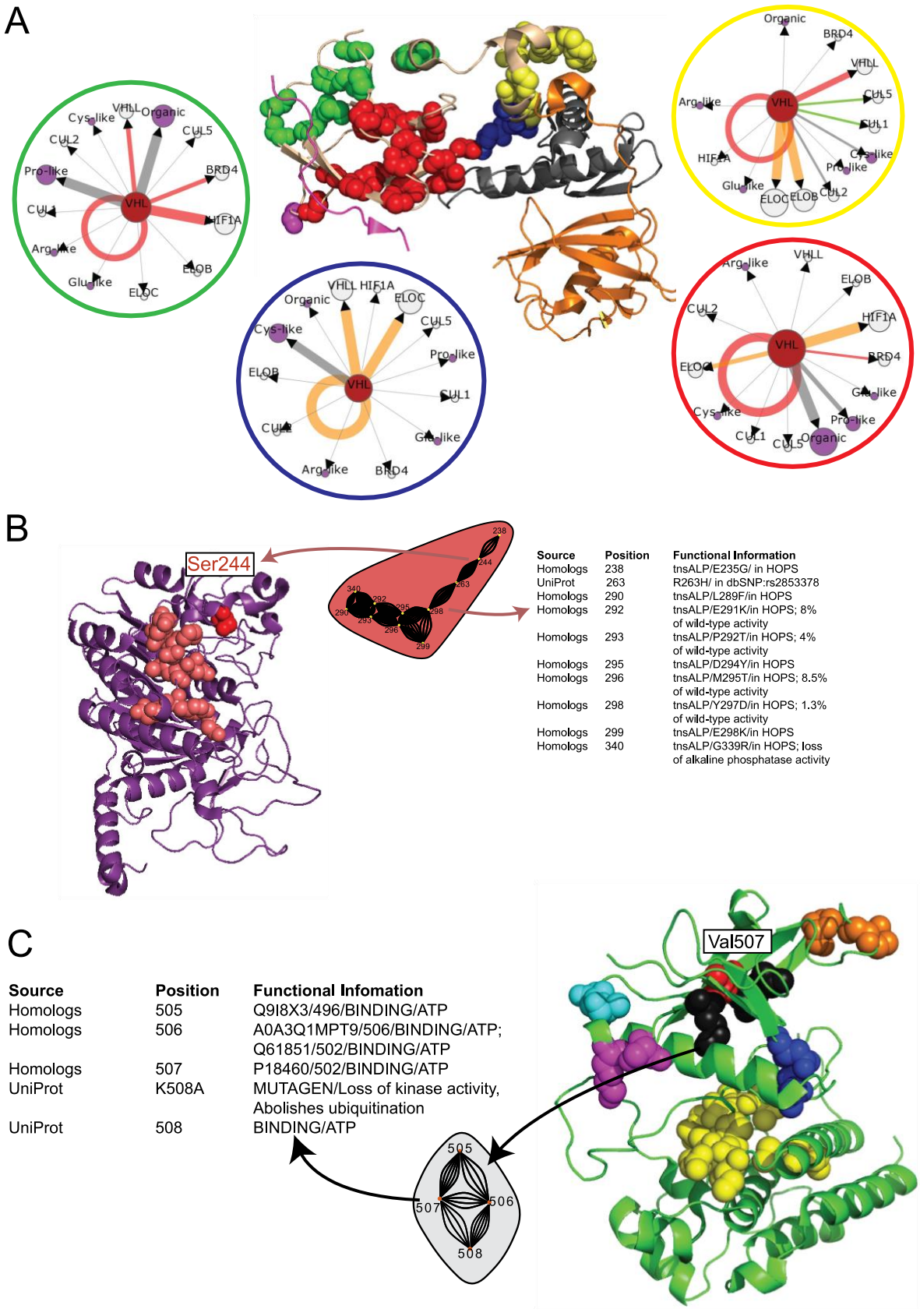


Figure 3G. A. VHL functional clusters. Centre: Structure of VHL (wheat, PDB: 6bvb) in complex with HIF-2α (pink), Elongin-B (grey) and Elongin-C (orange). Functional clusters of VHL are

indicated in different colours (red, green, blue, yellow). For each cluster a visual representation of Mechismo predictions are shown. These show perturbed interactions between VHL and putative interactors, highlighted by coloured edges (orange = enabling & disabling effects, red = disabling, green = enabling). **B. ALPP Ser244Gly.** Left: Protein structure (PDB: 1ZEB) of ALPP. Ser244 (red) and corresponding cluster residues (salmon) are highlighted. Centre: Functional clusters calculated for ALPP. Right: Functional information for residues sharing the same cluster as Ser244. **C**. **FGFR Val507Met**. As in B. Functional clusters of FGFR3 (PDB: 6pnx) are highlighted. Val507 (red) and corresponding cluster residues (black) are shown.

3.3.5 Conclusion and outlook

The analysis of hereditary disease variants shows that there is a focus on perturbation of DNA repair domain positions. However, just as somatic cancer can hardly be analysed or understood as a single disease, it became apparent that the same is clearly true for hereditary (cancer) variants.

Assessing the severity of missense variants usually requires extensive literature research. Many data points are readily available and can be used to arrive at guided conclusions on the effect of a given variant, or on the mechanistic consequences of the variant. However, this data is often not easily accessible. The distance-based functional clustering presented here gathers a variety of information about the protein of interest and about homologous proteins. This information is then used to cluster residues based on intramolecular distances. The approach can also be extended to include other functional information.

Currently, this method retrieves the data and computes the results within ~ 1 minute after inputting a series of variants. This processing time can be reduced by pre-computing several aspects of this approach, and by accessing some of the currently required online databases locally. Future work on this approach will focus on benchmarking and improving internal parameters, as well as including more data points. The initial ROC AUC value of 0.705 is promising, since little work has yet been done on optimizing this metric. It is therefore likely that work focusing on clustering disease-causing variants more effectively will enhance the predictive power of this approach further. For example, using only Alphafold 3D structures and setting a minimum cluster size already improves the AUC to 0.843 (+19.6%). One reason for this is the more homogenous methodology with which the underlying 3D structures were generated. In

contrast, experimental PDB structures of a single protein will often differ in resolution, precision and coverage of the target protein. A next step could be to make this tool available via a web app, for instance using R shiny (see **Fig. 3H** for a possible feature overview).

Integrating this method to the data points calculated and described in Chapter II would give a more complete view of variant consequences, both biochemically (for the respective variant) and by putting them in context (see ALPP), hence creating a powerful tool for studying variant mechanism.

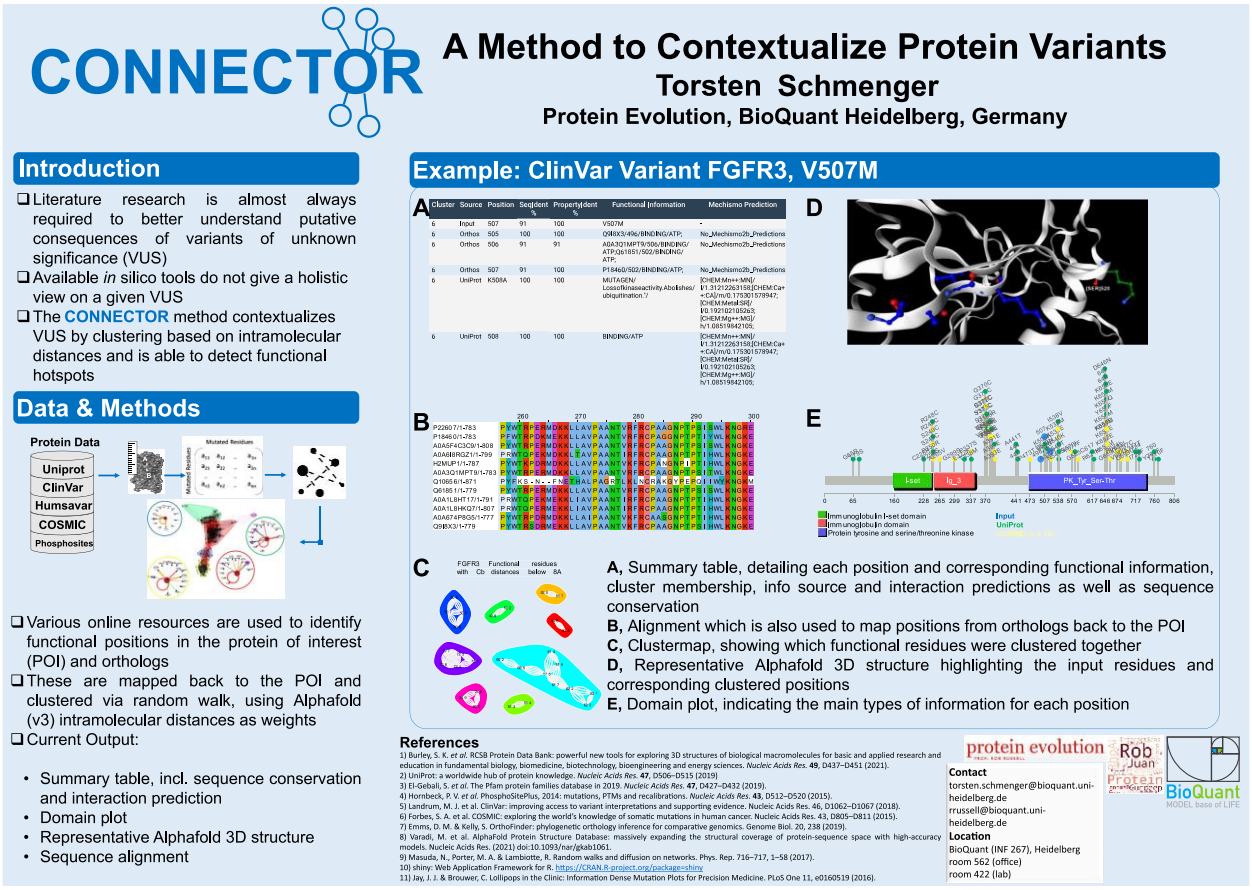


Figure 3H. Exemplary poster displaying features of CONNECTOR, the future web app for distance-based functional clustering. The clustered data would be accompanied by a reference structure (using Alphafold, D) with clusters highlighted, a sequence alignment of all used sequences (B), a graphical cluster map (C) and a domain plot showing known variant information on the protein of interest (E).

Chapter IV: Loss of Function Variants of Transforming Protein RHOA Show Heterogeneous Behaviour

4.1 Introduction

In silico methods can help advancing the understanding of biological processes. Computationally derived information can also be used to guide experimental biology, saving time and resources. The following chapter discusses one such example, where *in silico* and *in vitro* methods are used to supplement each other.

RHOA, a small GTP-binding protein (21 - 25 kDa) of the RHO (RAS homolog) family, is a master regulator of several crucial cellular functions including maintenance and modulation of the cytoskeleton, cell morphology and motility as well as cell proliferation.

The 3D structure is broadly similar to that of other GTPases like HRAS. It consists of six-stranded β-sheet surrounded by 6 α-helices, all interconnected with loops[202]. RHO family members (e.g. RHOA, RHOB, RHOC, CDC42, RAC1) have an additional insertion in form of a 3-turn helix from Asp124 to Gln136, between a loop from B5 to A4[202,203] (**Fig. 4A**, **A**). Like other GTPases RHO family members, also including CDC42 and RAC1 (**Fig. 4A**, **C** and Appendix 8.1), act as “molecular switches” typically by cycling between a GDP-bound inactive state and a GTP-bound active state[204] (**Fig. 4A**, **B**). This cycle is regulated by guanine nucleotide exchange factors (GEFs), accelerating GDP to GTP exchange, and GTPase-activating proteins (GAPs), stimulating GTP hydrolysis rate. GTP hydrolysis is typically two orders of magnitude faster than GDP/GTP exchange[205]. Guanine nucleotide dissociation inhibitors (GDIs) can then bind to prenylated RHO-GDP proteins in the cytosol, preventing their association with the cell membrane, a prerequisite for RHO protein function, while also stabilizing RHO GTPases[206]. Members of the GEF protein family can only activate RHO GTPases after GDI proteins are dissociated through GDI displacement factors (GDFs). The GTP-bound active state of GTPases is subjected to conformational changes in flexible loops called switch I (29-42, RHOA numbering) and II (62-68)[207], enabling RHO family members to selectively interact with downstream partners[206,208] (**Fig. 4A**, **D**).

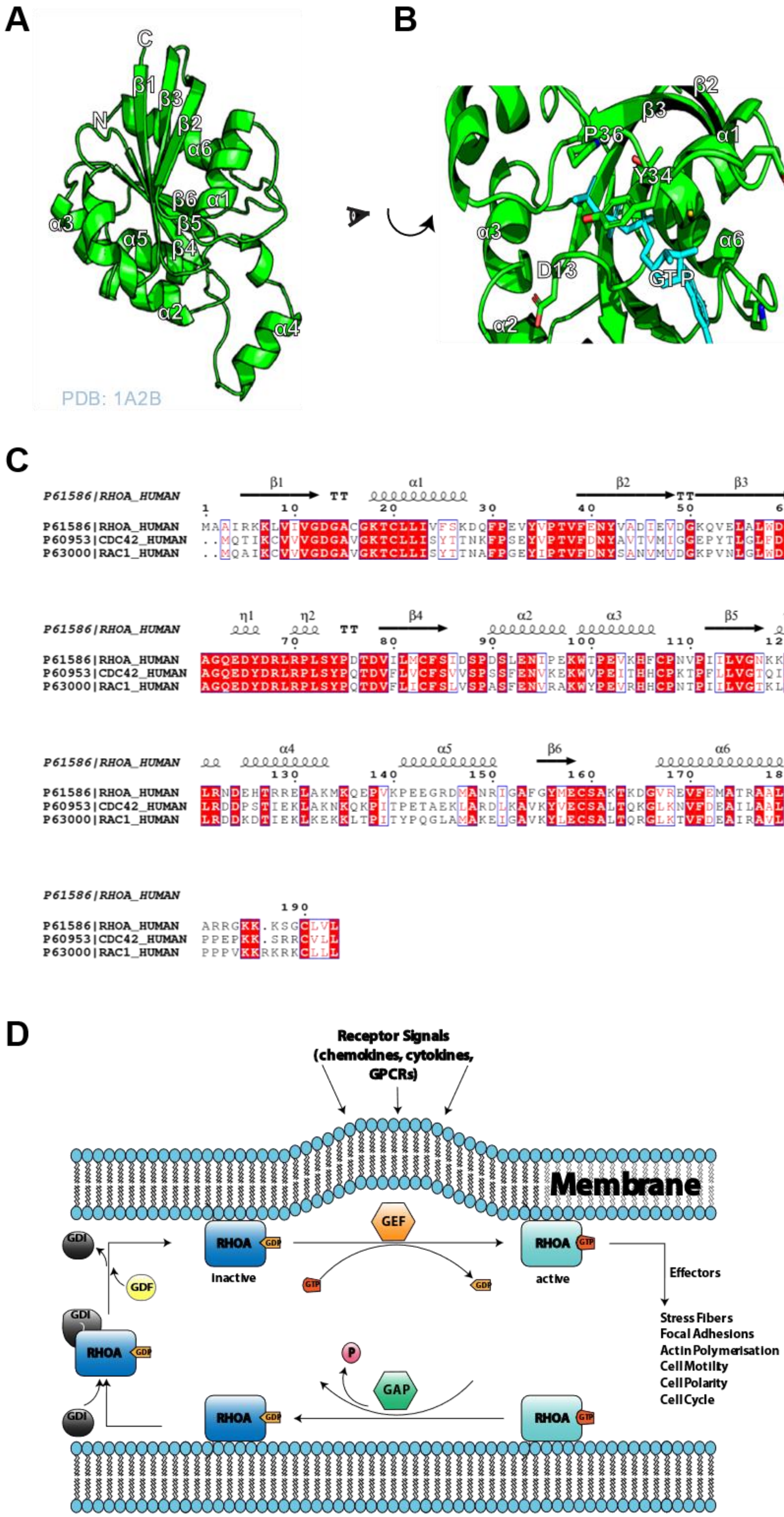

A
C
N
β1
β3
β2
α6
β6
α1
α3
α5
β5
β4
α2
α4
PDB: 1A2B
B
β2
β3
P36
α1
Y34
α3
α6
D13
GTP
α2
C
P61586|RHOA_HUMAN
P61586|RHOA_HUMAN MAAIRKKLVIVGDGACGKTCLLIVFSKDQFPEVYVPTVFENYVADIEVDGKQVELALWDT
P60953|CDC42_HUMAN ..MQTIKCVVVGDGAVGKTCLLISYTTNKFPSEYVPTVFDNYAVTVMIGGEPYTLGLFDT
P63000|RAC1_HUMAN ..MQAIKCVVVGDGAVGKTCLLISYTTNAFPGEYIPTVFDNYSANVMVDGKPVNLGLWDT
P61586|RHOA_HUMAN AGQEDYDRLRPLSYPDTDVILMCFSIDSPDSLENIPEKWTPEVKHFCPNVPIILVGNKKD
P60953|CDC42_HUMAN AGQEDYDRLRPLSYPQTDVFLVCFSVVSPSSFENVKEKWVPEITHHCPKTPFLLVGTQID
P63000|RAC1_HUMAN AGQEDYDRLRPLSYPQTDVFLICFSLVSPASFENVRAKWYPEVRHHCPNTPIILVGTKLD
P61586|RHOA_HUMAN LRNDEHTRRELAKMKQEPVKPEEGRDMANRIGAFGYMECSAKTKDGVREVFEMATRAALQ
P60953|CDC42_HUMAN LRDDPSTIEKLAKNKQKPITPETAEKLARDLKAVKYVECSALTQKGLKNVFDEAILAALE
P63000|RAC1_HUMAN LRDDKDTIEKLKEKKLTPITYPQGLAMAKEIGAVKYLECSALTQRGLKTVFDEAIRAVLC
P61586|RHOA_HUMAN ARRGKK.KSGCLVL
P60953|CDC42_HUMAN PPEPKK.SRRCVLL
P63000|RAC1_HUMAN PPPVKKRKRKCLLL
D
Receptor Signals (chemokines, cytokines, GPCRs)
Membrane
GEF
RHOA
GDP
inactive
GTP
active
Effectors
GDI
GDF
Stress Fibers
Focal Adhesions
Actin Polymerisation
Cell Motility
Cell Polarity
Cell Cycle
P
GAP

Figure 4A. RHOA Overview. A. RHOA showing secondary structure elements labelled. B. GTP analogue complexed with RHOA. **C. Alignments of RHOA, CDC42 and RAC1**. Alignments were generated with Clustal Omega, and secondary structure elements, displayed above the sequence, were overlaid using the Espript3 web server. Conserved residues are highlighted in red. **D. RHO GTPase cycle of activity**. Proteins of the RHO family typically cycle between an active (GTP bound) and inactive (GDP bound) state. Prenylation and tethering to the membrane is required for RHOA activity, and the different cycle states are mediated and assisted by a variety of proteins, including GAP and GEF proteins, as well as GDI and GDF proteins. For a detailed description see main text.

Interaction specificity is further increased by an overabundance of GTPase cycle regulating proteins, with 145 GEFs and GAPs acting on only 10 – 20 RHO GTPases[209]. The difference in numbers suggests that a tight regulation of all proteins by post-translational modifications (PTMs) is required to prevent dysregulation of RHO GTPases[210]. Differentially regulated RHO proteins play an important role in human disease, although their role in human cancer remains elusive. While RHOA and other GTPases are frequently found overexpressed in human cancers, activating mutations on the other hand are rare events[16], and current knowledge is insufficient to classify RHOA as either a tumour suppressor or an oncogene[211]. However, alterations of RHOA function are linked to carcinogenesis[206], activation of invasion and metastasis in gastric cancer[212], and connections between RHOA signalling with inflammation and cardiac disease are also the subject of ongoing research[213–215].

A frequent somatic cancer mutation of RHOA is Gly17Val, thought to be a loss-of-function (LoF) mutation due to the resulting inability of RHOA to bind GTP[216]. LoF variants are often found in cancer samples, though gain-of-function (GoF) variants, while seen in some instances[217], are not seen with anything close to the classical GoF variants of HRAS or KRAS[16]. The consequences of RHOA dysregulation appear to be highly context dependent, and the seemingly equivalent consequences of LoF or GoF variants for human health and disease are intriguing. One possible explanation is that RHOA LoF variants could be overcompensated by RHOB[218], which shares 85% of its sequence with RHOA[219]. There is, however, a notable difference between RHOA and RHOB in the C-terminus of RHOB, hinting at distinct functions[211].

Other intriguing RHOA, sometimes mutated, positions are Tyr34 and Tyr42, both located in the switch I region of RHOA (**Fig. 4B**). Phosphorylation of the switch I residue Tyr34, a known contact of RHOA effector proteins[204] by Src kinase[220], prevents the binding of RHOA to interaction partners. Phosphorylation of Tyr42 has the opposite effect, as it enhances the binding of VAV2, a RHOA GEF[206]. Both residues have been found mutated to cysteines in cancers (Tyr34Cys, Tyr42Cys), with cysteines putatively mimicking the negative charge of phosphorylation after oxidation of the cysteine sidechain[206], for example as a consequence of oxidative stress, a common co-occurrence of many diseases, including cancer[221]. Glu40 and Leu69 are also often found in contact with interaction partners, with Glu40 binding to the RHO-binding-domain of ROCK1 and Leu69 forming hydrophobic contacts with many RHOA interaction partners[204,206]. Perturbation of residues Tyr34, Glu40, Tyr42 or Leu69 are all thought to be LoF.

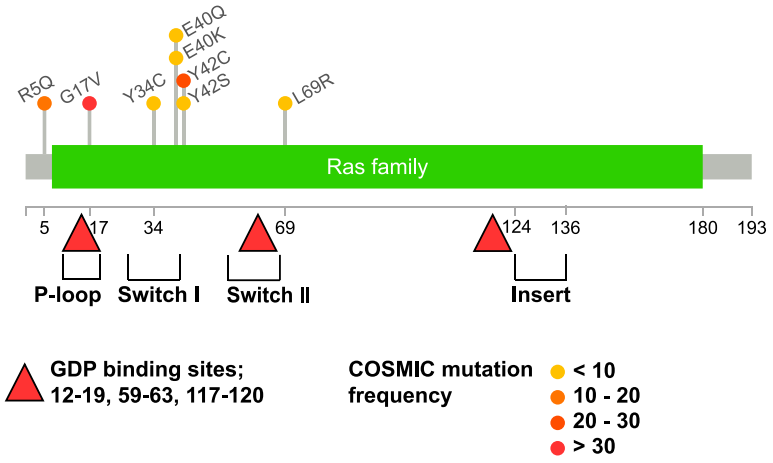


Figure 4B. Location of well-characterized RHOA functional features & variants. GDP binding sites are indicated as red triangles, variants as coloured dots according to their COSMIC frequency. Structural features (switch I and II, P-loop and the RHO family specific insert) are shown below the domain plot[64].

This project aims to explore the mechanism of how RHOA variants perturb interactions with regulatory proteins. The consensus on these variants is a LoF phenotype. In contrast RHOA has many functions, for example rearrangement of the cytoskeleton[213,222], and the debate whether RHOA is a tumour suppressor or an oncogene is still ongoing[211,223,224]. Computational investigations and subsequent experimental analysis of variants reveals a spectrum of effects on protein-interactions

and on cell phenotypes that suggests complex context-specific roles for RHOA variants in cancer.

4.2 Material

4.2.1 Inhibitors and antibiotics

Name	Company	Order No.
Ampicillin	Sigma Aldrich	A9518-5G
Gentamicin	Sigma Aldrich	G1914-5G
Kanamycin	Sigma Aldrich	K1377-5G
Mitomycin C	Sigma Aldrich	10107409001
Penicillin-Streptomycin	Thermo Fisher Scientific	15140122
Zeocin	Invivogen	Ant-zn-1

4.2.2 Medias and supplements for mammalian cells

I. Culture medium

High glucose DMEM (GlutaMAX, Thermo Fisher Scientific) containing 10 % (v/v) fetal bovine serum (FBS) (Thermo Fisher Scientific) as well as 100 U/mL penicillin and 100 µg/mL streptomycin (see 2.1.1) was used. If not otherwise stated no other medium was used.

II. Freezing medium

Cells were frozen in DMEM as in I. with the addition of 10 % (v/v) dimethyl sulfoxide (DMSO, Sigma Aldrich).

4.2.3 Media and supplements for cultivation of *E. coli*

I. Culture medium

20 g/L LB Broth (Sigma Aldrich) was diluted in H_2O and autoclaved at 121 °C. Before using the medium for liquid culture of *E. coli* cells the mixture was supplemented with either Zeocin (25 µg/mL), Kanamycin (50 µg/mL), Gentamicin (20 µg/mL) or Ampicillin (100 µg/mL).

II. Outgrow medium

SOC medium (New England BioLabs) was added to freshly transformed *E. coli* cells.

III. Agar Plates

LB-Agar was prepared with 20g/L LB Broth and 10 g/L agar (VWR). After dissolving both powders in water, the mixture was autoclaved at 121 °C. Afterwards, the solution was cooled down to approximately 55 °C before adding antibiotics and pouring round petri dishes (VWR).

4.2.4 Media and supplements for cultivation of S. cerevisiae

I. Culture medium

50 g/L YPD Broth (Sigma Aldrich) was diluted in H_2O and autoclaved at 121 °C.

II. Dropout Medium

For preparation of 500 mL medium (2x concentrated) 6.7 g Yeast Nitrogen Base without amino acids (Sigma Aldrich), 20 g glucose (Merck) and yeast synthetic dropout supplements (Sigma Aldrich).

Synthetic Dropout Supplements	Amount [g]
-Histidine -Leucine -Tryptophan	1.46
-Leucine -Tryptophan	1.54
-Leucine -Tryptophan -Uracil	1.46

The mixture was then filtered sterile using DURAPORE PVDF (.22 µm) stericups (Sigma Aldrich).

III. Agar Plates

To cultivate yeast on solid agar plates a 2x concentrated YPD or dropout medium was mixed with the same volume of autoclaved 4 % (w/v) agar, yielding a final agar concentration of 2 % (v/v). I poured plates in square petri dishes (VWR).

4.2.5 Plasmids

Name	Supplier	Resistance
pDONR/Zeo	Thermo Fisher Scientific	Zeocin
pDest22	Thermo Fisher Scientific	Ampicillin
pDest32	Thermo Fisher Scientific	Gentamicin
pcDNA3.1(+)/Zeo	Thermo Fisher Scientific	Zeocin
pcDNA3.1(+)/Zeo_hRHOA	Thermo Fisher Scientific	Zeocin
pGADT7 AD	Invitrogen	Ampicillin
pGBKT7	Invitrogen	Kanamycin
pEXP22/RalGDS wt	Invitrogen	Ampicillin
pEXP32/KREV1	Invitrogen	Gentamicin
pDONR/Zeo_RHOA	*	Zeocin

* kindly prepared by Dr. Oliver Wichmann, a former Postdoc of AG Russell

Plasmid maps can be found in Appendix 8.2.

4.2.6 DNA Sequences

Table 1: List of DNA inserts used to create DNA plasmids			
	Modification		
Insert Name	Protein Position	Amino Acid Change	Nucleotide Change
RHOA	Wildtype		
RHOA	61	Ala → Asp	gcc → gAc
RHOA	24	Val → Phe	gtg → TtC
RHOA	75	Pro → Arg	ccc → cGc
RHOA	34	Tyr → Cys	tac → tGc
RHOA	40	Glu → Lys	gag → AGg
RHOA	40	Glu → Gn	gag → Cag
RHOA	42	Tyr → Cys	tac → tGc
RHOA	42	Tyr → Ser	tac → tCc
RHOA	5	Arg → Gln	aga → CAa
RHOA	14	Gly → glu	ggc → gAA
RHOA	14	Gly → Val	ggc → gTc
RHOA	17	Gly → Glu	gga → gAa
RHOA	17	Gly → Val	gga → gTa
RHOA	69	Leu → Arg	ctg → cGg
NRAS	wildtype		
GNAQ	wildtype		
ARHGEF25	wildtype		
DIAPH1	wildtype		
ARHGAP20	383 - 551	wildtype	
PAK1	wildtype		
ROCK1	948 – 1323	wildtype	
ITSN1	12237 - 1571	wildtype	
ARHGEF6	238 – 550	wildtype	
DOCK7	1373 – 2140	wildtype	

The complete exon sequences are given in section ‘8.3 Complete DNA Sequences’.

.

4.2.7 Small interfering (si) RNAs

I used Silencer ® Select Validated siRNA (Thermo Fisher Scientific) and Silencer ® Select Negative Control No. 1 (Thermo Fisher Scientific) to target RHOA.

Target	siRNA ID	Ref. Seqs.	Sequence (5'-3')
RHOA	s758	NM_001313941.1 NM_001313943.1 NM_001313944.1 NM_001313945.1 NM_001313946.1 NM_001313947.1 NM_001664.3	CACAGUGUUUGAGAACUAUtt

4.2.8 Transfection reagents

Name	Company
Lipofectamine 2000	Thermo Fisher Scientific
Lipofectamine 3000	Thermo Fisher Scientific
Lipofectamine RNAiMAX	Thermo Fisher Scientific

4.2.9 Primers (PCR, cloning, qPCR, sequencing)

Primers were ordered from Thermo Fisher Scientific or Eurofins Genomics.

Table 2: List of qPCR Primers		
Target	**Forward primer (5'-3')**	**Reverse primer (5'-3')**
RHOA #1	GTGTCTTGCTATGTTGCCC	ACTCTACCTGCTTTCCATCC
RHOA #2	CCCAGACAGATCTTGTACTCC	TACCTGCTTTCCATCCACC
RHOA #3	TTCCATCGACAGCCCTGATAGTTTA	CACGTTGGGACAGAAATGCTTG
ACTIN	CACCATTGGCAATGAGCGGTTC	AGGTCTTTGCGGATGTCCACGT
GAPDH	GTCTCCTCTGACTTCAACAGCG	ACCACCCTGTTGCTGTAGCCAA
RAC1	ATGTCCGTGCAAAGTGGTATC	CTCGGATCGCTTCGTCAAACA
CDC42	TTGCTTGTCGGGACCCAAAT	GGCGGAACACTCCACATACT
ROCK1	AACATGCTGCTGGATAAATCTGG	TGTATCACATCGTACCATGCCT
ARHGDIA	CAGGAAAGGCGTCAAGATTG	GTCAGGAACTCGTACTCCTC
DIAPH3	GAAACACGGTTGGCAGAGTCT	GTGGCCGTAGTCTCTTCACA

IGF1 #1	GTGCGGAGACAGGGGCTTT	ACTTGGCGGGC TTGAGAGG
IGF1 #2	CTCTTCAGTTCGTGTGTGGAGAC	CAGCCTCCTTAGATCACAGCTC
TNFα	CTCTTCTGCCTGCTGCACTTTG	ATGGGCTACAGGCTTGTCACTC
CCND1	CCCTCGGTGTCCTACTTCAA	GTGTTCAATGAAATCGTGCG
NFκB1	GCAGCACTACTTCTTGACCACC	TCTGCTCCTGAGCATTGACGTC
SOX9	AGGAAGCTCGCGGACCAGTAC	GGTGGTCCTTCTTGTGCTGCAC
HIF1α	CATAAAGTCTGCAACATGGAAGGT	ATTTGATGGGTGAGGAATGGGTT
c-MYC	TGAGGAGACACCGCCCAC	CAACATCGATTTCTTCCTCATCTTC

Table 3: List of PCR Primers		
Target	**Forward primer (5'-3')**	**Reverse primer (5'-3')**
RHOA	GGAGTCCACGGTCTGGTC	CCTTCACAAGACAAGGCA
pDONR/Zeo	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACC
pDest22	GATGATGAAGATACCCCAC	TCGAGACCTTCCGCTT
pDest32	AGTGCGACATCATCATCG	CGTTTTAAAACCTAAGAGTCAC
pcDNA3.1	ACCTGACGTCGACGGATCGGGAGATCTC	CCGATCCGTCGACGTCAGGTGGCACTTTTC

Table 4: List of Mutagenesis PCR Primers (Yeast)		
Target	**Forward primer (5'-3')**	**Reverse primer (5'-3')**
RHOA Y34C	CCAGAAGTTTgCGTTCCAACC	GAATTGGTCCTTGGAGAAAAC
RHOA E40K	AACCGTTTTTaAAAACTACGTTG	GGAACGTAAACTTCTGGG
RHOA E40Q	AACCGTTTTTcAAAACTACGTTG	
RHOA Y42C	TTTGAAAACTgCGTTGCCGATATTG	AACGGTTGGAACGTAAAC
RHOA Y42S	TTTGAAAACTcCGTTGCCGATATTG	

Table 5: List of Mutagenesis PCR Primers (Mammalian Cells)		
Target	**Forward primer (5'-3')**	**Reverse primer (5'-3')**
RHOA Y34C	CCCGAGGTGTgCGTGCCCACC	GAACTGATCCTTGCTGAACACGATCAGC
RHOA E40K	CACCGTGTTCaAGAATTACGTGGCC	GGCACGTACACCTCGGGG
RHOA E40Q	CACCGTGTTCcAGAATTACGTGGCC	
RHOA Y42C	TTCGAGAATTgCGTGGCCGAC	CACGGTGGGCACGTACAC
RHOA Y42S	TTCGAGAATTcCGTGGCCGACATC	
RHOA + FLAG tag	TGTCTGGTGCTGGATTATAAAGATGATGATGATAAA	TGGACTAGTGGATCCTTATCATTTATCATCATCATCTTTATAATC

4.2.10 Antibodies

Table 6: List of Antibodies				
Target	**Origin**	**Dilution/ Concentration**	**Supplier**	**2nd Antibody**
Immunoblotting				
RHOA	mouse	1:500	Thermo Fisher Scientific	mouse
ACTIN	mouse	1:5000		mouse
GAPDH	rabbit	1:5000		rabbit
RAC1	rabbit	1:2000		rabbit
mouse-HRP	goat	1:5000		-
rabbit-HRP	goat	1:5000		-
Immunofluorescence				
pTyr20-FITC	rabbit	2 µg/mL	Thermo Fisher Scientific	-

4.2.11 Buffers and solutions

I. 10x gel electrophoresis running buffer

For 1 L of a 10x stock solution, 250 mM Tris (30.2 g, Carl Roth), 1. 92 M Glycin (144g, Sigma Aldrich) and 1 % (w/v) sodium dodecyl sulfate (SDS, 10g, Sigma Aldrich) were mixed by me and brought to a final volume of 1 L with deionized H_2O. A working solution was prepared by mixing 100 mL of the final 10x stock solution with 900 mL deionized H_2O.

II. 10x TBS

For 1 L of a 10x stock solution I mixed 1 M Tris (121 g, Carl Roth) and 1.5 M NaCl (90 g, Carl Roth) and brought to a final volume of 1 L with deionized H_2O, pH was adjusted to 7.5. I prepared a working solution by mixing 100 mL of the final 10x stock solution with 900 mL deionized H_2O.

III. TBST

A working solution of 1x TBST was created by mixing 100 mL of 10x TBS with 900 mL deionized H_2O. 0.1 % (v/v) Tween-20 (Sigma Aldrich) was then added to the solution.

IV. 4x resolving buffer

For the creation of 500 mL 4x resolving buffer I dissolved 1.5 M Tris (90 g, Carl Roth) and 0.4 % (w/v) SDS (2 g, Sigma Aldrich) in deionized H_2O and pH adjusted to 8.8.

V. 2x stacking buffer

To prepare 500 mL 2x stacking buffer I dissolved 250 mM Tris (15.14 g, Carl Roth) and 0.2 % (w/v) SDS (1 g, Sigma Aldrich) in deionized H_2O and pH adjusted to 6.8.

VI. 4x SDS sample buffer

50 mL of SDS sample buffer was composed of 40 % (v/v) glycerol (20 mL, Sigma Aldrich), 200 mM Tris (15.4 g, Carl Roth), 8 % (w/v) SDS (4 g, Sigma Aldrich) and 0.004 % (w/v) Bromphenolblue (2 mg, Sigma Aldrich). The mixture was brought up to 50 mL with H_2O and pH adjusted to 6.8.

VII. 4 % Formaldehyde

100 mL of a 4 % (w/v) Paraformaldehyde solution was prepared by heating up 80 mL phosphate buffered saline (PBS, Thermo Fisher Scientific). I then added 4 g paraformaldehyde (Sigma Aldrich) and stirred the mixture. The pH was slowly raised until paraformaldehyde was fully dissolved. The pH was then adjusted to 6.8 and the solution brought to a final volume of 100 mL with PBS. Aliquots were made and stored at – 20 °C.

VIII. TFB 1

100 mL of TFB 1 buffer was created by mixing 30 mM potassium acetate (Sigma Aldrich) with 100 mM rubidium chloride (Sigma Aldrich), 10 mM calcium chloride (Sigma Aldrich), 50 mM manganese (II) chloride and 15 % (v/v) glycerol. The pH was adjusted to 5.8 and the mixture was filter sterilized.

IX. TFB 2

100 mL of TFB 2 buffer was created by mixing 10 mM MOPS (Sigma Aldrich) with 10 mM rubidium chloride (Sigma Aldrich), 75 mM calcium chloride (Sigma Aldrich), and 15 % (v/v) glycerol. The pH was adjusted to 6.5 and the mixture was filter sterilized.

X. Others

Buffer or solution	Supplier
RIPA Lysis and Extraction Buffer	Thermo Fisher Scientific
M-PER™ Mammalian Protein Extraction Reagent	Thermo Fisher Scientific
SimplyBlue Safestain	Thermo Fisher Scientific
TAE buffer	Carl Roth
Sterile Water	VWR
Imperial Protein Stain	Thermo Fisher Scientific

4.2.12 Chemicals

Name	Company
2-Propanol	Sigma Aldrich
Albumin	Carl Roth
Ammoniumperoxodisulfat (APS)	Carl Roth
Aquaresist	VWR
Ethanol (EtOH) absolute	Sigma Aldrich
Ethanolamine	Sigma Aldrich
Ethylenediaminetetraacetic acid (EDTA)	Sigma Aldrich
EtOH 99%, 1 % petroleum ether	Central University Deposit
Fluoromount Aqueous Mounting Medium	Sigma Aldrich
Fluoroshield with DAPI	Sigma Aldrich
Gelatine	Sigma Aldrich
Methanol (MeOH)	Central University Deposit
N-Tetramethylethylenediamine B (Temed)	Sigma Aldrich
Polyethylene glycol (PEG) 3350	Sigma Aldrich
Ponceau S Dye	Sigma Aldrich
Rotiphorese Gel 30	Carl Roth
Skim milk powder	Carl Roth
Sodium azide	Sigma Aldrich
Triton X-100	Sigma Aldrich
Trypsin-EDTA (0.05%)	Thermo Fisher Scientific

4.2.13 Enzymes & ready-to-use premixes

Enzyme name	Company
Antarctic phosphatase	New England Biolabs
DpnI	New England Biolabs
KpnI-HF	New England Biolabs
Phusion-HF	New England Biolabs
PstI-HF	New England Biolabs
SpeI-HF	New England Biolabs
T4 Polynucleotide Kinase	New England Biolabs

Ready-to-use premix name	Company
1kb Plus DNA Ladder	Thermo Fisher Scientific
Adenosine 5’-Triposphate (ATP)	New England Biolabs
Fast SYBR Green Master Mix	Thermo Fisher Scientific
Halt protease inhibitor cocktail	Thermo Fisher Scientific
KLD Enzyme Mix	New England Biolabs
PageRuler Plus Prestained Protein Ladder	Thermo Fisher Scientific
Phalloidin-Atto 590	Sigma Aldrich
Ultrapure salmon sperm DNA	Thermo Fisher Scientific

4.2.14 Kits

Name	Company
ECL Western Blotting Substrate	Pierce
High-Capacity cDNA Reverse Transcription Kit	Thermo Fisher Scientific
MicroAmp Optical Adhesive Film Kit	Thermo Fisher Scientific
NucleoSpin Plasmid Mini kit	Macherey & Nagel
Q5 Hot Start HF PCR Kit	New England Biolabs
Qubit Protein Assay Kit	Thermo Fisher Scientific
QuiaQuick gel extraction kit	Qiagen
Quick Dephosphorylation Kit	New England Biolabs
Quick Ligation Kit	New England Biolabs
RNeasy Mini Kit	Qiagen

4.2.15 Equipment & devices

Name	Company
Nikon Ti-HCS microscope	Nikon
Environmental Box	Oko Lab
Centrifuge S804R	Eppendorf
C1000 Thermal Cycler	BioRad
IncuLine Incubator	VWR
Roller Mixer SRT6D	Stuart
Nano Photometer	Implen
Scales	Sartorius
Pipettes	Eppendorf
iBase Gel	Invitroen
MS3 Digital Vortexer	Ika
Mini Star Centrifuge	VWR
Gel Doc EZ Imager	BioRad
PS-M3D Shaker	Grantbio
Unimax 1010 Shaker	Heidolph
PowerPac Basic	BioRad
Waterbath	VWR
Dry Bath System	Starlab
Incu Shaker Mini	Benchmark
Heraeus Fresco17 Centrifuge	Thermo Fisher Scientific
iBlot Gel Transfer	Thermo Fisher Scientific
Innova 42 Shaker	New Brunswick Scientific
Hera Cell 150 Incubator	Thermo Fisher Scientific
Hera Safe Cell Culture Cabinet	Thermo Fisher Scientific
CKX41 Microscope	Olympus
Vortex Genie 2	Scientific Industries
Allegra X-12 Centrifuge	Beckman Coulter
Premium Freezer	Liebherr
StepOne Plus	Applied Biosystems
Azure 400 Visible Fluorescence Western Blot Imaging System	Azure Biosystems

4.2.16 Consumables

Name	Company
50 mL self-standing centrifuge tubes	VWR
Corning cell scrapers	Sigma Aldrich
Corning Cryovials	Sigma Aldich
Corning T75 Cell Culture Flask	Sigma Aldrich
Counting Chamber Coverslips	VWR
Coverslips	VWR
Coverslips and Coverslip Pickup tool	ibidi
Culture-Insert 3 Well, in µ-dish, 35 mm, high	ibidi
Disposable cuvettes	VWR
E-gel EX agarose gels, 1%, 2%	Thermo Fisher Scientific
Glass beakers	VWR
Glass-bottom dish, 35 mm	ibidi
iBlot transfer stack, nitrocellulose, regular size	Thermo Fisher Scientific
iBlot transfer stack, pvdf, regular size	Thermo Fisher Scientific
Inocluation loops	VWR
MicroAmp 96-well Support Base	Thermo Fisher Scientific
MicroAmp Fast Optical 96-well Reaction Plates	Thermo Fisher Scientific
Mini-Protean Short Plates	Biorad
Nitrile gloves	Central University Deposit
Nunc 100mm EasyDish	Thermo Fisher Scientific
Nunc 15 mL centrifuge tubes	Thermo Fisher Scientific
Nunc Cell-Culture treated multidishes, 6-well	Thermo Fisher Scientific
Nunc EasYFlask Cell Culture Flasks, T25	Thermo Fisher Scientific
PCR tubes	Biozym
Petri dishes, square or round	VWR
Pipetting Reservoir	Biozym
Qubit assay tubes	Thermo Fisher Scientific
RNase-free Microfuge tubes, 1.5 mL	Eppendorf
SafeSeal SurPhob pipette tips, 10 µL - 1000 µL	Biozym
Serological Pipettes, 5 mL, 10 mL, 25 mL	Sigma Aldrich

Steriflip-GP sterile centrifuge tube top filter unit	Sigma Aldrich
SupremeRun Sequencing Barcodes	Eurofins
Wide neck bottles	VWR

4.2.17 Cell lines

I. Mammalia Cells

Human osteosarcoma U2OS cells were kindly provided by the lab of Prof. Dr. Perihan Nalbant (University of Duisburg-Essen).

II. Bacteria

One Shot TOP10 chemically competent *E. coli* were bought from Thermo Fisher Scientific. To propagate empty Gateway vectors TOP10 ccdB survival *E. coli* were obtained from Thermo Fisher as well.

III. Yeast

S. cerevisiae MaV203 Competent Yeast Cells were obtained as part of the ProQuest Two-Hybrid System (Invitrogen).

4.2.18 Software & databases

Name	Company/ Source
Benchling	https://www.benchling.com/
BioGrid	https://thebiogrid.org/
BLAST	https://blast.ncbi.nlm.nih.gov/Blast.cgi
COSMIC	https://cancer.sanger.ac.uk/cosmic
Expasy ProtParam	https://web.expasy.org/protparam/
ExPaSy Translate	https://web.expasy.org/translate/
Fiji	https://imagej.net/software/fiji/
ImageJ	https://imagej.nih.gov/ij/
Inkscape	Inkscape Project
Jalview	https://www.jalview.org/
Mechismo3	http://mechismo3.russelllab.org/
Microsoft Office	Microsoft

NEBase Changer	https://nebasechanger.neb.com/
NIS-Elements AR	Nikon
Protein Data Bank	https://www.rcsb.org/
Pubmed	https://pubmed.ncbi.nlm.nih.gov/
PrimerBank	https://pga.mgh.harvard.edu/primerbank/
PyMol	Schrödinger, Inc.
Python 2.7.18	Python Software Foundation
R Studio	https://www.rstudio.com/
R-4.1.1	The Comprehensive R Archive Network
The Cancer Genome Atlas	https://portal.gdc.cancer.gov/
UniProt	https://www.uniprot.org/

4.3 Methods

4.3.1 Cell Culture

I. Osteosarcoma Cells

I maintained U2OS cells in standard culture medium. Overexpression and respective control cells were generated with the pcDNA3.1 vector system and maintained under permanent presence of zeocin (50 µg/mL).

To passage cells, I removed standard culture medium and I washed cells with 8 mL PBS. After addition of 1 mL Trypsin-EDTA cells were then incubated for 10 mins at 37 °C. Detachment of cells was confirmed using a microscope and 4 mL standard culture medium was added to stop the trypsin reaction. 300.000 cells were kept for propagation and cells were split 1:10 by keeping 500 µL of cell suspension and adding 9.5 mL standard culture medium for a total volume of 10 mL.

II. Determination of cell number

I washed cells with PBS, detached them with trypsin and resuspended the cells in standard culture medium. 15 µL of the cell suspension was loaded into a Neubauer counting-chamber (VWR) and cells were counted. The mean cell number of all four corner squares, each consisting of 16 small squares, was multiplied by 10.000 to arrive at a cell count/mL. The multiplication factor of 10.000 is based on the volume of a corner square of 0.1 µL (w= 1 mm, l= 1mm, h=0.1 mm).

III. Cell cryo-conservation and re-cultivation

After washing, cell detachment and counting 3.000.000 U2OS cells were pelleted by centrifugation at 300 g for 10 minutes. I resuspended cells in 1 mL of freezing medium and transferred into a cryovial. The vial was cooled down in 100% 2-propanol using *Mr. Frosty Freezing Container* (Thermo Fisher Scientific) to -80 °C for ≥ 24 hours. Afterwards the cryovials were stored at -160 °C in liquid nitrogen.

For re-cultivation I thawed cells at 37 °C and resuspended them in standard culture medium. Medium was completely exchanged 24 hours after thawing.

4.3.2 Modulation of gene expression

I. siRNA-mediated knockdown

Silencer ® Select Validated siRNA was used to target human RHOA. First 150 µL OptiMEM medium (Gibco) was mixed with 3 µL of 10 µM concentrated siRNA. In a second tube 150 µL OptiMEM medium was mixed with 9 µL Lipofectamine RNAiMAX. I mixed both tubes and incubated at room temperature for 5 minutes. The resulting siRNA-lipid complex was added dropwise to 400.000 plated U2OS cells in 6-well plates yielding a final siRNA concentration of 15 nM. Cells were incubated for 48 hours and then used for experiments.

II. Plasmid-based upregulation of RHOA expression

Vectors with the pDNA3.1(+)/Zeo backbone were used to transiently overexpress RHOA wildtype or mutants in U2OS cells. Transfection was achieved my mixing 150 µL OptiMEM medium with 2.5 µg plasmid DNA and 5 µL of P3000 reagent. In a second tube 150 µL OptiMEM medium was mixed with 5 µL Lipofectamine 3000. I combined both mixtures and then incubated them at room temperature for 15 minutes. Afterwards the lipid-DNA complexes were added dropwise to 400.000 plates U2OS cells in 6-well plates. Medium was exchanged after 24 hours and 50 µg/mL Zeocin was added. Overexpression was confirmed via Western Blot.

4.3.3 Polymerase chain reaction (PCR) based applications

I. Standard Phusion PCR

For basic amplification of DNA I followed the guidelines provided by the supplier of Phusion Polymerase (NEB). Primer sequences were generally looked for in PrimerBank, PubMed or designed manually using Benchling. PCR cycles were run on the C1000 Thermal Cycler.

II. Q5 Mutagenesis PCR

I designed primers using the NEBase changer tool (see Table 4 & 5), to mutate either pDONR/Zeo_hRhoA or commercially bought pcDNA3.1(+)/Zeo_hRhoA from wildtype to several mutants. Mutagenesis proceeded following the instructions by the manufacturer. To confirm the success of the PCR 5 µL PCR product mixed with 1 µL 6x Purple Dye was loaded into a 1 - 2 % agarose gel and analysed. The PCR product was then transformed into chemically competent TOP10 *E. coli* and the presence of

the desired nucleotide changes was confirmed by SupremeRun sequencing (Eurofins).

4.3.4 Quantitative reverse transcription (qRT)-PCR

Cells were treated as described above. I isolated total RNA using Qiagen RNeasy RNA isolation kit, following the standard instructions. 1 µg isolated RNA of each sample was used for cDNA-synthesis using the High-Capacity cDNA Reverse Transcription Kit according to the standard procedure. cDNA was then submitted to gene expression analysis on a StepOne Plus device. Samples were diluted in triplicates both 1:10 and 1:100. A single reaction was comprised of 5 µL diluted cDNA, 0.3 µL primer mix (5 µM forward and 5 µM reverse primer, diluted in nuclease-free water, see Table 2) and 7.5 µL Fast SYBR Green Master Mix. The mixture was brought up to 15 µL with 2.2 µL nuclease-free water. I initially set cycling reactions to 95 °C for 20 seconds, 40 cycles of 95 °C for 3 seconds and 60 °C or 30 seconds, followed by 95 °C for 15 seconds. Afterwards, a melt curve analysis was performed starting with 60 °C for 1 minute followed by stepwise increases of 0.3 °C until 95 °C, for 15 seconds. Relative gene expression of specific genes was calculated by comparison of Ct-values between gene of interest and indicated housekeeping gene(s).

4.3.5 Preparation of cell lysates

Cells were detached with trypsin and spun down for 10 minutes at 300 rpm before washing the pellet twice with 5 mL ice-cold PBS. Depending on the size of the pellet, 20 to 50 µL lysis buffer (RIPA Lysis and Extraction Buffer or M-PER™ Mammalian Protein Extraction) were used to resuspend the pellet on ice. Afterwards, the sample was left on ice for 30 minutes followed by centrifugation at 13.000 g for 15 minutes at 4 °C. I discarded resulting pellets and the supernatant was stored at – 80 °C.

Determination of protein concentration was performed following the instructions of the Qubit Protein Assay kit. The required amount of total µg protein was mixed with 4x SDS sample buffer before storage at – 20 °C.

4.3.6 Immunoblot

I. Gel Casting

I casted gels following the supplier's instructions (BioRad). To create a 12 % SDS gel 24 mL Rotiphorese Gel 30, 15 mL 4x resolving buffer and 19.5 mL H_2O were mixed with 54 µL Temed and 900 µL 10 % (w/v) APS. The mixture was poured into the gelcaster and covered with isopropanol. After 30 minutes the isopropanol was removed by decanting the gelcaster and the stacking gel was prepared by mixing 5 mL Rotiporese Gel 30, 20 mL 2x stacking buffer and 13 mL H_2O with 28 µL Temed and 600 µL 10 % (w/v) APS. The mixture was poured atop the polymerized resolving gel and a comb was added. After polymerization of the stacking gel the gels were either directly used or wrapped in wet tissue and stored at 4 °C for up to two weeks.

II. Blotting

Protein samples were incubated at 95 °C for 5 minutes and 1 µL 14.3 M β-Mercaptoethanol (Sigma Aldrich) was added to the samples. Following brief centrifugation, I loaded the samples onto an SDS gel and separated via electrophoresis, using 15 mA per gel for 20 minutes and 30 mA per gel until the end. Electrophoresis would be stopped once the running front reached the end of the gel. Proteins were blotted onto nitrocellulose membranes with 20 V for 5 minutes using the iBlot device and following the manufacturer's instructions.

III. Antibody Detection

After washing in TBST once, membranes were blocked in TBST containing 5 % (w/v) bovine serum albumin (BSA, Carl Roth) for 2 h – 3 h at room temperature on a shaker. I diluted the primary antibodies in blocking solution as given in Table 6 and incubated the membranes at 4 °C overnight on a shaker. Membranes were washed three times in TBST for 5 minutes and incubated with secondary antibodies in TBST at the indicated dilutions for 1 h at room temperature. After washing the membrane three more times in TBST for 5 minutes immunodetection was performed following the manual of the ECL Western Blotting Substrate and using the Azure 400 Visible Fluorescence Western Blot Imaging System.

4.3.7 Immunofluorescent staining

Cells were seeded in ibidi glass bottom dishes and incubated in standard conditions. After 24 h I fixed the cells in 4 % (w/v) formaldehyde at room temperature for 20 minutes. Cells were washed three times in PBS and excess formaldehyde was quenched in 10 mM ethanolamine in PBS for 5 minutes. To increase cell permeability the sample was incubated for 5 minutes in 0.2 % (v/v) Triton-X100 (Sigma Aldrich) in PBS. Samples were blocked for 2 hours at room temperature with 1 % (w/v) Gelatine in PBS followed by overnight incubation at 4 °C in the dark with 2 µg/mL pTyr20-FITC antibody in blocking buffer. The next day, around 90 minutes before imaging, the blocking-buffer-antibody mixture was removed and 0.25 nmol Phalloidin-Atto 590 in PBS was added to the sample, followed by an incubation of 70 minutes at room temperature without light. The samples were then rinsed three times with 0.05 % (v/v) Tween in PBS for 5 minutes each, followed by adding mounting media (without DAPI when pTyr20-FITC antibody was used, else with DAPI) and imaging on a Nikon Ti-HCS microscope. Co-workers then renamed the images to hide their origin and to allow for blinded analysis. Phosphotyrosine counts and cells sizes were determined manually measuring the number of dots or the cells area using the FIJI software. The Corrected Mean Particle Fluorescence (CMPF) was calculated according to the **formula 2**:

$$CMPF = \sum Pixelbrightness_{Signal} - \left(\sum Area_{Signal} \frac{\sum Pixelbrightness_{Background}}{n_{Backgroundpixel}}\right) \qquad (2)$$

For each analysed cell I took 5 to 10 rectangular shaped background measurements and averaged them. Then, the integrated density of phosphotyrosine particles was determined by summing all pixel values of the respective signal. The resulting value was then corrected by the product of the signals area and the mean background. By analysing the images blinded, and by background-correcting each individual cell, I attempted to remove any analysis bias, or effects simply caused by differing staining efficiency between experiments or even between individual cells of the same slide.

4.3.8 *In vitro* gap-closing assay

Three times 70 µL cells (c = 300.000 cells/mL) were seeded into culture-insert 3 well dishes (in µ-dish, 35 mm, high) and incubated overnight under standard cell culture conditions. The following day the insert was carefully removed using tweezers and

culture medium was brought up to a final volume of 2 mL. I imaged samples on a Nikon Ti-HCS microscope at 37 °C, 5 % CO_2 for 20 hours in presence of 1 µM Mitomycin C, with a maximum of four different conditions being tested, due to technical limitations. Pictures were taken every 5 minutes over the whole duration of the experiment. The final cell number was determined immediately after the imaging period ended. I adjusted gap closing speeds to 100.000 cells and normalized them to the gap closing speed of WT cells from the same run.

4.3.9 Microbiological methods

I. Transformation of chemocompetent *E. coli*

DNA was diluted from 50 ng/µL to 100 ng/µL and chemocompetent TOP10 *E. coli* were thawn on ice. 1 µL DNA was transferred to an empty and prechilled RNase-free microfuge tube. 50 µL chemocompetent cells were added directly onto the plasmid DNA and incubated for 30 minutes on ice. Afterwards I heat shocked cells in a water bath at 42 °C for 30 seconds followed by incubation on ice for 5 more minutes. 200 µL SOC medium was added to the samples and cells were incubated for 1 hour at 37 °C, 200 rpm. 100 µL of the cell suspension was then plated on prewarmed agar plates, in the presence of the respective antibiotics, using glass beads.

II. Maintenance of chemocompetent *E. coli*

A 3 mL culture TOP10 *E. coli* was inoculated and incubated at 37 °C, 200 rpm, overnight. The next day the overnight culture was diluted 1:200 into 200 mL – 400 mL fresh LB medium. The cells were incubated at 37 °C and 200 rpm until an OD_{600} of 0.48 was reached, typically after 3 – 4 hours of incubation. The culture was then transferred into an appropriate amount of 50 mL centrifuge tubes and chilled on ice for 10 minutes. I pelleted the cells thereafter at 4 °C for 10 minutes at 3700 RCFmax. Supernatant was removed and cells were resuspended in 20 mL cold TFB 1 buffer for each 100 mL of bacterial culture before chilling the cells on ice for 5 minutes. Cells were pelleted once more through centrifugation and supernatant was discarded. For each 100 mL culture 4 mL TFB 2 buffer was added and I carefully resuspended the cells. Upon completion of another 15 min incubation on ice cells were carefully aliquoted in 100 µL units. 50 µL cells were immediately used for transformation (see above) using a control plasmid, to confirm chemocompetence of the freshly prepared cells.

III. Creation of glycerol stocks for long-term storage of transformed bacteria

I mixed 500 µL cultured cells, typically from an overnight culture, with 500 µL 50 % (v/v) glycerol. The resulting glycerol stock was then stored at – 80 °C.

IV. Preparation of plasmid DNA

I picked either a single colony or took a small number of frozen cells from a glycerol stock and transferred them into 3 mL of LB-medium containing the appropriate amount of antibiotic. The tube was then incubated for 12 - 16 hours at 37 °C and shaking at 200 rpm. Cells were pelleted and medium discarded. DNA was isolated following the manual of the NucleoSpin Plasmid Mini kit and DNA concentration was determined using a Nano photometer. 10 µL DNA with a concentration of 50 ng/µL was sent to Eurofins Genomics to confirm the integrity of the plasmid DNA via sequencing.

4.3.10 Yeast-Two-Hybrid assay

I. Double-Transformation

I double-transformed all interaction pairs into MaV203 yeast cells. Several yeast colonies were inoculated into 50 mL YPD medium and grown at 30 °C and 200 rpm overnight. The next day 30 mL overnight culture was added to 300 mL fresh YPD medium, in which cells were grown until an OD_{600} of 0.4 – 0.6 was reached. Cells were placed in 50 ml tubes and pelleted. The supernatant was discarded and the cells resuspended in sterile water and pooled into one tube. Afterwards, cells were spun down and supernatant was removed. Cells were resuspended in 1.5 mL 1x TE/1x Liciumacetate (LiAc, Sigma Aldrich). Meanwhile 100 ng of each interaction pair plasmid was mixed with 100 µg carrier DNA (Salmon Sperm DNA, Thermo Fischer Scientific). 100 µL yeast competent cells were added to each tube containing plasmid DNA and mixed well. 600 µL PEG/LiAc was added and each sample was vortexed at high speed for 10 seconds followed by 30 minutes incubation shaking with 200 rpm at 30 °C. Afterwards, 70 µL sterile DMSO was added and samples were mixed by inverting, followed by a heat shock performed for 15 minutes at 42 °C. Cells were then chilled on ice for 2 minutes and briefly spun down. Supernatant was discarded and cells readied for plating by resuspension in 500 µL sterile TE buffer (Sigma Aldrich).

II. Yeast-Two-Hybid

First, I grew freshly transformed cells were at 30 °C on agar plates lacking leucine and tryptophan (SC-Leu-Trp). After 2-3 days 3 colonies of each sample were resuspended in 100 µL sterile saline and stored in a 96-well plate. Using sterile 96-needle replicators cells were transferred onto rectangular SC-Leu-Trp agar plates additionally lacking histidine and containing differing amounts of 3-aminotriazol (3AT), typically between 10 – 50 mM. After 5 – 10 days I took photos and assessed the phenotype.

4.3.11 Protein Affinity Purification

RHOA pcDNA3.1/Zeo constructs with FLAG-tags were sent to Tübingen. There, Tobias Leonhard, Dr. Karsten Boldt and Prof. Dr. Marius Ueffing assisted me by performing protein affinity purification as previously described[106]. In short, constructs were overexpressed in HEK293T cells and after cell lysis proteins were enriched using the FLAG-tag. After elution and precipitation, the samples were subjected to mass spectrometric analysis in the Ueffing lab.

4.3.12 Protein Affinity Purification Analysis

The analysis of intensity values derived from protein affinity purification was analysed with a combination of python 2.7 and R. First, $\log_2$-transformed intensity values were used to calculate p-values between RHOA WT and variants, with intensity values of 0 set to 10^{-10}. I corrected these p-values using the Benjamini-Hochberg method[225]. Intensity values would only be kept and ratios between WT and mutants calculated when FDR ≤ 0.05 and not more than one zero was present in the WT or mutant intensity triplicates. Ratios produced this way could indicate weakened or strengthened interactions. However, since many cases surfaced where either the WT or mutant displayed triple zeroes, showing consistency within given samples, and at the same time demonstrating intensities in some samples. Such cases were manually given a value of -1, when a given interaction was seen in WT cells but not in variant overexpressing cell lines (‘interaction lost’), or a value of +5, when a given interaction was not seen in WT cells but observable in variant overexpressing cell lines (‘interaction gained’).

4.4 Results

4.4.1 RHOA variants in cancer show signs of tissue specificity

COSMIC registers several hundred missense variants for RHOA. The GTP-binding altering variant Gly17Val is the most frequent (with over 700 counts). There are many other, presumably LoF, variants scattered around the sequence. Generally, they cluster in the switch I (29-42) and switch II (62-68) regions with rarer variants (counts < 10) at the C-terminus (**Fig. 4C**, **A**).

RHOA variants are common in haematopoietic & lymphoid tissue and stomach cancers. Gly17 variants to Val, Leu or Arg are only found in haematopoietic and lymphoid tissue, while other variants affecting this position are seen in other cancers, for example the Gly17Ala in lung (3 times) and Gly17Glu in breast cancer (> 60 times) (**Fig. 4C**, **B**). This suggests tissue and cell type specific functions for RHOA and at least some of its pathogenic variants, despite RHOA not being differentially expressed amongst those tissues. There is, however, emerging evidence that RHOA expression varies dependent on cell type[63].

I was in particular interested in variants affecting RHOA protein-protein interactions, located primarily at the end of the switch I region. While they are seen far less often than the G17V variant, I hypothesised that the frequency and the variants tissue distribution would still be sufficient to deduce novel insights into the mechanisms of RHOA function. For example, Tyr42Cys was observed 54 times and Glu40Lys 6 times, with the remaining variants of interest (Tyr34Cys, Glu40Gln, Tyr42Ser and Leu69Arg) being observed between 10 – 20 times. These variants were chosen due to their respective frequency and based on amino acid properties. For instance, a mutation from glutamic acid into lysine replaces a negative with a positive charge. While glutamic acid does not appear to be conserved between proteins of the RHO family, the negative charge at this position is conserved, as CDC42 and RAC1 both have another negatively charged amino acid at this position, aspartic acid (**Fig. 4A**, **C**). Tyrosine residues are of special interest, as they might be the target of phosphorylation.

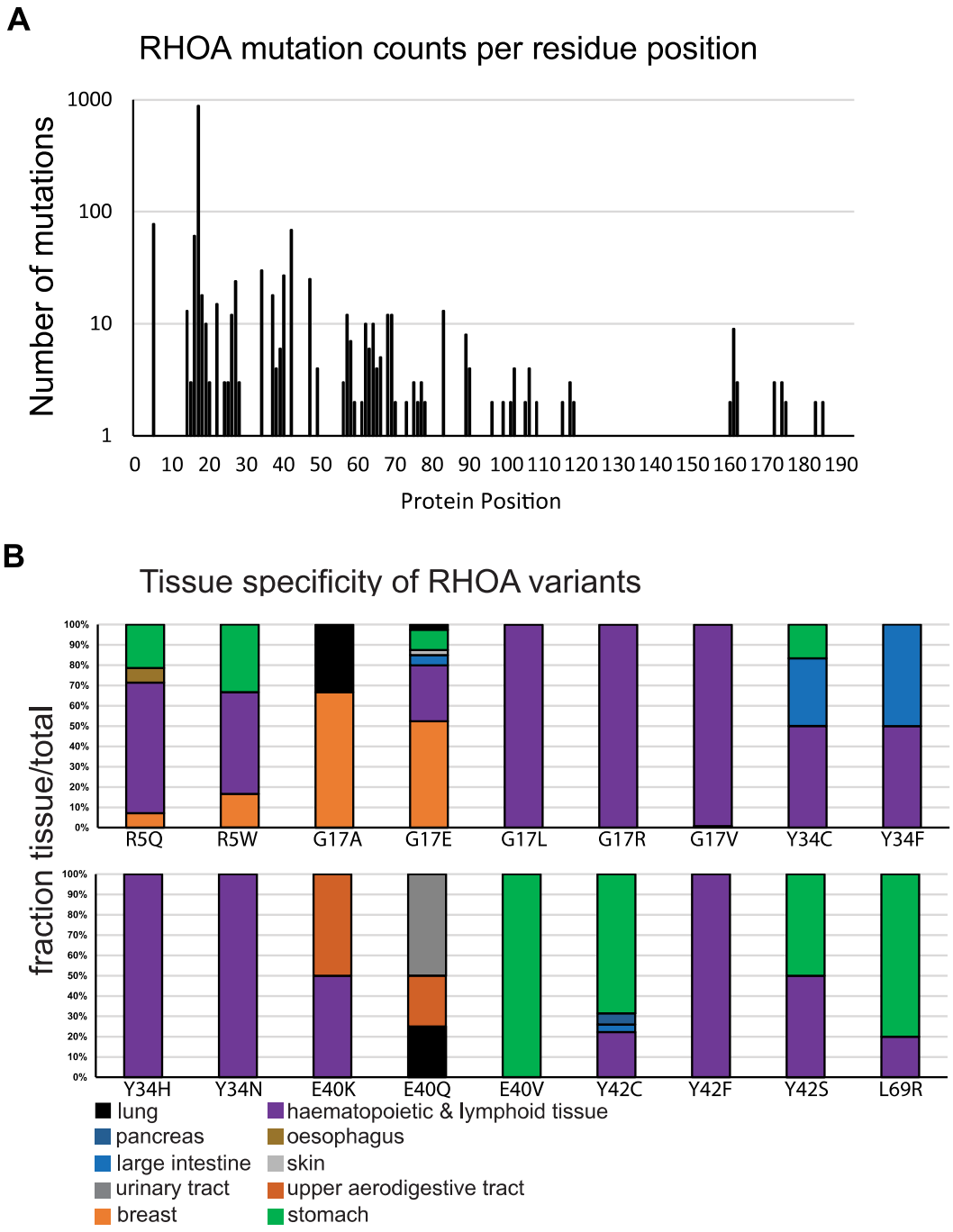


Figure 4C. Tissue Specificity of RHOA Variants in Cancer. A. Mutation frequency in cancer. Counts were taken from COSMIC and the sum of counts for all variants affecting a particular position are displayed. Y-axis is in logarithmic scale. **B. Tissue distribution of RHOA cancer variants.** Tissue distributions are shown for some RHOA variants, including different Gly17 variants. Amino acids are shown using the 1-letter code.

4.4.2 Canonically inactivating variants show heterogenous effects on protein-protein interactions

I predicted the impact of mutations on protein-protein (as well as protein-DNA and protein-chemical) interactions using Mechismo, a tool that uses sequence and 3D-structure information[55] to calculate a score that is positive when interactions are enabled, or negative when interactions are disabled. Surprisingly I found differences in how certain PPIs are affected (**Fig. 4D**).

For example, Tyr34Cys and Leu69Arg are clustered together with similar Mechismo scores, as both variants are largely disabling interactions with GEF (e.g. ARHGEF1, ARHGEF3, ARHGEF4, ARHGEF9, etc.) and GAP proteins (ARHGAP1, ARHGAP5, ARHGAP35, etc.). Tyr34Cys (and not Leu69Arg) is additionally predicted to disable the interaction with ARHGAP15 and Leu69Arg (and not Tyr34Cys) to disrupt interactions with PAK and DIAPH effector proteins. Leu69Arg is also predicted to enable interactions with the effectors ARHGAP20 and DOCK.

Tyr42Cys and Tyr42Ser are predicted to be very similar, both only weakly perturbing a handful of interactions. Tyr42Ser nevertheless has a stronger predicted disabling effect on ITSN1/2, while Tyr42Cys more strongly disables DOCK effectors.

As might be expected owing to the reversal of charge Glu40Lys has more extreme predicted perturbations compared to the Glu40Gln variant. There are multiple enabled PPIs for both variants including RAC1/2/3 or ROCK1/2 (Glu40Gln). The interaction between RHOA Glu40Lys and ARHGEF25 is also predicted to be disabled, but predicted to be unperturbed for Glu40Gln. While both variants seem to be disabling towards ITSN1 and ITSN2 the Glu40Lys variant displays the strongest disabling scores of all the PPIs observed (between -6 and -8).

This analysis suggests that RHOA variants display a heterogenous pattern of how they affect interactions between RHOA and its partner proteins (**Table 7**). While the observed phenotype of many RHOA variants might suggest an inactivated protein, these findings indicate that the underlying mechanisms might be different, depending on cell type, tissue and ultimately on the RHOA variant.

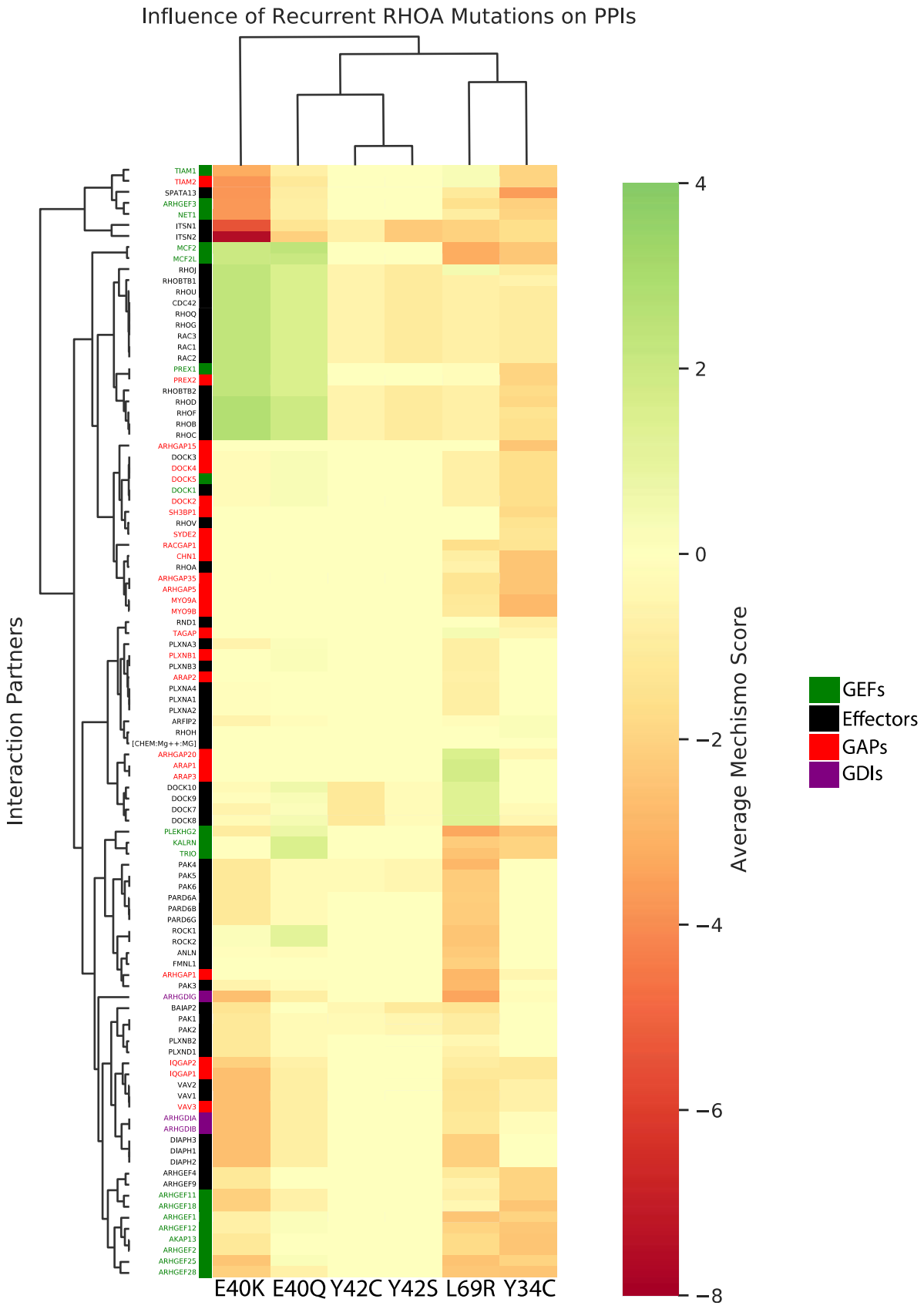


Figure 4D. *In Silico* Analysis of perturbed RHOA PPIs. Indicated RHOA variants were subjected to analysis with Mechismo3. Cell colours correspond to the average Mechismo score

(if an interaction between a RHOA variant and another protein involve multiple residues), with negative scores suggesting a disabled and positive score predicting enabled interactions.

Table 7: Summary of predicted perturbed PPIs.

Variant	Enabling	Disabling
Tyr34Cys		GEFs, GAPs, ARHGAP15
Glu40Lys	RAC1/2/3	ARHGEF25, ITSN1/2
Glu40Gln	RAC1/2/3, ROCK1/2	ITSN1/2
Tyr42Cys		DOCKs
Tyr42Ser		ITSN1/2
Leu69Arg	ARHGAP20, DOCKs	GEFs, GAPs, PAK, DIAPH

Multiple occurrences of interacting proteins are highlighted by colour.

4.4.3 RHOA variants do not influence cell proliferation in osteosarcoma cells

Cell proliferation is a hallmark of cancer[226] and assessing the impact of RHOA variants on proliferation is an important marker for further downstream experiments. I transfected U2OS cells (see 4.3.2.II) and seeded overexpressing cell lines into 6-wells. Transfected cells displayed a tendency to die within few days when under selection pressure for more than a week. Notably control cells would have already succumbed to the antibiotic stress and therefore a failed transfection and hence insufficient resistance to the selective antibiotic is unlikely to be an explanation for this sudden death. Surviving cell lines also only show relatively weak RHOA overexpression (Appendix 8.4). These both suggest that overexpressing RHOA might have cytotoxic effects, ultimately killing cells.

I determined cell numbers in 24 h intervals for a total of 72 h (see 4.3.1 II). Due to the stability of siRNA-mediated RHOA knockdowns it was only possible to count cell numbers over the course of 48 h (Appendix 8.4) when knocking down RHOA. To make results more reliable one well of a 6-well plate was used per time point, and the starting cell number used to normalize the cell count at given time points was determined immediately after seeding.

Proliferation of U2OS cells was not perturbed when wildtype RHOA (WTOE) was overexpressed. Likewise, temporary removal of RHOA expression did not affect cell proliferation over the observed time frame (**Fig. 4E**, top left). While a slight tendency to slower cell proliferation at 72 h after seeding could be seen for cells overexpressing

RHOA Tyr34Cys, the effect was minor and not significant (**Fig. 4E**, top centre). I observed no differences for cells expressing RHOA Glu40Lys or Glu40Gln, neither when compared to WT or knockdown nor when compared to each other (**Fig. 4E**, top right). Furthermore, I saw no statistically significant changes in cell proliferation when overexpressing RHOA Tyr42Cys or Tyr42Ser (**Fig. 4E**, bottom left) or Leu69Arg (**Fig. 4E**, bottom right).

While these results were all negative in nature for the time frame observed, they meant that, when performing gap-closing assays, any observed effect on cell velocity would unlikely be related to altered cell proliferation.

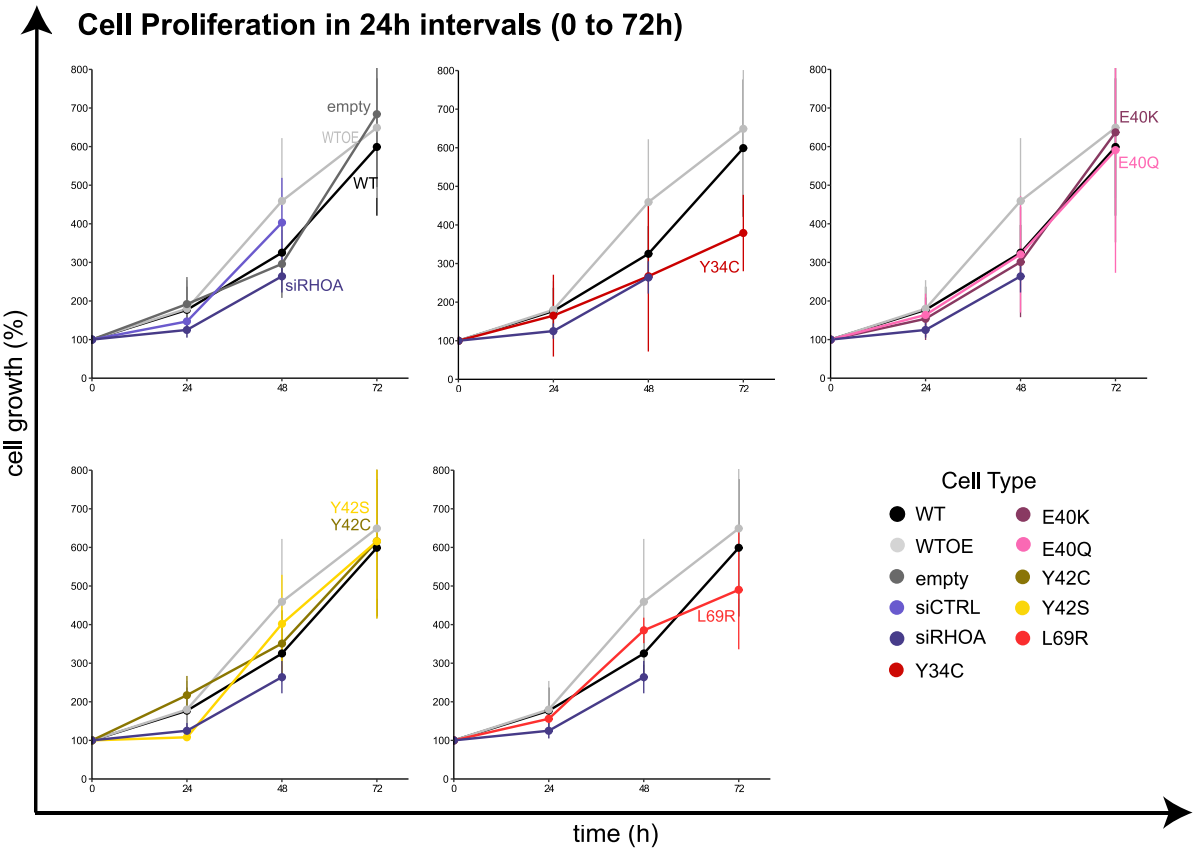


Figure 4E. Cell Proliferation Assay. Cells were seeded and cell numbers were determined at 24h, 48h and 72h and normalized to the starting cell number seeded into the well. Each subplot compared the indicated cell lines against WT, WTOE and RHOA knockdown cells. N = 3.

4.4.4 Cell velocity is perturbed in RHOA variants

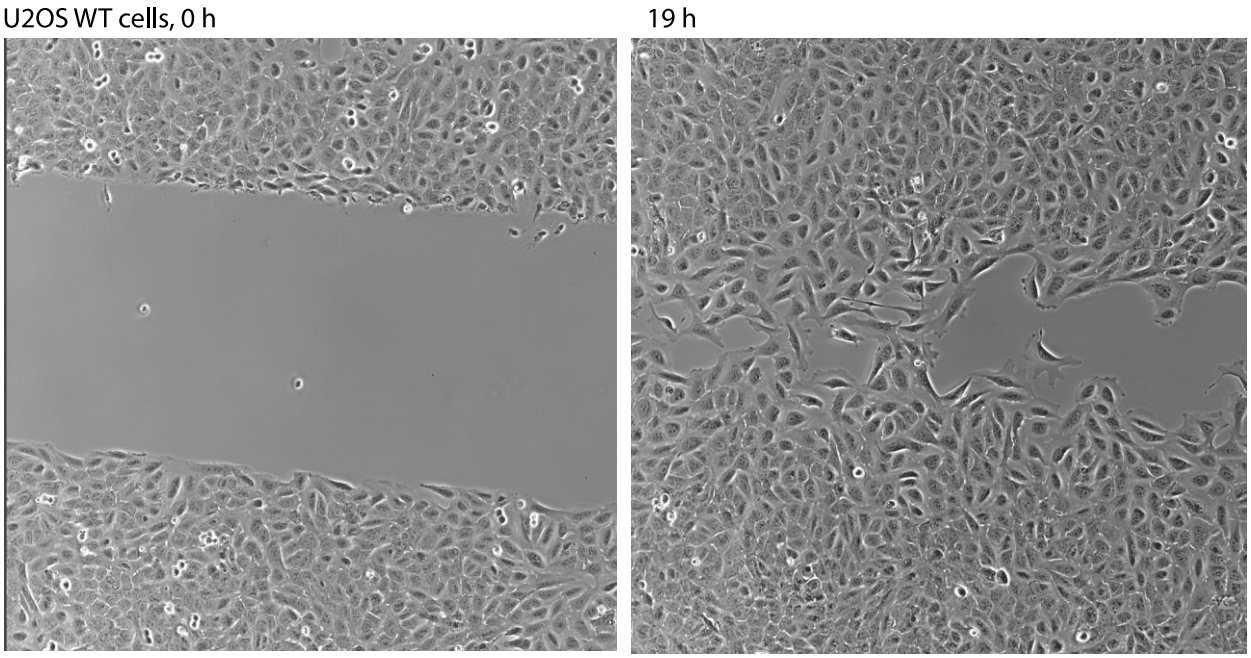
I seeded osteosarcoma cells into ibidi culture-insert 3 well dishes (**Fig. 4F**, top left) and imaged every 5 minutes for at least 19 hours (**Fig. 4F**, top right; also see 4.3.8). The average gap closing speed of untransfected U2OS WT cells was measured as 0.41 µm/min and lies within the typical range reported before (0.21 – 0.49 µm/min[227]).

Cells overexpressing wildtype RHOA displayed a gap closing speed reduction down to 29.6 % compared to WT cells and repeatedly failed to close the gap when observed for up to 23 hours. This effect was rescued in RHOA Tyr34Cys and Leu69Arg overexpressing cells, with gap closing speeds of 87.7 % and 104.3 %, respectively. Since the siRNA-mediated knockdown of RHOA was stable for at least 72 hours it was possible to include this condition for this analysis. Similar to the effects seen in RHOA Tyr34Cys and Leu69Arg cells, a WT-like phenotype with a gap closing speed of 115.9 % was observed when RHOA was depleted. Overexpression of variants Glu40Lys or Glu40Gln showed a slight, but significant, increase in cell motility to circa 170% with no difference between the two. I made similar observations for Tyr42Cys or Tyr42Ser with significant increases in cell velocity up to 154 % and 136 %, respectively.

I was able to see a phenotype when WT RHOA was overexpressed but not when RHOA was either knocked down or the variants Tyr34Cys or Leu69Arg were overexpressed. Expression of either Glu40Lys, Glu40Gln, Tyr42Cys or Tyr42Ser showed neither WT or overexpressed-WT behaviour, instead, these mutants displayed a slight but significantly enhanced ability to close the gap over the course of imaging (**Table 8**).

Table 8: Summary of gap-closing speeds.

Cells	Glu40 Lys	Glu40 Gln	Tyr42 Cys	Tyr42 Ser	siRHOA	Leu69 Arg	WT	Tyr34 Cys	WT OE
rel. Speed (%)	170	170	154	136	116	104	**100**	88	30



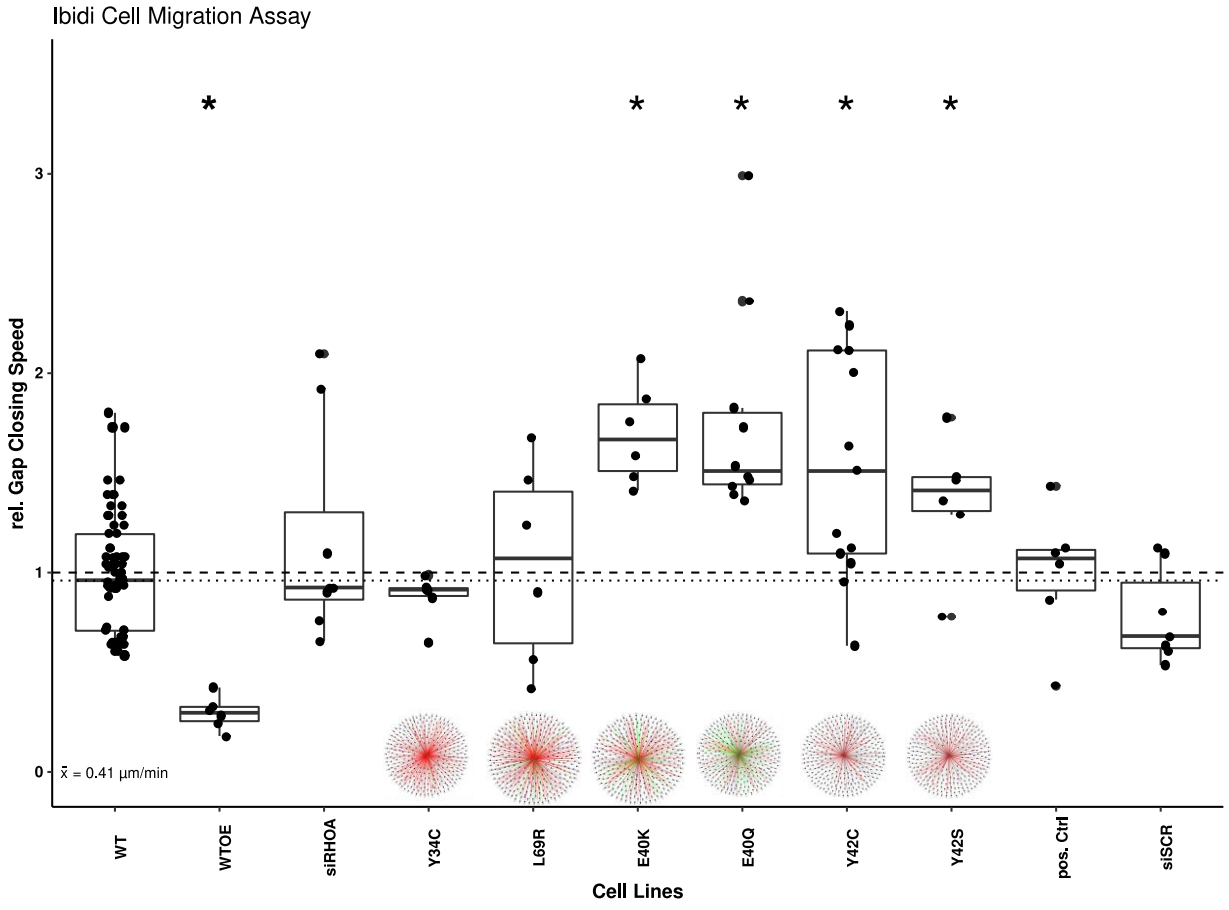


Figure 4F. Gap Closing Assay and Cell Velocities. Top left, wildtype U2OS cells at the beginning of the experiment. The size of the gap, introduced through the ibidi dish setup, can be measured within the Fiji software due to internal calibrations of the Nikon Ti-HCS microscope. **Top right,** wildtype U2OS cells after 19 hours of incubation under typical cell culture conditions (5 % CO_2, 37 °C), with most of the gap closed. **Bottom,** Cell velocities of the indicated cell lines

(x-axis) are compared against the average cell velocity of U2OS WT cells, set to 1 (i.e. 100 %). The fine-dashed line indicates the median of all U2OS WT cells while the dashed line indicates their average gap-closing speed. Mechismo graphical output is shown below to indicate the severity of predicted disabling effects.

4.4.5 RHOA knockdown can perturb phosphorylation events at focal adhesions

RHOA is a major player in cellular actin dynamics, especially due to its antagonistic relationship with RAC1, where anterior RAC1 activity counteracts posterior RHOA activity[228]. RHOA is involved in the formation and maintenance of focal adhesions and stress fibres, playing an important role during directed cell movement (**Fig. 4G)**. A previous study was able to detect changes to focal adhesions in RHOA knockout fibroblasts[229], as focal adhesions are heavily linked to actin dynamics[230] and a frequent location of phosphorylation, for example through focal adhesion kinase-1 mediated paxillin phosphorylation[231].

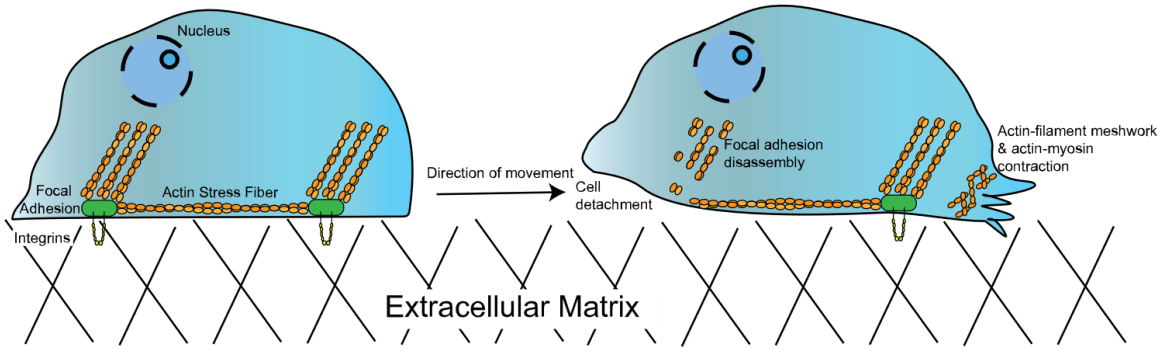


Figure 4G. Focal adhesions play an important role in cell locomotion. The assembly and disassembly of focal adhesions is dynamically regulated by RHOA in interplay with RAC1, and while assembled focal adhesions mediate the contact of cells to the ECM, it is their disassembly that allows cells to move upon anterior actin-myosin contraction.

The ability of cells to move in a given direction is key for cell migration, invasion and the formation of metastases, a hallmark of cancer[226]. To assess this, osteosarcoma cells were grown on ibidi dishes filamentous actin (F-Actin) was stained with Phalloidin-Atto 590. Phosphorylation events were detected with a FITC-linked primary antibody directed against pTyr20.

I detected bright focal adhesions (**Fig. 4H,** white arrows) near the plasma membrane of U2OS cells, often linked to and present at the terminal tips of actin fibres. RHOA WTOE cells displayed a relative mean CMPF of 88 % compared to WT cells (**Fig. 4H**, row 3 and 4, and **Fig. 4I**, top). A decrease in relative CMPF was measurable in RHOA knockdown (CMPF = 55.9 %) and Tyr34Cys overexpressing cells (CMPF = 61.7 %) (**Fig. 4H**, row 5 to 8, and **Fig. 4I**, top). Additionally, knockdown cells appeared slightly bigger than WT (**Fig. 4I**, centre) and displayed fewer pTyr signals. While Tyr34Cys cells seemed unaffected in terms of size they also showed fewer pTyr signals (**Fig. 4I**, bottom).

Overexpression of either RHOA Leu69Arg, Glu40Lys, or Glu40Gln did not significantly change the brightness of pTyr signals at focal adhesions, but the Tyr42Cys variant showed a minor reduction in pTyr brightness (CMPF = 80.8 %).

Taken together, I could show that U2OS cells overexpressing either RHOA Tyr34Cys or Tyr42Cys mutants produce tyrosine phosphorylation events at focal adhesions slightly less pronounced than WT, mimicking the phenotype observed when RHOA was knocked down (**Table 9**).

Table 9: Summary of pTyr brightness.

Cells	**Glu40 Lys**	**Leu69 Arg**	**WT**	**Glu40 Gln**	**WTOE**	**Tyr42 Cys**	**Tyr34 Cys**	**siRHOA**
CMPF (%)	129	111	**100**	91	88	81	62	56

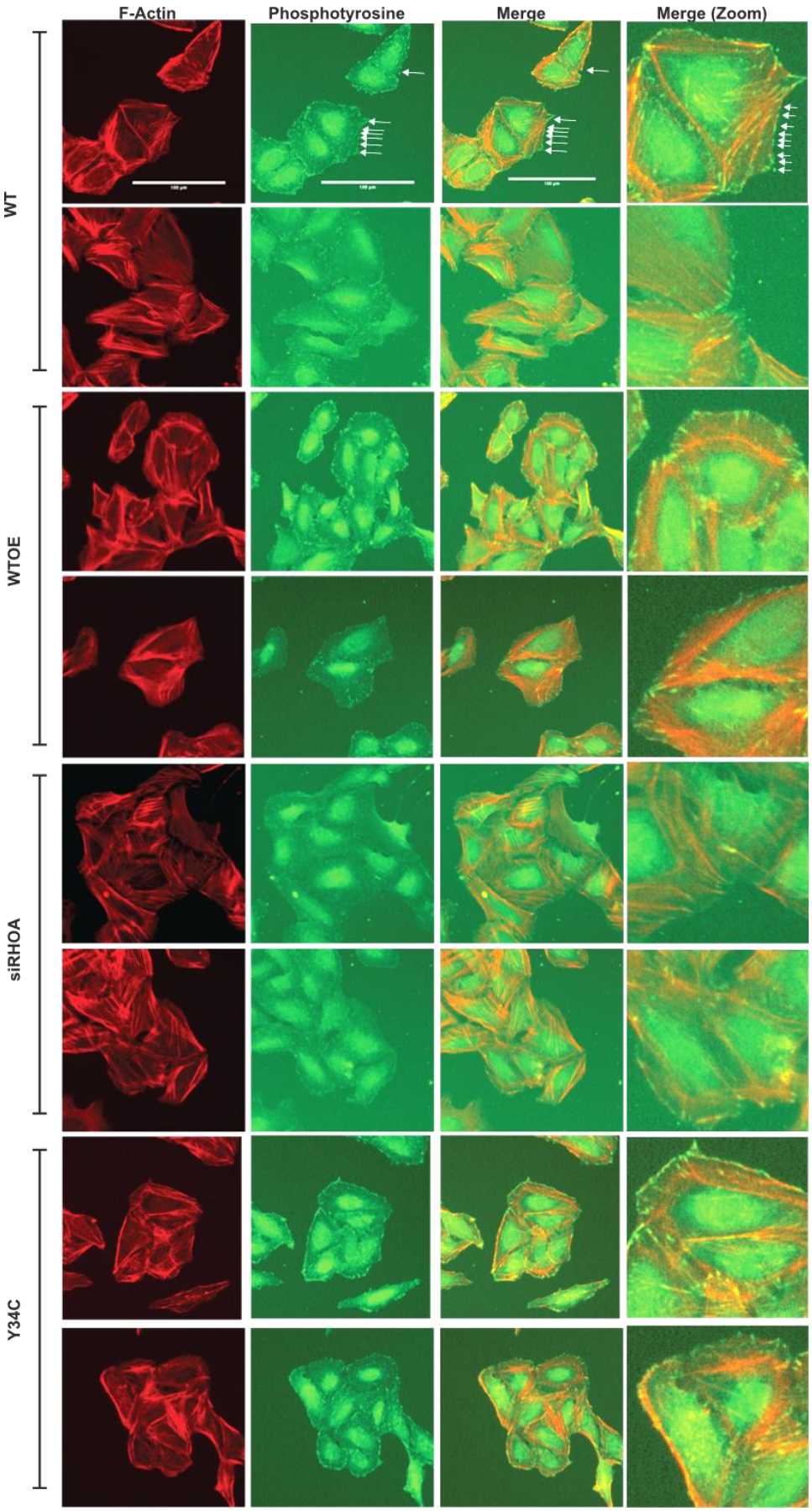
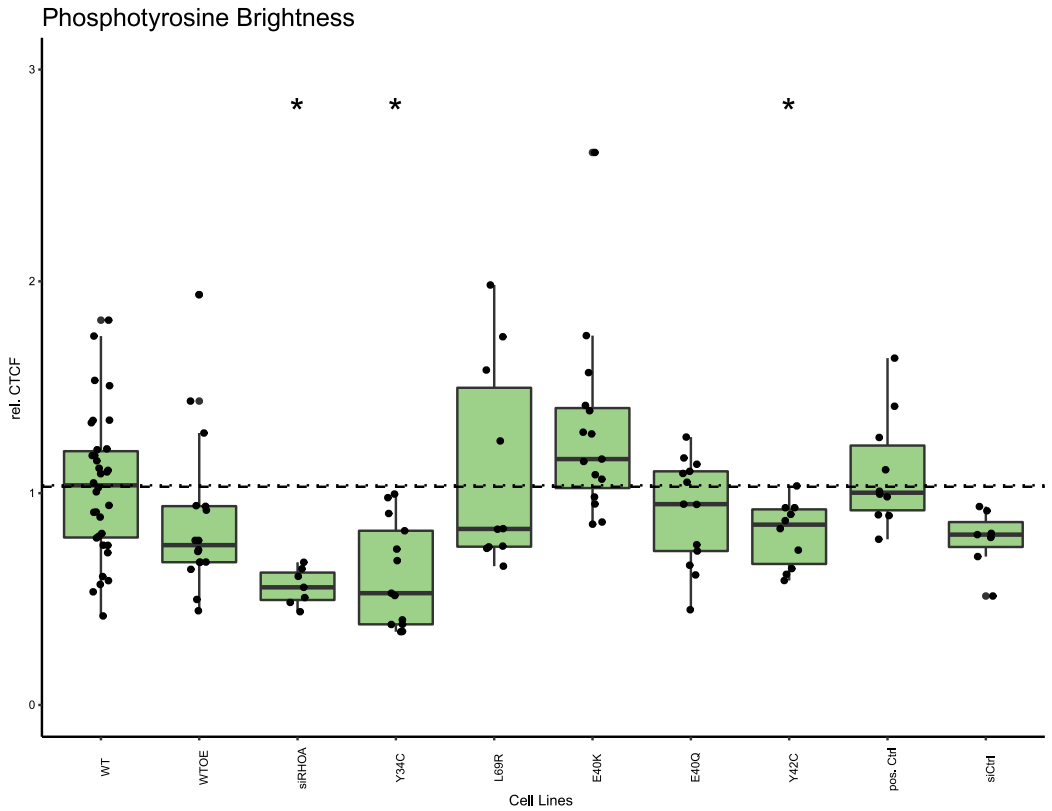
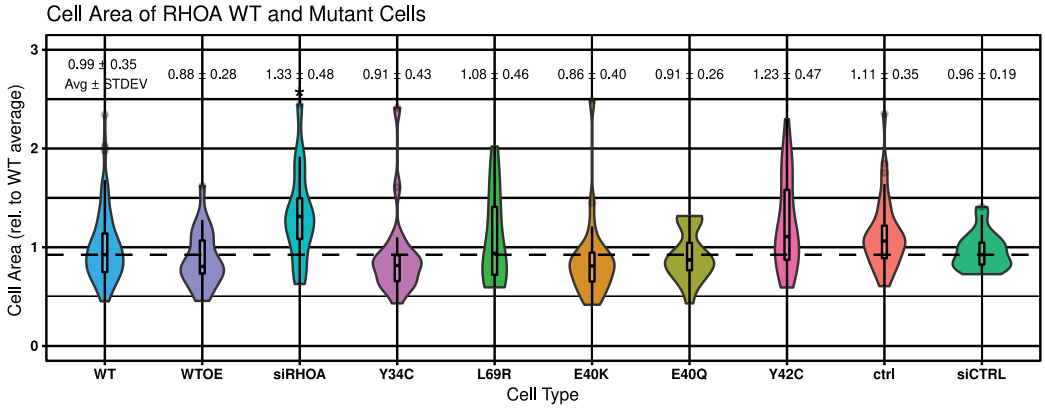

F-Actin
Phosphotyrosine
Merge
Merge (Zoom)
WT
WTOE
siRHOA
Y34C

Figure 4H


Phosphotyrosine Brightness
rel. CTCF
Cell Lines
WT
WTOE
siRHOA
Y34C
L69R
E40K
E40Q
Y42C
pos. Ctrl
siCtrl


Cell Area of RHOA WT and Mutant Cells
0.99 ± 0.35
Avg ± STDEV
0.88 ± 0.28
1.33 ± 0.48
0.91 ± 0.43
1.08 ± 0.46
0.86 ± 0.40
0.91 ± 0.26
1.23 ± 0.47
1.11 ± 0.35
0.96 ± 0.19
Cell Area (rel. to WT average)
WT
WTOE
siRHOA
Y34C
L69R
E40K
E40Q
Y42C
ctrl
siCTRL
Cell Type

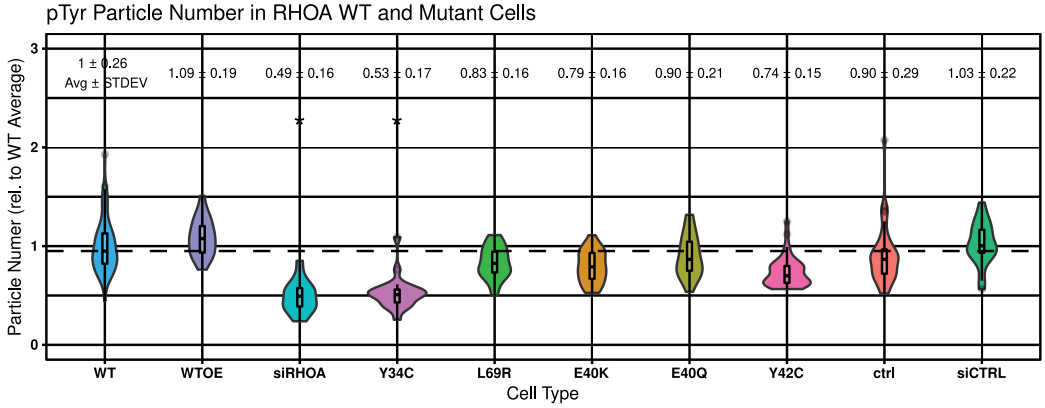

pTyr Particle Number in RHOA WT and Mutant Cells
1 ± 0.26
Avg ± STDEV
1.09 ± 0.19
0.49 ± 0.16
0.53 ± 0.17
0.83 ± 0.16
0.79 ± 0.16
0.90 ± 0.21
0.74 ± 0.15
0.90 ± 0.29
1.03 ± 0.22
Particle Numer (rel. to WT Average)
WT
WTOE
siRHOA
Y34C
L69R
E40K
E40Q
Y42C
ctrl
siCTRL
Cell Type

Figure 4I

Figure 4H. Fluorescence staining of phosphotyrosines at focal adhesions. U2OS cells were stained with Phalloidin-Atto 590 (F-Actin, red) and anti-pTyr20-FITC (green) as described in 4.3.7. Focal adhesions are visible as bright green spots close to the plasma membrane and often at the end of actin fibres (white arrows). Camera: Nikon DS-Qi2 Direct, Microscope: Nikon Ti2 (Brightfield & Widefield Fluorescence), NA = 0.75. Exposure times: 100 ms (Actin) and 3 s (pTyr20).

Figure 4I. Brightness and number of phosphotyrosines depend on RHOA status. Top, phosphotyrosine brightness in relative CMPF measured per cell as described in 4.3.7 (page 68). Dashed line indicates the median wildtype value. **Centre**, cell area determined using the FIJI software and manually measuring the area in μm^2, with the WT average set to 1. Dashed line indicates the median WT value. **Bottom**, number of phosphotyrosine spots per cell. Overlapping spots were counted as 1 unless the number of overlapping focal adhesions could be clearly distinguished. Dashed line indicates the median WT value.

4.4.6 Gene expression of RHOA knockdown cells differs from gene expression of RHOA variants

In order to learn more about the downstream effects of different mutants and why some RHOA variants only sometimes produce a phenotype mimicking that of RHOA knockdown cells, I performed gene expression analysis using qRT-PCR. After consultation of the literature, I chose to monitor the gene expression of WT and RHOA knockdown cells for RHOA, RAC1, ROCK1, CDC42, ARHGDIA, DIAPH3, IGF1, TNFα, CCND1, NFκβ, SOX9, HIF1α and c-MYC[232,233].

The gene expression data was unaltered for all but three genes: CDC42, ARHGDIA and CCND1. qRT-PCR was performed for these genes using cDNA prepared from RNA of the various RHOA overexpressing cell lines described above.

Gene expression for ARHGDIA roughly doubled in siRHOA, Leu69Arg and Tyr42Cys cells but remained around WT level for the remaining investigated cell lines (**Fig. 4J**, left). For CCND1 an increase up to 266 % could be determined for siRHOA cells, with 179 % for Tyr42Cys being second (**Fig. 4J**, centre). Lastly, a reduction of CDC42 gene expression was observed for siRHOA cells, while CDC42 went up to 176 % in RHOA Tyr42Cys cells.

Overall, I found that RHOA knockdown affects the gene expression of ARHGDIA, CCND1 and CDC42, with only the Leu69Arg and Tyr42Cys overexpressed variant cell lines could mimic some of these effects.

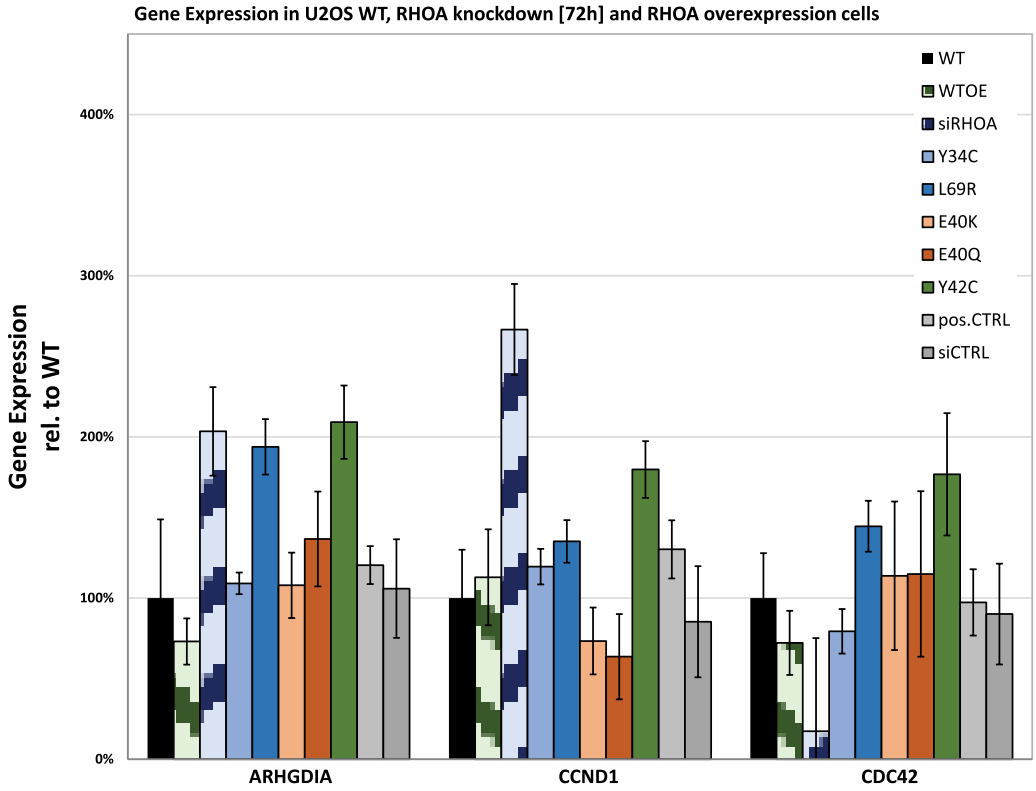


Figure 4J. Gene Expression Analysis in U2OS cells with RHOA knockdown or overexpression. U2OS cells total RNA was subjected to qRT-PCR analysis. Determined ∆ct-values were used to compare U2OS WT gene expression to either RHOA knockdown or RHOA overexpressing cell lines, with GAPDH as a house-keeping gene. WT expression was set to 100 %. N = 3

4.4.7 RHOA protein-protein interactions are differentially perturbed between variants

A selection of putative interaction partners was drafted based on an earlier Mechismo analysis (similar to **Fig. 4D**), where I included proteins of the GEF (ARHGEF6/12/25), GAP (ARHGAP20), GDI (ARHGDIA) families, as well as interesting effector proteins, either selected due to their functional role within RHOA signalling pathways (ROCK1,

DIAPH1) or due to predictions of perturbed interactions (PAK1, ITSN2, DOCK7). DNA constructs with added FLAG-tags and sent a selection of candidates (Tyr34Cys, Glu40Lys, Tyr42Cys and Leu69Arg) to the Ueffing lab in Tübingen, where Tobias Leonhardt performed protein affinity purification experiments with HEK293 cells overexpressing the respective RHOA variants (**Fig. 4K**).

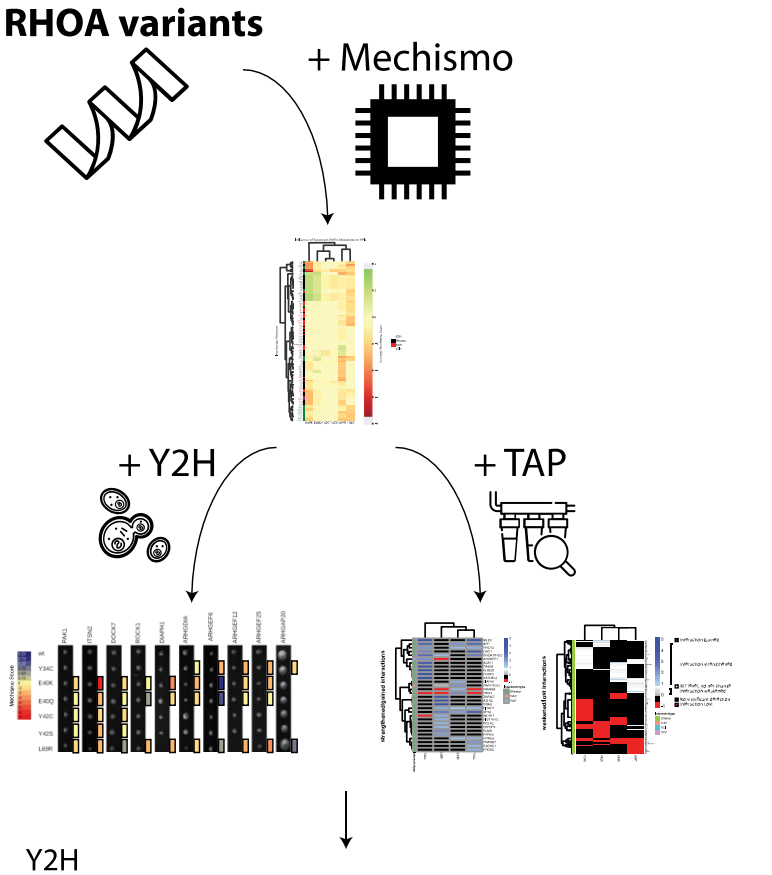
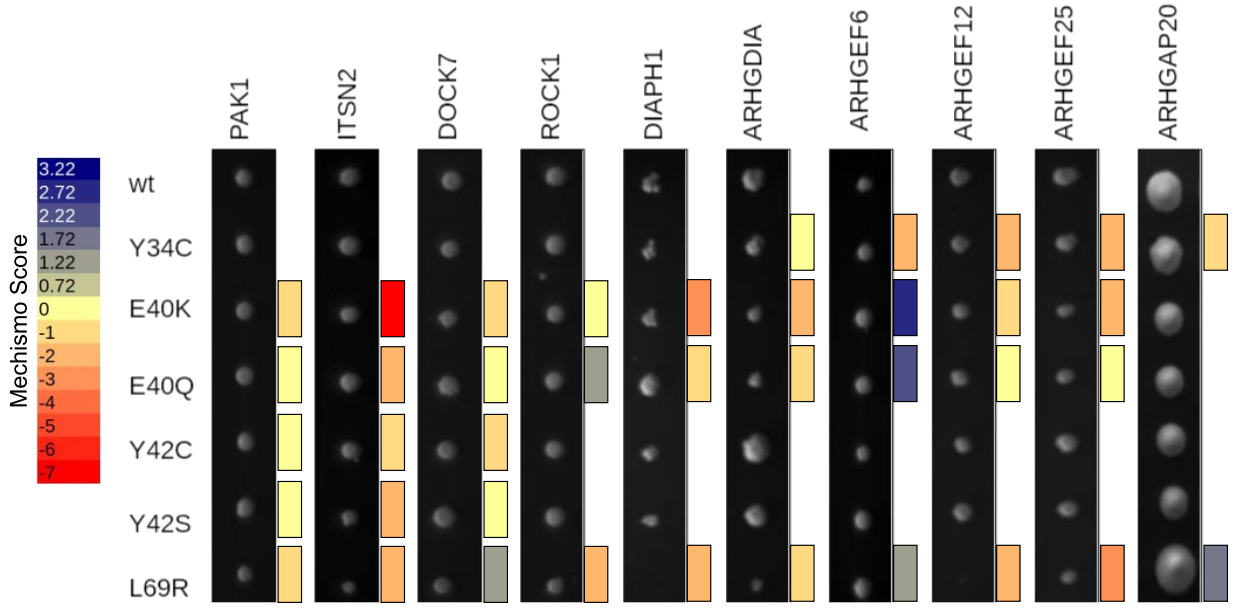


Figure 4K. Experimental Overview of RHOA PPI Exploration. Several PPIs were specifically investigated in Y2H based on earlier *in silico* findings. A more holistic approach was conducted by tandem affinity purification of RHOA and interaction partners.

Difficult to measure colony sizes and the tendency that they do not correlate well with the strength of given PPIs render interpretation of Y2H results a challenge[234]. Nonetheless,

the Leu69Arg variant displayed a decrease in colony size matching my predictions in many instances. Additionally, both Glu40Lys and Tyr42Ser affect the PPI with ITSN2 (but not Glu40Gln and Tyr42Cys, **Fig. 4L**)



Variant	Enabled	Disabled
Tyr34Cys	-	-
Glu40Lys	-	ITSN2, ARHGDIA, ARHGAP20
Glu40Gln	DIAPH1	ARHGDIA, ARHGAP20
Tyr42Cys	ARHGDIA	ARHGAP20
Tyr42Ser	ARHGDIA	ITSN2. ARHGAP20
Leu69Arg	ARHGEF6, ARHGAP20	PAK1, ITSN2, DOCK7, ROCK1, DIAPH1, ARHGDIA, ARHGEF12, ARHGEF25

Figure 4L. Yeast-Two-Hybrid assays. (**top**) Example colonies are shown for each interaction between RHOA and its putative interaction partners. Predicted Mechismo scores are displayed next to the respective colonies (right side, coloured rectangles), with negative scores suggesting a disabled interaction, and positive scores hinting at enabled interactions. N = 3. (**bottom**) Tabular summary of Y2H findings.

TAP results showed that Tyr34Cys lost the most interactions with 97 lost interactions compared to WT (i.e. interactors seen only in WT and not in the variant). While Glu40Lys

and Tyr42Cys showed a number of lost interactions that was higher than the observed number for Leu69Arg, they were significantly lower than what had been seen in Tyr34Cys with 61 lost interactions for both variants and 49 lost interactions for Leu69Arg (**Fig. 4M**, top, red cells).

It was more or less equally rare for variants to gain interactions (i.e. interactors seen only in variants and not in WT). Glu40Lys gained none, Leu69Arg gained three (ASHL2, GBAS and ZNF207), Tyr34Cys one (MLEC) and Tyr42Cys five (DUSP3, KATNAL2, KDM3B, MLEC and RAB18, **Fig. 4M**, bottom).

The WDR17 interaction is notable because it is lost in Tyr42Cys yet possibly strengthened in Leu69Arg. The interaction with FAM65A, also known as Rho family-interacting cell polarization regulator 1 (RIPOR1), is lost in both Leu69Arg and Tyr34Cys, but slightly enhanced in Glu40Ls. Similarly, the interaction with Rhotektin (RTKN) is strengthened in Glu40Lys but lost in Tyr34Cys, Tyr42Cys and Leu69Arg. Glu40Lys strengthens the interaction with RAP1GD1, whereas Leu69Arg weakens it (**Table 10**).

Taken together, these results demonstrate the highly heterogenous nature of phenotypes observable in RHOA variants. While this does not contradict the canonical opinion that these RHOA variants, which are also seen in human health and disease, are inactivating, it does raise the question on whether the inactivating mechanism behind these variants is the same, and if the mechanism is more than just simply destabilizing the protein, as is suggested to be the main mechanism behind disease-associated mutations[34]. It is clear that certain variants (e.g. Leu69Arg) are largely distinct from others (e.g. Tyr42Cys, Tyr34Cys, **Fig. 4N**).

Table 10: Summary of TAP results.

Variants	Tyr34Cys	Glu40Lys	Tyr42Cys	Leu69Arg
Lost PPI	97	61	61	49
Gained PPI	1	0	5	3

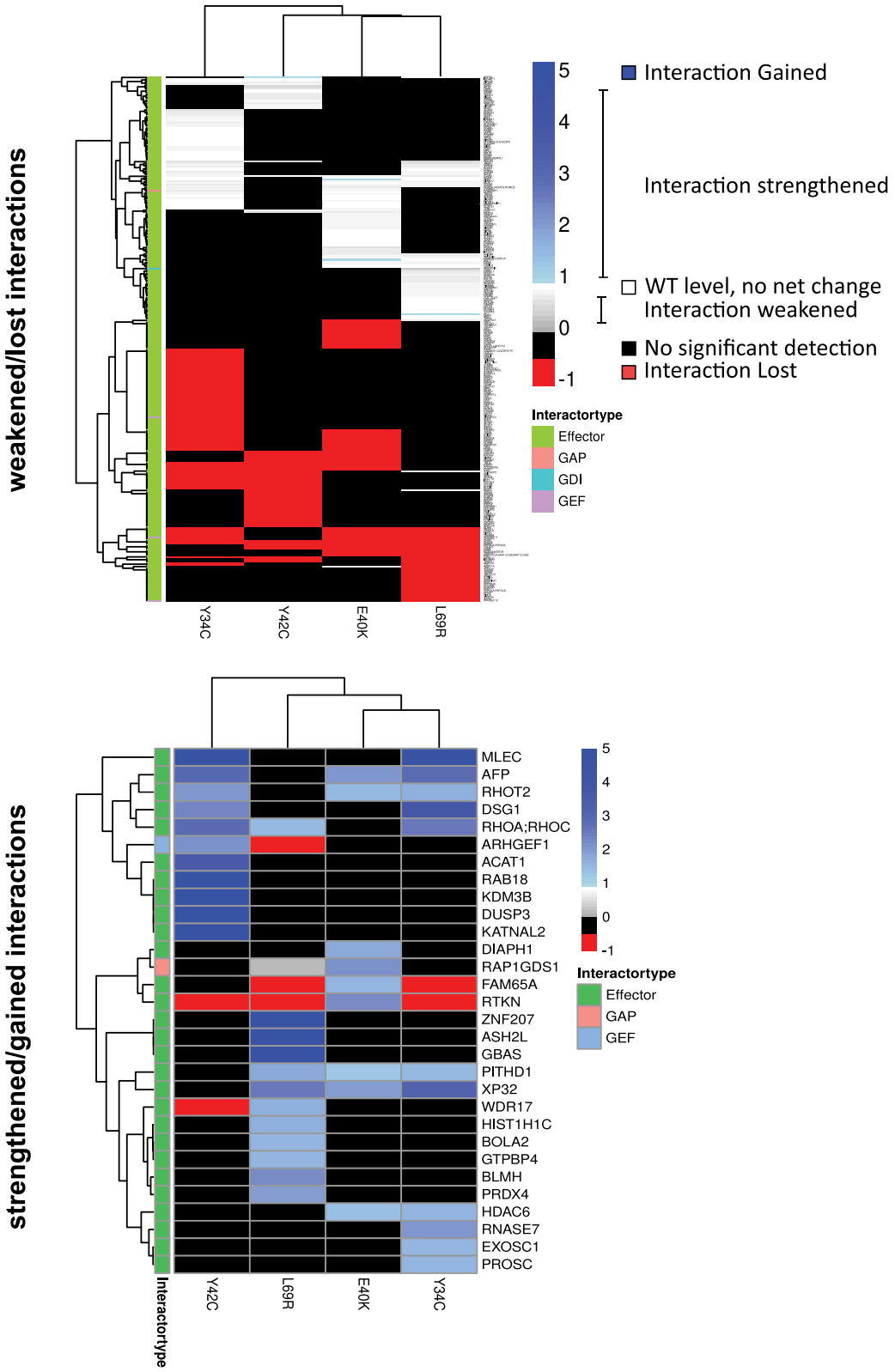


Figure 4M. Protein Affinity Analysis. Top, Cluster map showing primarily weakened or lost interactions derived from protein affinity data and calculated as described in 4.3.12. Cell colours: Black= non-significant detection; grey= weakened interaction; red= lost interaction; blue= strengthened or gained (dark blue) interactions. Clustering of interaction partners and RHOA

variants along the y- and x-axis was performed using the k-means clustering algorithm within the R-package pheatmap[235] . **Bottom,** Cluster map showing primarily strengthened or gained interactions.

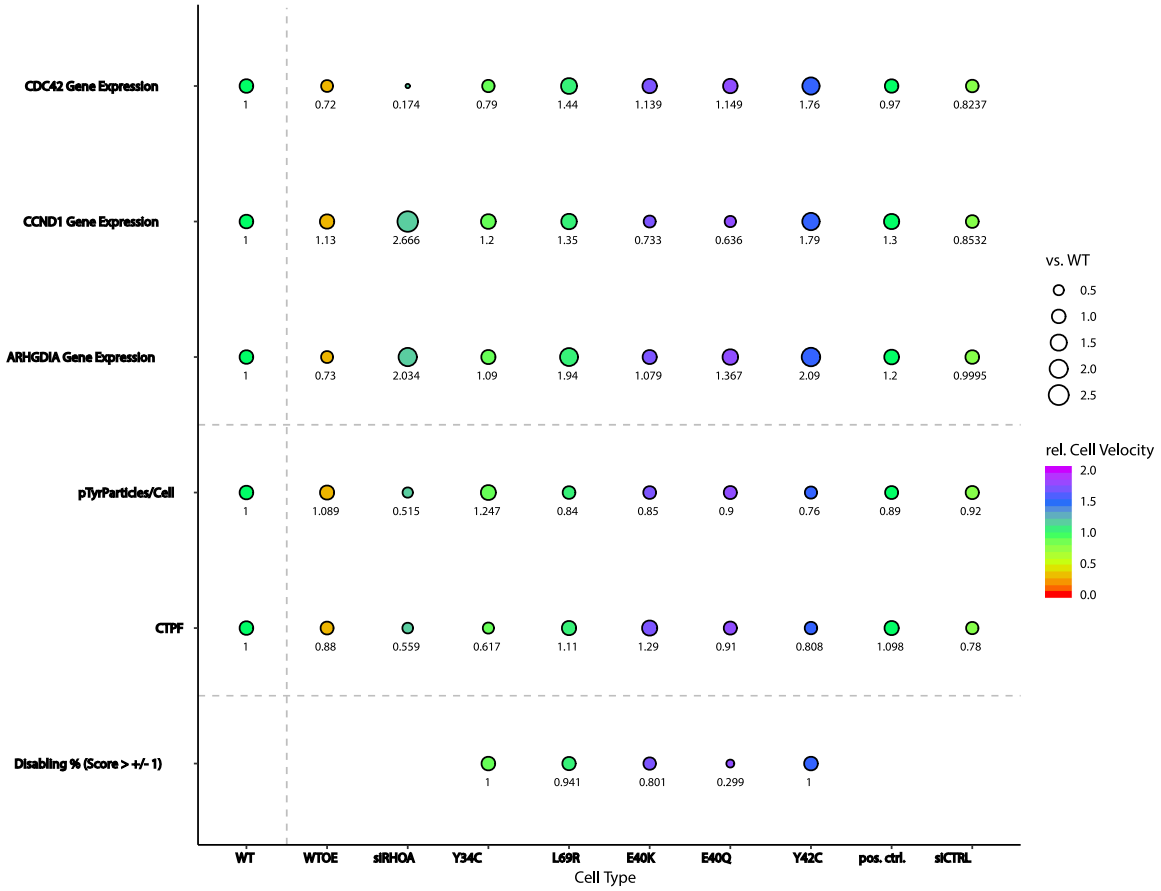


Figure 4N. Summary of RHOA Experiments. This summary shows results from several experiments (gap-closing assay, gene expression, pTyr phosphorylation).

4.5 Discussion

4.5.1 The role of RHOA in cancer and the impact of RHOA variants on PPIs

The small GTPase RHOA has many functions that are both crucial for normal tissues, as well as for cancer. These include cell migration, polarization, proliferation and survival, and RHOA is putatively involved in all stages of cancer progression.[224,236] However, RHOA variants were not thought of as major cancer drivers until recently, when the protein was found to be mutated in several malignancies, including gastric cancer[237,238], breast cancer[211] or T-cell lymphoma[239]. Nonetheless perturbed RHOA expression appears to be prognostic for some, but not all, cancers. Overexpression, for example, is

associated with poor survival in metastatic colorectal cancer[240], while reduced expression seems to enhance breast cancer tumorigenesis[241].

The findings presented here not only show that somatic RHOA mutations are frequently seen in cancer (**Fig 4C**, **A**), but that there is also specificity to which mutations are seen in which cancer type (**Fig 4C**, **B**). This, taken together with the observation that perturbation of RHOA gene expression can be linked to cancer progression highlights the duality of RHOAs importance for human health and disease and asks the question whether RHOA does have tumour suppressive or oncogenic functions[211].

A possible explanation is that RHOA acts as a hub protein, that is, a protein acting as a central focal point for many different PPIs[242]. From the variety of functions RHOA is involved in it can be assumed that this would include a number of different PPIs.

It was thus of interest to investigate whether the target RHOA variants (Tyr34Cys, Glu40Lys/Gln, Tyr42Cys/Ser and Leu69Arg) would indeed demonstrate differing phenotypes, especially since RHOA variants are often hypothesized to be LoF variants[243,244].

In silico analyses found that Leu69Arg and Tyr34Cys are quite similar in their predicted effects on PPIs and indeed I found many perturbations between these RHOA variants and GEF or GAP proteins. This is surprising, as Tyr34Cys is a phosphosite, mediating the interaction between RHOA and effector proteins, while Leu69Arg is mostly suggested to affect interaction with GEFs[206]. Especially the putative effect of these variants on both GEFs and GAPs is interesting, as GEF proteins are generally required for small GTPase activation, while GAP proteins cause a shift towards the inactivated GDP-bound state. These effects seem quite contrary, as a disabling effect between RHOA-GEFs should lead to less active RHOA molecules (i.e. causing a quasi-LOF phenotype), while a disabled binding between RHOA and GAPs would perhaps accumulate (activated) RHOA-GTP. It is noteworthy that some presumed perturbations for Leu69Arg could be confirmed in Y2H assays, including the disabling effect on ITSN2, DIAPH1, ARHGDIA, ARHGEF12 and ARHGEF25 as well as enabling effects on DOCK7, ARHGEF6 and ARHGAP20, giving credibility to these *in silico* results.

RHOA Glu40 mediates the binding to the Rho-binding domain of ROCK proteins, important effectors for RHOA signalling[215,245]. Glu40Lys seems to not impact the interaction between RHOA and ROCK1/2. In contrast Glu40Gln potentially enables these

interactions, and a consequence of enabled RHOA-ROCK binding should be enhanced RHOA downstream signalling. More fitting to the hypothesized LoF phenotype of RHOA variants is the putatively disabled interactions between RHOA Glu40Lys and several GEFs, for instance ARHGEF25 and ARHGEF3, could also be shown in Y2H assays, for example between both Glu40Lys and Glu40Gln with ARHGEF12 and ARHGEF25. Nonetheless, overexpression of human RHOA in yeast might not be the perfect environment to observe nuanced changes in protein function.

4.5.2 Gains and losses – consequences of RHOA variants

The Leu69Arg presumably lost interaction with both ARHGEF11 and ARHGEF12 in tandem affinity proteomics experiments, suggesting that this variant might be affected by impaired activation. Interactions with RTKN, a presumed oncogene (due to RTKN mediating apoptosis resistance[246,247]), as well as RIPOR1, were lost, too. Novel interactions gained by the variant Leu69Arg (ZNF207, ASH2L and GBAS) are difficult to interpret. These could be an indication of wrongly localized RHOA, perhaps as an artefact of significantly overexpressing Leu69Arg. However, this effect should have been seen in any one of the remaining cell lines. ZNF207 for example is known to have stabilizing effects on at least one other protein[248]. This could mimic the protective effects of GDI binding to small RHO GTPases, which protects the latter from degradation[249].

RHOA Tyr34Cys shows the highest number of lost interactions. The removal of the Tyr34 phosphosite did indeed abolish many interactions between RHOA and effector proteins (even ARHGEF11 and ARHGEF2), fitting the canonical function of Tyr34. Of interest is the enhanced interaction with DSG1, a protein involved in cell-cell adhesion[250], in turn requiring the cellular actin network to function, one of the main RHOA signalling targets.

The interaction between RHOA Glu40Lys and DIAPH1 appeared to be slightly strengthened, a result that was not previously observed in Y2H assays. Although research on DIAPH1 gained traction (at the time of writing DIAPH1 yields only ~ 249 hits on PubMed since 1993, with 158 hits since 2012) the protein has putative functions in brain development[251], ciliogenesis[252] and cytoskeletal organisation[253] – suggesting an important role for DIAPH1 in human health and disease, and underlines the heterogenous effects of RHOA variants, as only the Glu40Lys variant affects this interaction (but not

Tyr34Cys, Tyr42Cys or Leu69Arg). Noteworthy are also the presumed enhanced interactions with both RIPOR1 and RTKN. These interactions were mostly detectable in RHOA WT cells and were lost in other variants (RIPOR1: Tyr34Cys and Leu69Arg, RTKN: Tyr34Cys, Tyr42Cys, Leu69Arg) but strengthened in Glu40Lys, highlighting that RHOA variants seem to have nuanced effects on signalling networks.

While there is little evidence available to suggest a direct interaction between AKAP11 and RHOA it was observed that such an interaction might exist in RHOA WT cells, however it is lost in RHOA Tyr42Cys overexpressing cells. AKAP proteins belong to a family of anchor proteins with involvement in autism spectrum disorder, where it could be shown that AKAP13 regulates RHOA, which is beneficial for neurite outgrowth[254], linking RHOA to AKAP proteins. The gained interaction of RHOA Tyr42Cys to DUSP3 is noteworthy. DUSP3 is a phosphatase with a specificity for phosphotyrosines that has a large spectrum of different substrates, making DUSP3 an important player in human maladies[255]. If RHOA pTyr42 is a substrate of DUSP3, then the removal of pTyr42 might trap DUSP3. With DUSP3 being unable to dephosphorylate the RHOA Tyr42Cys variant DUSP3 might not be readily available to dephosphorylate other substrates.

Due to the nature of the experiments exploring PPIs (*in silico* analyses, Y2H and TAP) some discrepancies have to be expected. However, not only did TAP cover a much larger range of PPIs, an additional advantage is that human RHOA has been overexpressed in a human derived cell line, giving these results extra credibility.

4.5.3 The gap closing ability of U2OS cells is impaired when RHOA WT is overexpressed

I was not able to detect a significant proliferation difference of U2OS cells overexpressing RHOA WT or variants. The siRNA-mediated RHOA knockdown did not affect the short-term proliferation of U2OS cell, either. Due to the limited stability of the siRNA-mediated RHOA knockdown it was only possible to address cell proliferation over 48 h, however other conditions did also not produce significant results after 96 h (not shown).

Cell migration has important functions in human health and aberrations in cell migratory behaviour has previously been linked to diseases[256], with contributions of RHOA[257]. I saw a significant decrease in cell velocity (that equals gap-closing ability) for U2OS cells overexpressing RHOA WT. Cell locomotion depends on a dynamic interplay

between RAC1 at the leading edge and RHOA at the trailing edge[228]. One explanation could be that overexpression of RHOA WT inactivates RAC1 at the leading edge, therefore no movement would be observable. However, RAC1 protein levels in U2OS RHOA WTOE cells appeared unaffected (not shown). Another explanation could be that increased RHOA WT levels at the trailing edge compete with RAC1 activity at the leading edge to cancel each other out, yielding a quasi-zero net movement. RHOA dysregulation has recently been linked to reduced migration[258], however examples for increased migration and invasion[257] are plentiful, too. Excessive adhesion to the polymer of the cell culture dish through changes to focal adhesions[259] in RHOA WTOE cells could be yet another putative explanation for the observed phenotype.

Knockdown of RHOA did not affect the gap closing ability of U2OS cells. Overexpression of RHOA Tyr34Cys or Leu69Arg however rescued the RHOA WTOE phenotype. As knockdown of RHOA also showed U2OS WT behaviour it has to be assumed that both RHOA Tyr34Cys and Leu69Arg must be mainly inactivating RHOA function. Furthermore, it seems that absence of WT protein is not enough to stop U2OS cells from moving in my experimental setup, as shown by the WT phenotype of knockdown cells.

Overexpressing other variants (Glu40Lys, Glu40Gln, Tyr42Cys or Tyr42Ser) did not impair the gap closing ability of U2OS cells in a similar way as overexpressing the WT protein. However, since upon overexpression they also did not simply produce a phenotype mimicking that of WT U2OS cells but rather showed a slight but significantly increased cell motility these variants can also not be inactivating, as this would have produced a quasi-WT phenotype (as seen for Tyr34Cys and Leu69Arg). There must be a third option that is neither inactivation nor activation. RHOA is involved in a variety of different functions, many of them crucial to the survival of the cell, hinting at the idea of RHOA as a hub protein, where each variant affects a distinct set of PPIs that could result in a variety of observed phenotypes.

Tyrosine phosphorylation is a PTM with major significance for human health, with up to 30 % of oncogenes being tyrosine kinases. However, intracellular pTyrs are still comparatively rare[260,261]. Fortunately pTyrs can be detected due to their accumulation at focal adhesions, which is based on phosphorylation of focal adhesion kinase (FAK)[262] and paxillin[263].

Although the correct assembly of focal adhesions depends on RHOA[264] and pTyr accumulates at focal adhesions through FAK and paxillin phosphorylation, no change in phosphotyrosine brightness could be detected for U2OS cells overexpressing RHOA WT. This finding, together with the gap closing results for RHOA WTOE cells, might suggest that the antagonizing relationship between RHOA and RAC1 lead to the cells showing little net movement, and perhaps causes interference with focal adhesion recruitment of FAK and paxillin. A putative reason could be that RAC1 inactivates RHOA through p190 RhoGAP[265]. This would also explain why the number of presumed focal adhesions did not increase, despite more WT RHOA being present in RHOA WTOE cells. Besides, both phosphorylated FAK and phosphorylated paxillin are also able to suppress RHOA activity through p190 RhoGAP[266,267], most likely limiting the overall brightness and number of pTyr particles in the presence of excess RHOA.

Knockdown of RHOA did reduce pTyr brightness and also affected the number of pTyr particles, fitting to the known dependency of focal adhesion assembly and RHOA activity[264]. The same phenotype could be observed for RHOA Tyr34Cys (decreased pTyr number & brightness) and Tyr42Cys (decreased pTyr brightness), although the effect was less pronounced in the latter. The gap closing results had suggested an inactivating effect of RHOA Tyr34Cys and Leu69Arg, however the Leu69Arg variant did neither demonstrate a changed pTyr brightness nor number.

4.5.4 Knockdown of RHOA activates Cyclin D1 expression

Besides exploring PPIs it is possible to better understand RHOA signalling by studying gene expression. I selected a list of candidate genes based on literature and found that RHOA knockdown cells, as well as RHOA Leu69Arg and RHOA Tyr42Cys mutants displayed increased expression of ARHGDIA. Recent studies could show that increased RHOA activity is often accompanied with reduced ARHGDIA expression[268,269], it can be assumed that the reversed is true, too. It remains elusive as to why a change in ARHGDIA expression has not been observed for the RHOA Tyr34Cys variant, which showed putatively inactivating behaviour in other experiments. Perhaps the Leu69Arg and Tyr42Cys variants also, in addition to any edgetic effects, destabilize RHOA and, similar to a direct knockdown, lead to less available. ARHGDIA expression could be increased

as a means to protect RHOA from degradation, keeping RHOA more readily available, as has been shown for other GDIs[249].

The expression of Cyclin D1 (CCND1) demonstrates an increase in both RHOA knockdown and the Tyr42Cys variant. CCND1 is generally linked to cell proliferation and has been shown to be upregulated when RHOA activity is increased[270,271] and the activating RHOA G14V mutation promotes cell proliferation via CCND1 in epidermal stem cells[272]. There is clearly an established link between RHOA and CCND1, however, it remains elusive as to why a putatively inactivating variant (Tyr42Cys) would cause elevated CCND1 expression, especially since no proliferation effect had been observed in short-term cell culture experiments. Furthermore, CCND1 expression was also elevated in knockdown cells, highlighting once more the context and tissue dependent effects of RHOA.

4.5.5 Conclusion and outlook

The results presented in this chapter highlight the diversity of RHOA protein function in different cellular contexts. It can be deduced from literature that RHOA often has contradicting effects when comparing different tissues, for example RHOA activates proliferation in intestinal epithelial cells[273] but has an opposing effect in breast epithelial cells[274].

It could be shown that the investigated variants (Tyr34Cys, Glu40Lys/Gln, Tyr42Cys/Ser, Leu69Arg), despite all being considered to disrupt RHOA protein function, clearly do so in unique ways. This supports the idea of RHOA being a hub protein where each of the examined variants affects a unique set of edges, that is interactions between RHOA and interaction partners. Some of these edges might be directly affected (based on changes of the respective PPI strength), while others might only modulate gene expression. One example for this is RHOA Tyr34Cys and Leu69Arg, which both show a gap closing phenotype similar to RHOA knockdown cells when overexpressed. However, only Tyr34Cys also showed a reduced pTyr brightness – similar to RHOA knockdown, but not Leu69Arg. In contrast, the Leu69Arg variant demonstrated an upregulated ARHGDIA gene expression, similar to RHOA knockdown cells, but not Tyr34Cys.

Fully understanding the downstream network perturbations of RHOA WT and variants in U2OS cells would have exceeded the scope of this project. One idea to move

forward is using several different cell lines, as RHOA function depends on tissue and cellular context. Focal adhesions and actin filament imaging is a great opportunity to study RHOA, perhaps checking protein levels and phosphorylation status of paxillin and FAK via Western Blotting would be able to give additional answers. Generally observing cell proliferation for longer time periods or investigating gene expression patterns on a larger scale might improve the understanding of individual effects that each RHOA variant causes.

This sparked the idea to create a comparatively simple workflow to address both the cellular toxicity of overexpressed constructs, as well as judging on whether protein mutations affect protein activity. This approach is called Induced Cell Microarray Analysis (ICMA) and will be discussed in the next chapter.

Chapter V: Induced Cell Microarray Analysis (ICMA) – a simple workflow to detect loss/gain of function variants

5.1 Introduction

The other chapters in this thesis focus on computational and experimental methods to assess variant impact on protein function. This is motivated by the recent explosion in sequence and associated genetic variant data. When presented with a new variant (e.g. a variant of unknown significance), computational methods can offer some insights, but ideally findings should be experimentally confirmed. Unfortunately, traditional variant validation efforts have been extremely cumbersome, often involving years or even decades of research to establish (e.g.) whether a variant is gain- or loss- of function. For optimal clinical applicability, one needs rapid experimental tests that support or dismiss computational predictions. This chapter presents my attempts to design a gene-expression based workflow to identify gain-of-function variants in a matter of weeks.

The most well studied gain-of-function variants are found in cancers, where they most often affect key pathways activated by tyrosine kinases. This kinases transfer a phosphoryl group from ATP to tyrosines on a target protein[275]. This reversible process controls gene expression and protein-protein interactions[275] and dysregulation of this process has been shown to contribute to neurological disorders[275] and cancer[59,276]. The importance of both receptor tyrosine kinases (RTK) and non-receptor tyrosine kinases (nRTK) for human health and disease is highlighted by the fact that more than 25 % of worldwide drug discovery efforts are spent on this group of enzymes[277]. Examples are drugs that target either VEGF or its receptor, VEGFR, in renal cell carcinoma[278] or inhibitors of EGFR in lung cancer[279].

RTKs typically contain a N-terminal extracellular domain, a transmembrane domain as well as an intracellular kinase domain followed by a mostly unstructured C-terminus[280]. Activation of RTKs is facilitated upon ligand binding, which are often growth factors and each RTK can typically be activated by more than one ligand[280]. Upon ligand binding RTKs undergo conformational changes of the extracellular domain, making them prone to dimerization[281]. While several modes of dimerization are described[282] each mode ultimately activates the kinase domain. These domains are autoinhibited before receptor dimerization and trans-phosphorylation takes place upon dimerization, ultimately readying and activating the RTK[282] (**Fig. 5A**).

One possible outcome of RTK activation is the recruitment of PI3K to the membrane by binding to phosphotyrosine residues of RTKs[283]. The resulting PI3K/AKT/mTOR signalling influences cell cycle progression and cell growth, rendering it a relevant target for anti-cancer drugs[283,284] (**Fig. 5A**).

The RAS-RAF-MEK-ERK pathway is another well-studied cascade and is particularly important for several cancer hallmarks, including cell proliferation, differentiation and survival[285,286] (**Fig. 5A**). Perturbations in this evolutionary highly conserved[287] pathway are linked to a significant amount of human cancers and nearly all melanomas[285]. Targeted therapies with inhibitors against one of the proteins involved in RAS-RAF-MEK-ERK signalling, sometimes in combination with inhibitors against RTKs, have proven useful to increase progression-free survival of cancer patients[288,289].

Typical mutations in RAS affect the intrinsic GTPase activity. RAS proteins cycle between an inactive GDP-bound and an active GTP-bound state (similar to RHOA, see Chapter 4). Mutations of RAS proteins eliminate the GTPase activity, locking the protein in an GTP-bound active state, where RAS-GTP will bind to effector proteins[290]. MEK1, also called MAP2K1, functions as a mediator of signal transduction within the RAS-RAF-MEK-ERK cascade. Cancerogenic mutations often cluster around a N-terminal α-helix[291,292] that was shown to have an autoinhibitory function[293]. Disrupting the autoinhibition of MAP2K1 is very likely causing hyperactivation of the protein.

This project aimed to develop a straightforward method to assess hyperactivation of proteins involved in RTK signalling by means of tetracycline-induced cellular overexpression and microarray-based gene expression analysis. I tested the procedure on two well-known cancer variants (in HRAS and MAP2K1) and a third candidate recently identified in Lymphomas (Mozas, Lopez et al, submitted).

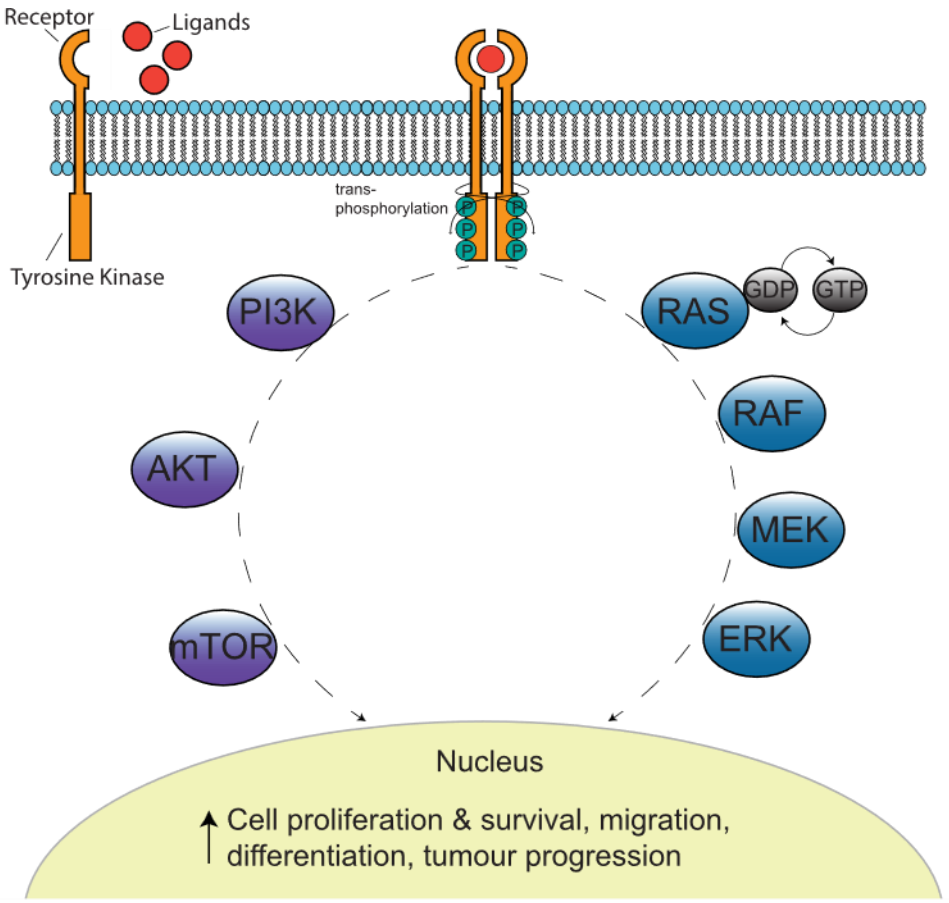


Figure 5A. Overview of receptor tyrosine kinase signalling. Receptor tyrosine kinases (RTK) dimerize upon ligand binding, causing trans-phosphorylation of the kinase domains from both receptor subunits. This in turn starts a signalling cascade, for example via PI3K-AKT-mTOR or RAS-RAF-MEK-ERK signalling, leading to increased cell proliferation, survival and enhanced tumour progression.

5.2 Material

Material and methods are as in Chapter 4.2 and 4.3, unless otherwise stated.

5.2.1 Antibiotics

Name	Company	Order No.
Blasticidin	Thermo Fisher Scientific	R21001
Geneticin	Thermo Fisher Scientific	10131035
Tetracycline	Thermo Fisher Scientific	A39246

5.2.2 Medias and supplement for mammalian cells

I. T-REx Standard Culture Medium

High glucose DMEM (GlutaMAX, Thermo Fisher Scientific) containing 10 % (v/v) FBS (Thermo Fisher Scientific) as well as 100 U/mL penicillin and 100 µg/mL streptomycin (see 2.1.1) was used. For successful culturing of T-Rex-293 cells 5 µg/mL blasticidin was added.

II. T-REx Tet- and PenStrep-free Culture Medium

Transfected T-REx-293 cells eventually receive 350 µg/mL geneticin. For this purpose high glucose DMEM (GlutaMAX, Thermo Fisher Scientific) containing 10 % (v/v) tetracycline-free FBS (Thermo Fisher Scientific) as well as 5 µg/mL blasticidin was used. Complete medium was filtered using 0.22 µm Stericups (Sigma Aldrich).

III. Freezing Medium

Cells were frozen in equal parts T-REx Standard Culture Medium (with blasticidin) and conditioned DMEM (medium taken from cultured cells, made cell-free by centrifugation) with addition of 10 % (v/v) dimethyl sulfoxide (DMSO, Sigma Aldrich).

5.2.3 Plasmids

Name	Supplier	Resistance
pT-REx-DEST30	Thermo Fisher Scientific	Ampicillin/Geneticin
pT-REx/GW-30 /LacZ	Thermo Fisher Scientific	Ampicillin/Geneticin

Plasmid cards can be found in Appendix 8.2.

5.2.4 Primers (qPCR)

Table 11: List of qPCR Primers

Target	Forward primer (5'-3')	Reverse primer (5'-3')
MAP2K1	GGTGTTCAAGGTCTCCCACAAG	CCACGATGTACGGAGAGTTGCA
HRAS	ACGCACTGTGGAATCTCGGCAG	TCACGCACCAACGTGTAGAAGG
PIM1	TCTACTCAGGCATCCGCGTCTC	CTTCAGCAGGACCACTTCCATG

5.2.5 DNA Sequences

Table 12: List of DNA inserts used to create DNA plasmids			
	Modification		
Insert Name	Protein Position	Amino Acid Change	Nucleotide Change
MAP2K1	Wildtype		
MAP2K1	56	Gln → Pro	cag → ccg
HRAS	Wildtype		
HRAS	61	Gln → Arg	cag → cgg
PIM1	Wildtype		
PIM1	23	Thr → Ile	acc → atc
PIM1	97	Ser → Asn	agc → aac
PIM1	127	Gln → Glu	caa → gaa

5.2.6 Ready-to-use premixes

Ready-to-use premix name	Company
Gateway™ BP Clonase™ II Enzyme mix	Thermo Fisher Scientific
Gateway™ LR Clonase™ II Enzyme mix	Thermo Fisher Scientific

5.2.7 Cell lines

I purchased Human T-REx™-293 cells from Thermo Fisher Scientific (#R71007).

5.3 Methods

5.3.1 Cell Culture

I. T-REx-293 cells

T- REx-293 cells were maintained in T-REx Standard Culture Medium in presence of 5 µg/mL blasticidin. Cells overexpressing genes after tetracycline induction were maintained in absence of penicillin and streptomycin, using qualified tetracycline-free FBS and under selection pressure by 350 µg/mL geneticin.

5.3.2 Modulation of gene expression

I. Tetracycline-induced upregulation

I used vectors with the pDest30 backbone to transiently overexpress wildtype or mutant proteins in T-REx-293 cells, stably expressing the Tet repressor (on pcDNA6/TR, Thermo Fisher) in the presence of 5 µg/mL blasticidin.

Transfection was performed in T-REx Standard Culture Medium as described in 4.3.2.II. Medium was then exchanged after 24 hours and 350 µg/mL Geneticin was added. Cells were split after 48 h – 72 h, depending on their confluence. Surviving cells typically reached 50 % confluency 144 h after transfection at which they would be induced with 1 µg/mL tetracycline. The Tet-on system[294,295] induces the production of transfected proteins in the presence of tetracycline. I harvested cells 24 h after tetracycline induction and target gene overexpression was confirmed via qPCR (also see 4.3.4).

5.3.3 Gateway Cloning

I. Sequence Design

I designed sequences using Benchling and the designed sequences were then ordered from Integrated DNA Technologies (IDT). attB1 and attB2 sites had to be added to the 5' and 3' end of the target gene exon sequence, as illustrated below:

ACAAGTTTGTACAAAAAAGCAGGCTTC + *5'-Sequence-'3* + 2x STOP + ACCCAGCTTTCTTGTACAAAGTGGT

with attB1 and attB2.

II. BP Reaction

I mixed 100 ng attB-DNA (purchased from IDT) with 100 ng pDONR/Zeo and filled up to a total volume of 4.5 µL with TE-buffer. Then, 0.5 µL BP Clonase II enzyme mix was added to the tube, mixed and incubated at room temperature overnight. 1 µL proteinase K was added to the mixture the next day. After incubation at 37 °C for 20 min the mixture was either stored at 4 °C or directly transformed into TOP10 *E. coli* on LB-Zeo as described in 4.3.9.I. DNA was isolated as in 4.3.9.IV and presence of the target gene was confirmed through sequencing.

III. LR reaction

The LR reaction was performed similar to the BP reaction. 100 ng pDONR/Zeo/Target were mixed with 100 ng pDEST30 and filled up to 4.5 µL total volume with TE-buffer. Then, 0.5 µL LR Clonase II enzyme mix was added to the tube, mixed and incubated at room temperature overnight. 1 µL proteinase K was added to the mixture the next day. After incubation at 37 °C for 20 min the mixture was either stored at 4 °C or directly transformed into TOP10 *E. coli* on LB-Amp as described in 4.3.9.I. DNA was isolated as in 4.2.9.IV and presence of the target gene was confirmed through sequencing.

5.3.4 Microarray Analysis

Tetracycline-induced cells were harvested and RNA was prepared using the RNeasy Mini Kit (Qiagen), including DNase (Qiagen) digestion, following the manufacturer's instructions. Overexpression of the target gene was confirmed by qPCR and 10 µL total RNA with a concentration of 50 ng/µL were prepared for submission. I then handed over the samples for microarray assays at and by the Genomics and Proteomics Core Facility (German Cancer Research Center [DKFZ], 69120 Heidelberg, Germany).

Upon receiving the raw data files an analysis pipeline was set up according to the "maEndtoEnd" R package[296]. To summarize, the raw data was loaded and data quality assessed through principal component analysis (PCA), using the R package "arrayQualityMetrics". After cleaning the data and removing any flagged chips the data was then background corrected and calibrated. To remove low-intensity signals the data was filtered by setting a threshold based on median intensities. The detected transcripts were annotated and translated into more human-readable gene names

before contrasts were defined, in order to group repetitions of a given sample condition together for comparison against wildtype or control conditions. These contrast groups were subjected to empirical bayes statistics for differential expression (eBayes) and results were extracted. Depending on the number of significantly upregulated genes (p-value ≤ 0.001) pathway and GO-term enrichment analysis was performed.

5.3.5 Mutational Footprint Analysis

I performed an analysis of signalling pathways based on the gene expression data collected from Microarray data by using a variety of different bioinformatics tools (doRothEA[297], decoupler (preprint), COSMOS[298], CARNIVAL[299]) and with the assistance of Prof. Dr. Julio Saez-Rodriguez and Dr. Aurélien Dugourd. This method aims to identify connections between genes in a gene expression dataset with help of a prior knowledge network[299]. The resulting sub-network is then improved by usage of an integer linear programming solver (IBM ILOG CPLEX[300]), uncovering pathway connections based on the gene expression data used as input.

5.3.6 Detection of Kinase Phosphorylation

For detection of human MAPK14/p38 Thr180 and Tyr182 phosphorylation an enzyme-linked immunosorbent assay (ELISA) was used (RayBiotech, #CBEL-P38-2). 40.000 T-REx HEK293 cells were grown, transfected, selected and induced (in triplicates) in wells of a 96-well plate (VRW, #734-0025), using a volume of 200 µL. The 96-well plate was coated with 20 µL of 0.1 mg/mL Poly-L-Lysine (Sigma Aldrich, #P9155-5MG) for 2 h at RT prior to seeding. 24 h after tetracycline induction I gently washed cells 3 times with 200 µL wash buffer A before fixation for 20 min with 100 µL fixing solution. The wash step was repeated and remaining fixing solution quenched with 200 µL quenching buffer. After an additional washing step, I blocked wells with 200 µL blocking buffer for 1 h at 37 °C. Another washing step was carried out, using washing buffer B before incubating the cells with 50 µL primary antibody for 2 h at RT. I incubated half of the wells with either α-p38 or α-phospho-p38 primary antibody. After an additional washing step, using washing buffer B, each well was incubated with 50 µL of an HRP-conjugated secondary antibody (mouse) for 1 h at RT. After washing with washing buffer B, each well was incubated with 100 µL TMB substrate for 30 min at RT in the dark. Afterwards 50 µL stop solution was added and absorbance at 450 nm was measured using a ‘Tecan Spark’ plate reader.

Then, all remaining solvent was removed from the wells and cells were washed with deionized water before being stained with 50 µL 0.1 % (w/v) crystal violet (Sigma Aldrich, #G2039-100G) for 20 min at RT. Cells were then again washed with deionized water before destaining with 100 µL 0.3 % (w/v) SDS for 15 min at RT. Absorbance was measured at 590 nm to address cell density.

I normalized resulting α-phospho-p38 values based on the average α-p38 signal and the relative cell density of each individual well determined by crystal violet staining. Outliers were determined using a Z-score transformation, following formula 3:

$$Z = \frac{x - \mu}{\sqrt{\frac{\sum (x_i - \mu)^2}{n-1}}} \qquad (3)$$

Where x is the value of a single measurement and µ corresponds to the mean of an experimental group. A z-score ≥ |3| was considered to be an outlier and the respective value was removed. The procedure was nearly identical for the detection of phosphorylated ASK1 (BioAssays, A102753), following the manufacturers protocol.

5.4 Results

5.4.1 A method to rapidly assess gene expression using mammalian cells as a tool

The method presented here aims to perform Gateway cloning of a gene of interest into a suitable vector with successive transfection into T-REx HEK293 cells, that stably express the Tet repressor in the presence of blasticidin. Cells are then harvested and total cell RNA is ultimately sent to gene expression profiling using microarray technology (**Fig. 5B**, **A**). The acquisition of biological triplicates in the cell culture can be rapidly achieved by performing transfection and selection of T-REx HEK293 cells in a small volume, i.e. 6-well culture dishes (or smaller). A small overall culture volume ensures that enough wells can be seeded with only a minimum of initially available cells, for example cultured in a typical T75 flask. Furthermore, a 6-well with a final confluence of only 65-70 % still yields enough total cell RNA for post-processing (alternatively: a full 12-well). A single cell culture experiment typically takes ~ 7 days from seeding to harvesting, which includes transfection, antibiotic selection and tetracycline-dependent induction of target gene expression (**Fig. 5B**, **B**). Furthermore, smaller culture volumes and the total requirement of fewer cells enable the performance of three experiments, using cells from three individual flasks for seeding, in parallel (**Fig. 5B**, **C**).

It is thus possible to produce biological triplicates in a matter of 4 to 5 weeks, starting with target gene sequence design. Additional time is required for the external gene expression screening, though the processing of the resulting data is almost instantaneous through R scripting.

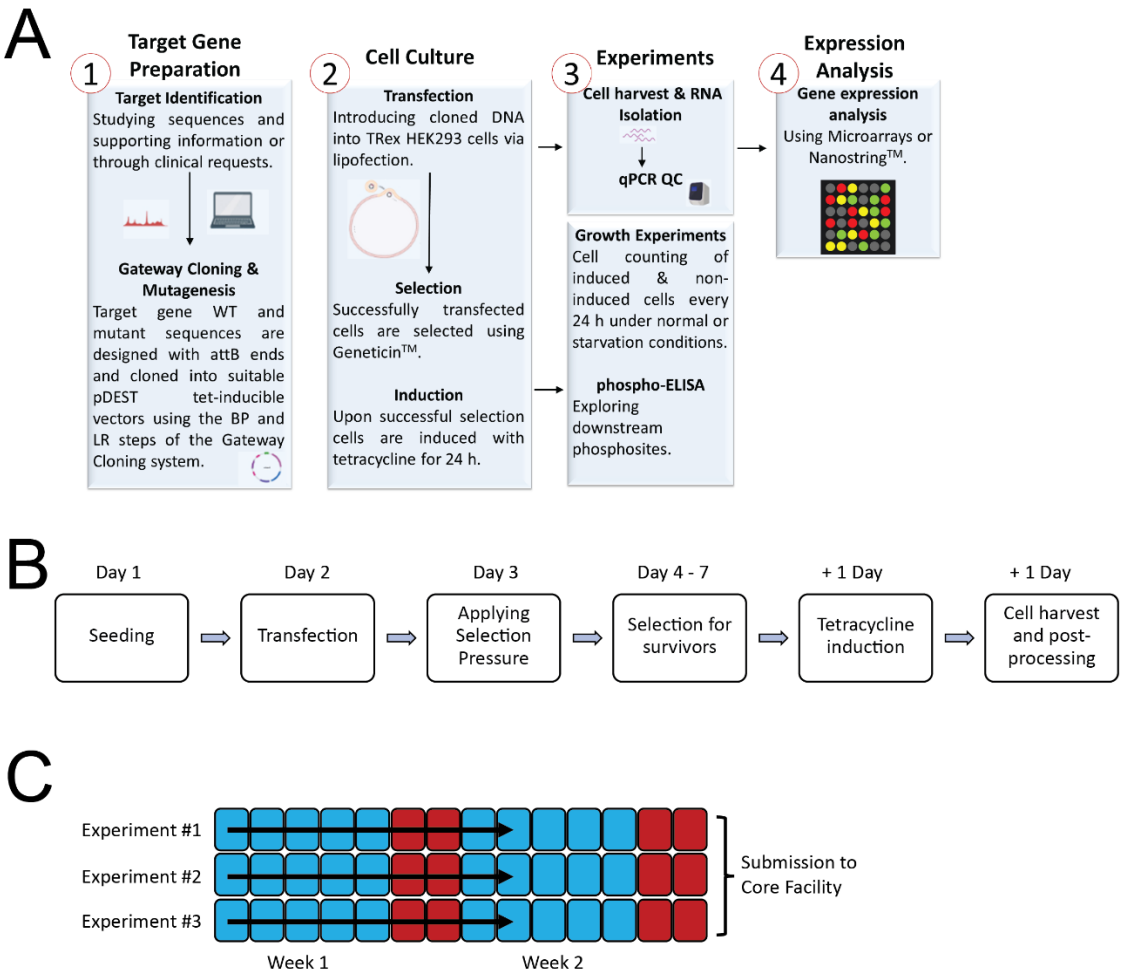


Figure 5B. Methodological overview. A. General method overview. Target gene preparation via cloning after a candidate has been selected (1). After transfection of DNA into T-REx HEK293 cells and selection, target gene expression is induced with tetracycline (2). Several experiments can be performed using induced cells, including growth assays or phospho-ELISAs. RNA isolation is used to conform target gene overexpression (3). Samples overexpressing the target gene are sent to gene expression analysis (4). **B. Outline of a typical cell culture experiment.** A typical cell culture experiment contains the initial seeding, transfection, selection, induction of target gene expression and cell harvest followed by total cell RNA extraction. **C. Scheduling for 3 cell culture experiments that run in parallel.**

5.4.2 Proof of principle 1: MAP2K1 Gln56Pro

Tetracycline-dependent overexpression of MAP2K1 WT in T-REx HEK293 cells (compared to the non-induced but transfected control) lead to an increase of both MAP2K1 (logFC = 2.17, p-value = 5.37×10^{-7}) and EPPIN-WFDC6 (logFC = 0.55, p-value

= 0.0009), with no additional genes observed to be differentially regulated (**Fig. 5C**, **A**). Cells overexpressing MAP2K1 Gln56Pro showed 19 differentially regulated genes (logFC ≥ |0.5|, p-value ≤ 0.001), including EGR1 (logFC = 1.49, p-value = 1.17×10^{-7}), EGR3 (logFC = 1.19, p-value = 3.67×10^{-5}) and ANXA1 (logFC = 0.72, p-value = 0.0001) (**Fig. 5C**, **B**). The final comparison between cells overexpressing MAP2K1 Gln56Pro against cells overexpressing MAP2K1 WT yielded 7 differentially expressed genes, including EGR1 (logFC = 1.56, p-value = 9.11×10^{-8}), RSPH4A (logFC = -0.67, p-value = 0.0003) and TFPI2 (logFC = 1.51, p-value = 5.83×10^{-6}) (**Fig. 5C**, **C**).

Using several computational tools and the assistance of the Saez-Rodriguez lab I was able to deduce a putative signalling network that provided several routes of EGR1 overexpression, namely via MAP2K1→MAPK3→EGR1 as well as MAP2K1→MAPK1→ELK1→ EGR1 and via MAPK3→STAT3→EGR1 (**Fig. 5C**, **D**).

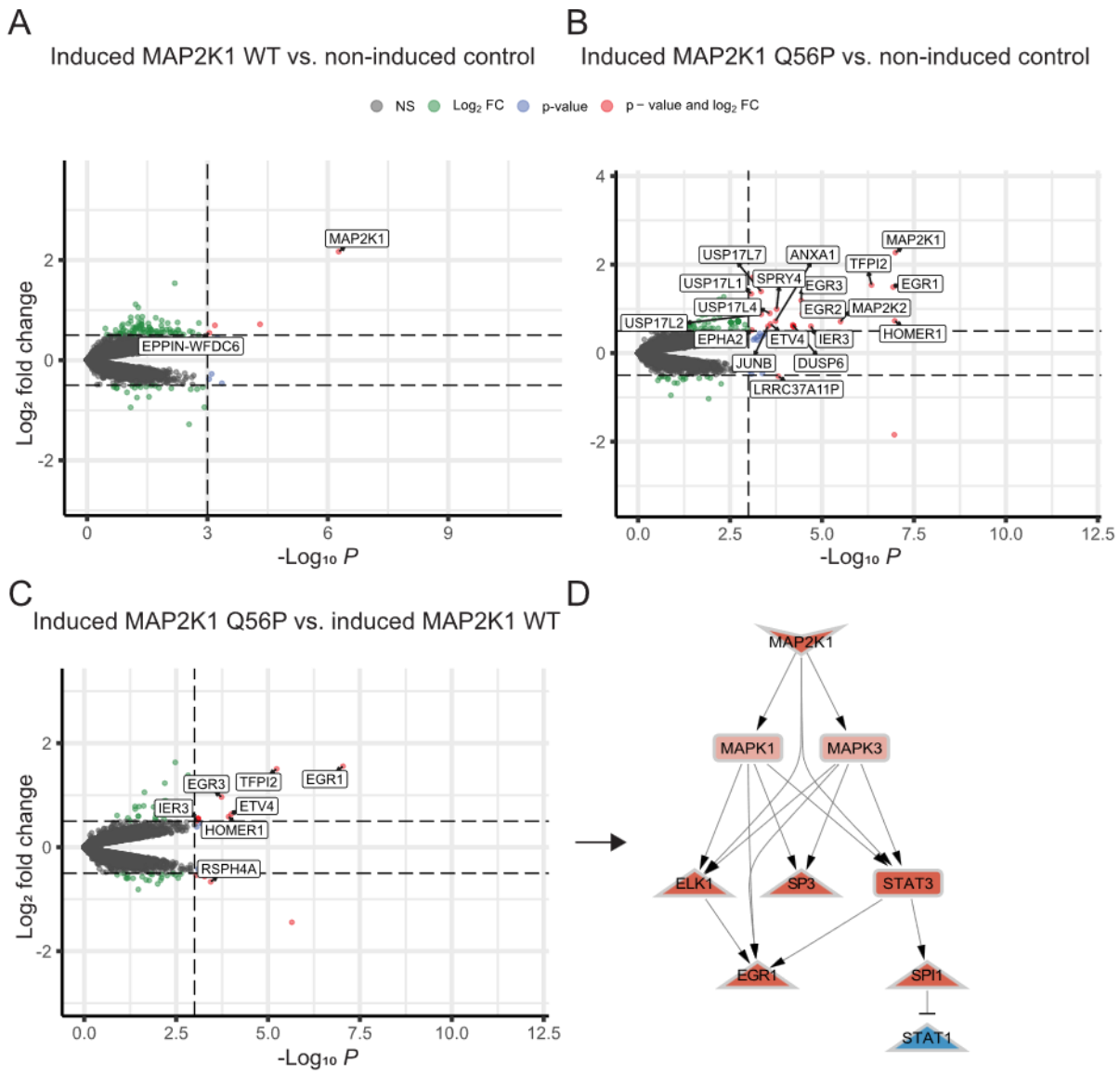


Figure 5C. MAP2K1 Gln56Pro overexpression yields differentially regulated genes that differ from MAP2K1 WT overexpression. A. MAP2K1 WT overexpression does not differentially regulate downstream genes. The top 20 genes with logFC ≥ |0.5| & p-value ≤ 0.001 (red dots) are labelled. **B. Overexpression of MAP2K1 Gln56Pro leads to overexpression of downstream genes.** As in A. **C. MAP2K1 Gln56Pro differs from MAP2K1 WT in its gene expression profile.** As in A and B. **D. Putative causal network of EGR1 overexpression.** Formatting according to E. Gjerga[301], where red (and salmon, though to a lesser degree) coloured shapes indicate upregulation (i.e. EGR1), blue coloured shapes indicate downregulation (i.e. STAT1). A downward pointing triangle indicates a measured node and an upward pointing triangle indicates a perturbation target.

5.4.3 Proof of principle 2: HRAS Gln61Arg

WT HRAS overexpression in T-REx HEK293 cells lead to a number of differentially regulated genes (n = 16) including EGR1 (logFC = 0.86, p-value = $1.5x10^{-5}$), MMP10 (logFC = 1.1, p-value = $4.66x10^{-5}$) and SPRY4 (logFC = 0.84, p-value = 0.0005) (**Fig. 5D**, **A**). The number of differentially regulated genes in cells overexpressing HRAS Gln61Arg was much higher (n = 150) and include MMP1 (logFC = 1.63, p-value = $5.22x10^{-8}$), MMP13 (logFC = 1.74, p-value = $1.19x10^{-7}$) and JUNB (logFC = 1.06, p-value = 0.0002) (**Fig. 5D**, **B**). When comparing cells overexpressing HRAS Gln61Arg with cells overexpressing HRAS WT the number of differentially expressed genes increases noticeably (n = 433). These include a number of matrix metalloproteinases (MMPs), e.g. MMP1 (logFC = 1.45, p-value = $1.87x10^{-7}$) and MMP10 (logFC = 1.34, p-value = $6.79x10^{-6}$). Other differentially expressed cancer related genes included CDKN1A (logFC = 1.12, p-value = $6.12x10^{-7}$) and CDHR1 (logFC = -1.01, p-value = 0.0001) (**Fig. 5D**, **C**). The putative causal network derived by CARNIVAL reveals a HRAS & MAPK1 based upregulation of JUN and STAT1 s well as transcription factor SP3 (**Fig. 5D**, **D**).

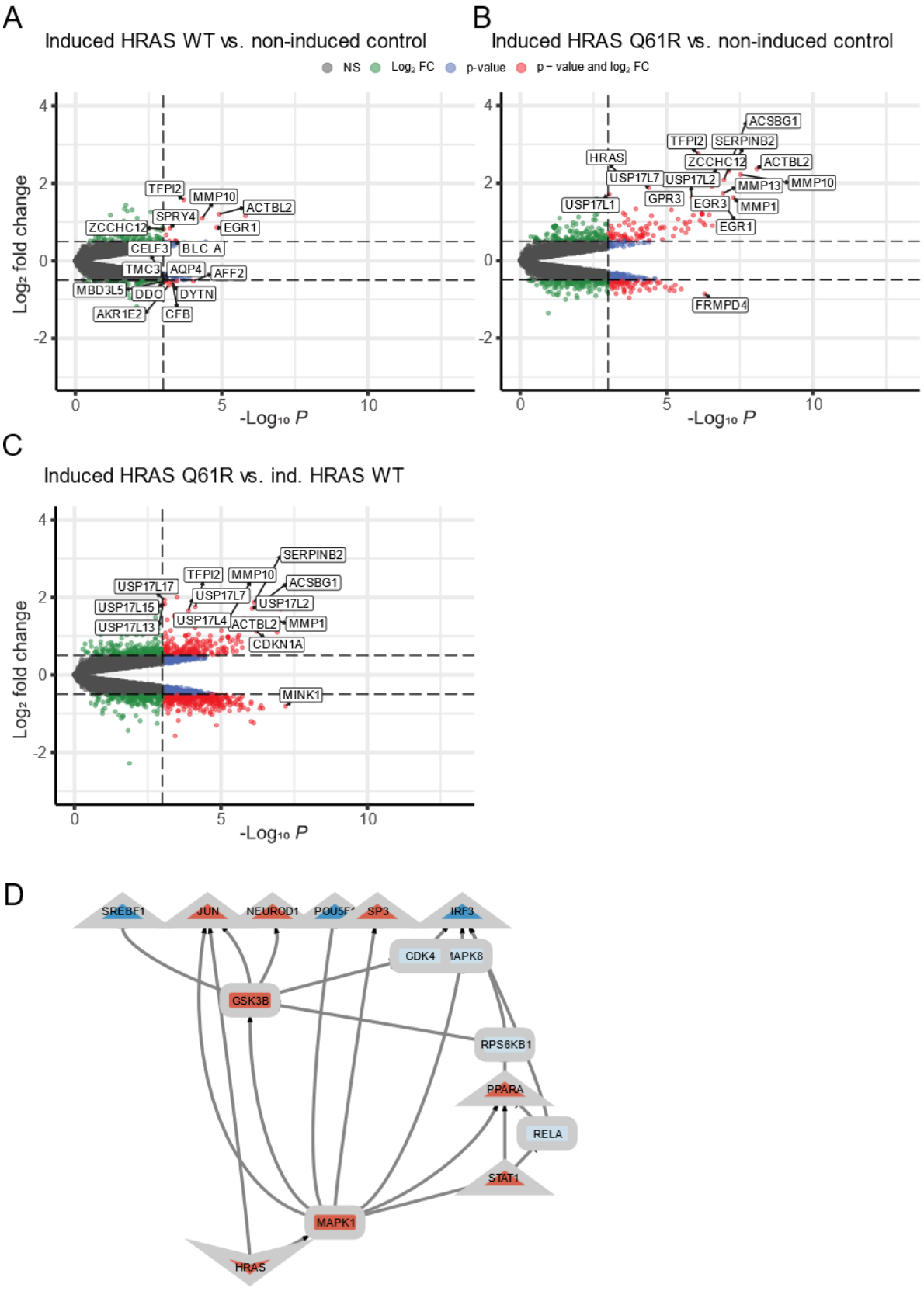


Figure 5D. HRAS Gln61Arg overexpression yields a significant number of differentially expressed genes that differ from HRAS WT overexpression. A. HRAS WT overexpression does show some differentially expressed downstream genes. The top 20 genes with logFC

≥ |0.5| & p-value ≤ 0.001 (red dots) are labelled. **B. Overexpression of HRAS Gln61Arg leads to overexpression of downstream genes.** As in A. **C. HRAS Gln61Arg differs from HRAS WT in its gene expression profile.** As in A and B. **D. Putative causal network of HRAS Gln61Arg downstream signalling.** Formatting as described in Fig. 5B, D.

5.4.4 PIM1 variants and a gain/loss of function

In contrast to HRAS or MAP2K1, the effects of perturbing PIM1 are not well understood. I conducted additional experiments to complement the gene expression profiling. First, cell proliferation in cells overexpressing PIM1 WT or PIM1 variants (Thr23Ile, Ser97Asn, Gln127Glu) was observed for 96 h, yielding no significant results. I performed ELISA assays to determine MAPK14/p38 and MAP3K5/ASK1 phosphorylation, as PIM1 can affect MAPK14/p38 phosphorylation through MAP3K5 signalling (**Fig. 5E**, **B**). While overexpression of PIM1 WT or the Thr23Ile variant had no effect on phospho-MAPK14 levels it was possible to observe a decrease of MAPK14 phosphorylation to 46 % in PIM1 Ser97Asn overexpressing cells, as well as an increase to 127 % in cells overexpressing PIM1 Gln127Glu (**Fig. 5E**, **A**).

PIM1 WT overexpression did not yield any noteworthy results. However, when PIM1 Thr23Ile was overexpressed a small number of genes were found to be differentially regulated (n = 6) when compared to PIM1 WT, including CCNA1 (logFC = 0.55, p-value = 0.0001) and IL13 (logFC = -0.55, p-value = 0.0006). Cells overexpressing Gln127Glu showed a greater number of differentially regulated genes (n = 18), including GH1 (loFC = 0.68, p-value = 0.0001) and ANKRD53 (logFC = 0.712, p-value = 5.78×10^{-5}). Overexpression of PIM1 Ser97Asn yielded the highest number of differentially expressed genes (n = 34), which include LMO2 (logFC = -0.8, p-value = 0.0006), ALX1 (logFC = 0.77, p-value = 0.0003) and KDM5D (logFC = -0.54, p-value = 0.0007) (**Fig. 5E, C** and **D**). The causal network putatively derived by CARNIVAL shows that PIM1 Ser97Asn negatively affects MAPK14 signalling, which in turn increases MAPK3 downstream signals (**Fig. 5E**, **E**).

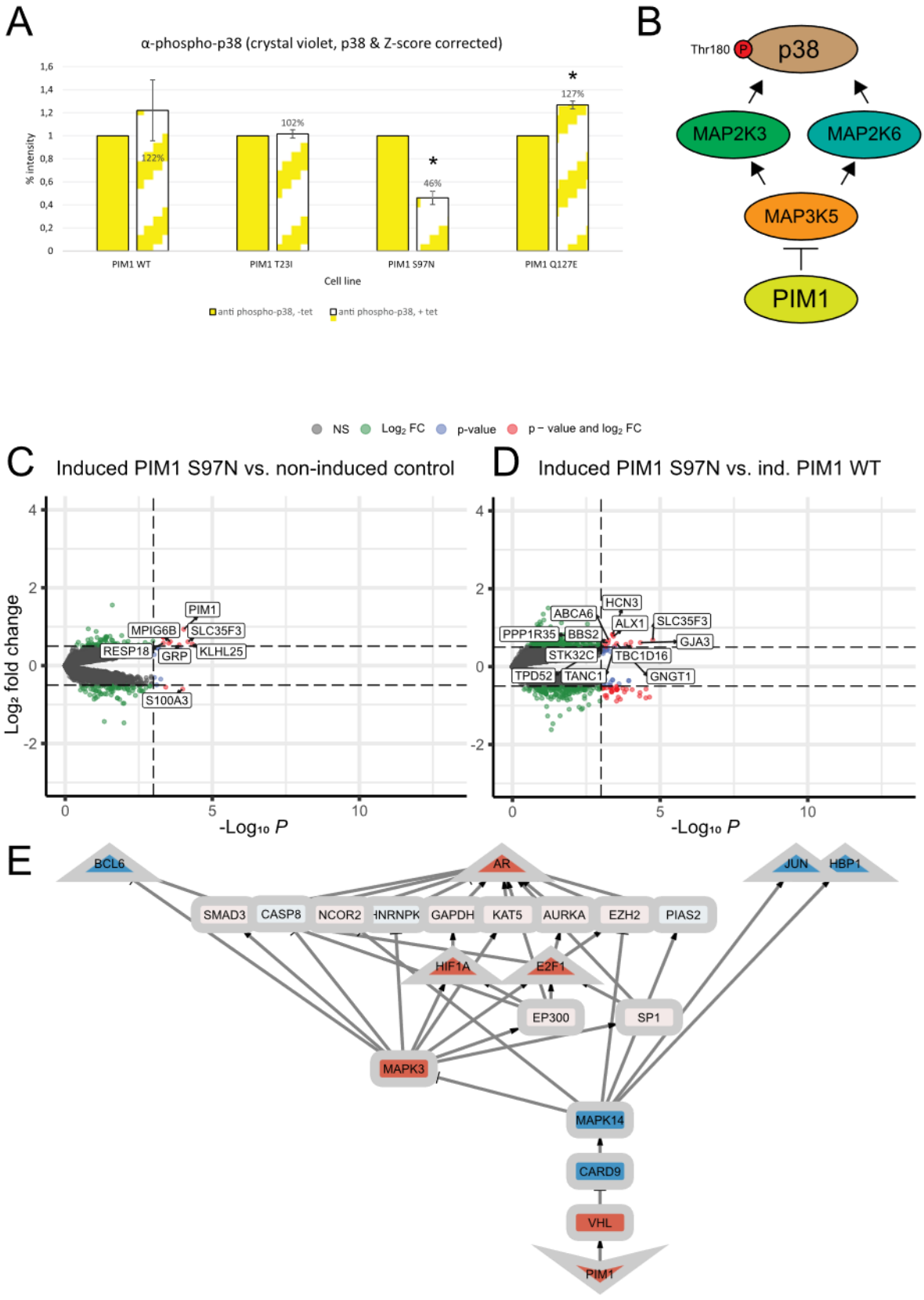


Figure 5E. PIM1 Ser97Asn overexpression shows signs of hyperactivity when compared to PIM1 WT overexpression. A. PIM1 Ser97Asn reduces MAPK14 phosphorylation. ELISA results of PIM1 WT and the variants Thr23Ile, Ser97Asn and Gln127Glu. **B. PIM1 negatively**

affects MAPK14/p38 signalling via MAP3K5 inhibition. PIM1 canonically inhibits MAP3K5 signalling. One consequence is a decrease in MAPK14/p38 phosphorylation through missing MAP3K5→MAP2K3/6 downstream signalling. **C and D. PIM1 Ser97Asn differs from control or WT in its gene expression profile.** The top 20 genes with logFC ≥ |0.5| & p-value ≤ 0.001 (red dots) are labelled. **E. Putative causal network of PIM1 Ser97Asn downstream signalling.** Formatting as described in Fig. 5B, D.

Additionally, I assessed the phosphorylation status of Ser83 ASK1 (also called MAP3K5, see **Fig. 5E**, **B**). This MAPK is located upstream p38 and a direct target of PIM1. PIM1 WT overexpression yielded slightly lower levels of ASK1 Ser83 phosphorylation (86 %, p = 0.07) the overexpression of PIM1 Ser97Asn, but not Thr23Ile or Gln127Glu, lead to significantly (p = 0.003) decreased levels of phosphorylated ASK1 (67 %) compared to the non-induced controls (**Fig. 5F**).

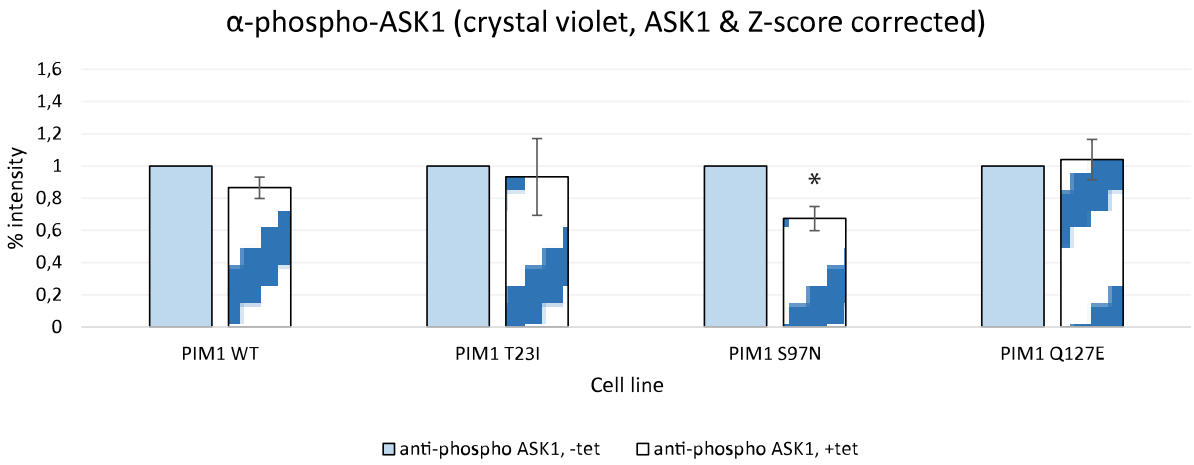


Figure 5F. PIM1 Ser97Asn overexpression displays lower ASK1 phosphorylation at Ser83. ELISA results of PIM1 WT and the variants Thr23Ile, Ser97Asn and Gln127Glu.

5.5 Discussion

5.5.1 Time consumption & cost can be minimized through usage of HEK293 cells and gene expression profiling, with implications for clinical treatment assessment

The time between cancer diagnosis and treatment is an important factor for cancer patient survival[302] and many weeks may pass between primary care and the start of an anti-cancer treatment regimen[303,304].

The here presented workflow is using T-REx HEK293 cells as a tool, rather than a tissue model. HEK293 cells are widely used for cell studies today. However, large scale efforts often reduce HEK293 cells to mere producers of recombinant proteins[305–307]. This approach uses modified HEK293 cells as a tool to generate gene expression data. These kind of data have been previously shown to be helpful for predicting prognosis or treatment response[308], and analysis of mRNA has recently been shown to outperform diagnosis based on genomics alone[309].

This workflow is reasonably fast (4-5 weeks for a triplicate), a time-frame that could be realistically helpful in treatment decisions. Moreover, my approach is moderately priced (~ 4000 € in material to compare a WT protein to a single mutant) while avoiding cytotoxicity effects caused by prolonged overexpression of target genes[310] or prolonged exposure to tetracycline[311]. Hence, this method could play a valuable role determining and improving anti-cancer treatment in the age of personalised medicine[312].

5.5.2 MAP2K1 Gln56Pro gene expression profile confirms loss of auto-inhibition

MAP2K1 is an important player in RAS/RAF/MEK/ERK signalling and thus of major significance in human cancer[285,313]. It is an established essential gene: for instance, deletion of MAP2K1 causes embryonic death in mice[314]. MAP2K1 mutations implicated in cancer are often found to be at or close to an N-terminal α-helix, in particular Gln56Pro, Lys57Glu and Lys57Asn[292,315]. This α-helix negatively regulates MAP2K1 activity, explaining why the protein has a smaller baseline activity compared to other protein kinases[316]. While MAP2K1 is normally activated by phosphorylation of Ser218 and Ser222[317] – similar to ERK[316] – disruption of the negative regulatory α-helix was shown to hyperactivate MAP2K1[318]. Serine mutations in the activation loop are not typically seen

in cancer[16] and disruption of MAP2K1 auto-inhibition have significant implications for cancer treatment, due to resistance to upstream targeting inhibitors (i.e. targeting RAS or RAF proteins[319,320]). Knowledge of which mutations disrupt MAP2K1 autoinhibition can give important adjustments to anti-cancer treatment, as effective inhibitors to target MAP2K1 directly are available[321].

As expected, given the autoinhibitory nature of MAP2K1, overexpression of MAP2K1 WT does not lead to any significant differential gene expression with the exception pf the read-through transcription product EPPIN-WFDC6, about which very little is known. However, overexpression of MAP2K1 Gln56Pro revealed a number of differentially expressed genes (both compared to control cells or to MAP2K1 WT). The Gln56Pro variant however is likely disrupting the proteins autoinhibition and thus increasing its activity. For instance, transcription factors EGR1 and EGR3 were shown to be regulated by MAPK signalling and have implications for human health[322–327]. While current knowledge about RSPH4A is sparse, it could be shown that defects in this ciliary protein can be causative of primary ciliary dyskinesia[328], and a similar protein, RSP3, is a known target of ERK signalling[329]. While the differential regulation of RSPH4A does not fit into the context of cancer there are at least some hints at the importance of RSPH4A for human health and its possible regulation downstream of MAP2K1.

Another up-regulated gene is ANXA1, a modulator of ERK signalling and often found to be downregulated in metastatic cancer[330,331]. While T-REx HEK293 cells are immortalized, they are not typical cancer cells and therefore overexpression of ANXA1 could be a natural response to overexpression of the more active MAP2K1 Gln56Pro protein. Likewise upregulation of the anti-angiogenic TFPI2[332] could be a natural response of MAP2K1 hyperactivation, and studies did show that expression of TFPI2 is directly linked to ERK activation[333,334].

Taken together these results suggest that MAP2K1 Gln56Pro is likely more active than MAP2K1 WT, and that it is possible to link some of the differentially expressed genes to MAP2K1 downstream signalling. These results are thus a first proof-of-principle of this workflow to yield reasonable results.

5.5.3 Hyperactivation of HRAS Gln61Arg enhances MAPK signalling

HRAS is one of three similar small GTPases (with KRAS and NRAS). While the subcellular location of RAS proteins affect their function[335] the majority of cancer mutations occur at protein positions (the same for all three homologs) 12, 13 or 61, with KRAS most frequently and HRAS least frequently mutated[335]. These mutations are located either in the switch II or the P-loop region[336] and typically affect Ras GTPase activity[335,337], locking the proteins in their active state. HRAS mutations are often observed in melanomas[338], however the HRAS Gln61Arg variant was also seen in epithelial-myoepithelial carcinoma[339] and a rare oestrogen receptor negative breast cancer[340]. Knowledge about HRAS mutations is important, as drugs are available to target MAPK and mTOR signalling downstream of HRAS activation[341].

One gene that I found to be upregulated upon overexpression of HRAS Gln61Arg was CDKN1A (also called p21). This gene is often silenced in cancer and thought of as a tumour suppressor by inhibiting cyclin-dependent kinase[342]. However, CDKN1A overexpression was also previously linked to oncogenic features. For instance, upregulated CDKN1A has been observed in non-small cell lung cancer[343] and in a variety of additional tissues. Here, it has been discovered that an increase in cell motility is based on a reduction of actin stress fibres, which is achieved by repressing RHOA activity[344].

Overexpression of HRAS WT lead to a slight increase in EGR1 and SPRY4 expression. The latter is a known inhibitor of MAPK signalling through interference with GTP-Ras formation[345]. The inhibitory effects of SPRY4 on MAPK signalling are further reinforced by the finding that a SPRY4 mutation in familial non-medullary thyroid cancer lead to increased proliferative effects[346], and that SPRY4 mediates tumour suppressive effects when both BRAF Val600Glu and NRAS Gln61Arg are overexpressed, mutations that normally are mutually exclusive[347]. The former EGR1 is linked to Src-RAS-RAF-MEK-ERK signalling[348] and expression of oncogenic HRAS was previously linked to a loss of EGR1 expression[349].

The gene CDHR1 was downregulated in cells overexpressing HRAS Gn61Arg. It is interesting as CDHR1 is a photoreceptor-specific cadherin, and while its involvement in cancer is not yet well understood a low expression of CDHR1 was previously found to be indicative of worse survival in glioma patients[350]. Amongst the upregulated genes after HRAS Gln61Arg overexpression is JUNB, a leucine zipper transcription factor found to

be overexpressed in lymphomas[351] and known to be affected by oncogenic HRAS signalling[352,353]. Matrix metalloproteases (MMPs) are important for softening the extracellular matrix, which favours metastases, one hallmark of cancer[226]. I found MMP1 and MMP10 to be upregulated upon HRAS Gln61Arg overexpression, and while links between upregulation of MMPs and KRAS activity or MAPK signalling are most well understood in animal studies[354–357], the finding of dysregulated MMPs through HRAS Gln61Arg harbours important prospects for improved understanding of HRAS related cancer syndromes.

In summary, the results presented here suggest that the HRAS Gln61Arg variant leads to increased HRAS activity, which agrees with literature, hereby acting as a second proof of principle.

5.5.4 PIM1 Ser97Asn increases protein activity

The Ser/Thr kinase PIM1 is highly expressed in hematopoietic, gastric, head and neck cells[358], as well as in several tumours including prostate cancer and lymphomas[358–360]. PIM1-3 are conserved proteins and demonstrate partial functional overlap[358]. PIM1 has a shorter and a longer isoform, and in addition to differences at the C-terminus it has also been shown that the shorter isoform localizes mainly in the nucleus or the cytosol, whereas the longer isoform accumulates at the plasma membrane[361]. PIM1 lacks a regulatory domain and is considered to be constitutively active[358], separating it from most of kinases. Despite not possessing typical regulatory features common to other kinases, it has been observed that PIM1 phosphorylation indeed affects protein activity and half-life, with the latter being < 5 min[358], presenting a challenge for PIM1 detection in healthy tissues[360]. PIM1 is connected to AKT signalling[358] and inhibits apoptosis, promotes proliferation and increases genomic instability[360]. Reports also suggest that PIM1 is a positive regulator of G2/M transition of the cell cycle[362] as well as enhancing cell motility and invasion[363], and overexpression of PIM1 is linked to poor prognosis of cancer patients[364,365].

Mutations in PIM1 are reported for a large fraction of human myeloid and lymphoid leukaemia and lymphomas[359,366,367], and I subjected the three somatic PIM1 variants (Thr23Ile, Ser97Asn and Gln127Glu) to further investigation are exclusively found in haematopoietic and lymphoid tissue[16]. Thr23 is a known phosphorylation site[368] and

Ser97 lies in a structural position where activating mutations are found in other kinases. Nothing is known for Gln127, though it lies on the surface of the kinase and thus could play roles in protein interactions[369].

MAP3K5 (also called ASK1) downstream signalling is linked to pro-apoptotic effects[370] and PIM1 is known to negatively regulate ASK1 and other pro-apoptotic proteins, hence promoting cancer cell survival[371]. While cell proliferation experiments did not yield any significant differences over the course of 96 h it was possible to detect changes in p38 phosphorylation for PIM1 Ser97Asn and PIM1 Gln127Glu variants, but not for PIM1 WT overexpressing cells. p38 phosphorylation is mediated via ASK1 downstream signalling[372]. Cells overexpressing PIM1 WT (and the Thr23Ile variant, lying at a known phosphosite[368]) however did not display significantly changed phospho-p38 levels despite showing the highest signal variation. If any of the investigated variants would show a higher activity than WT, then a reduction of phospho-p38 would be expected. Indeed, PIM1 Ser97Asn overexpressing cells showed a 54 % lower phospho-p38 signal, hinting at a stronger inhibition of ASK1 and hence lower phosphorylation of p38. The PIM1 Gln127Glu variant displayed an increase in p38 phosphorylation of 27 %. This result would indicate a less active PIM1 enzyme and any activating effects of this variant must be doubted.

Ser83 phosphorylation in ASK1 raises some questions. It has been shown that Ser83 is a direct target of PIM1, and that ASK1 phosphorylation at Ser83 inhibits ASK1 activity, therefore weakening outgoing pro-apoptotic signalling[371]. This observation would explain a reduced phosphorylation of p38, but a direct analysis of ASK1 Ser83 in cells overexpressing PIM1 Ser97Asn showed a reduction, not an increase, of ASK1 phosphorylation.

ASK1 is known to respond to stress. One explanation could be that PIM1 phosphorylates and stabilizes HIF1α[373], which in turn activates TNFα[374,375]. TNFα on the other hand has been shown to lead to decreased pASK1 levels via an unknown phosphatase[376,377]. Perhaps this non-canonical signalling via PIM1-HIF1α-TNFα has a stronger effect on ASK1 than PIM1 activity alone. In the absence of true hypoxia conditions, and accumulation of HIF1α might therefore lead to aberrant ASK1 signalling, which is linked to increased angiogenesis via VEGF[378]. However, accumulating HIF1α might also act as a perceived stress signal (again in absence of true hypoxia/stress conditions) which has been shown to activate PRL-3, which in turn dephosphorylates

p38[379]. This non-canonical signalling might also lead to cancer progression[380], rather than cell death (**Fig. 5G**). This complex example illustrates that signalling events are often more complicated than anticipated, and that non-canonical signalling is often able to explain disease-associated phenotypes.

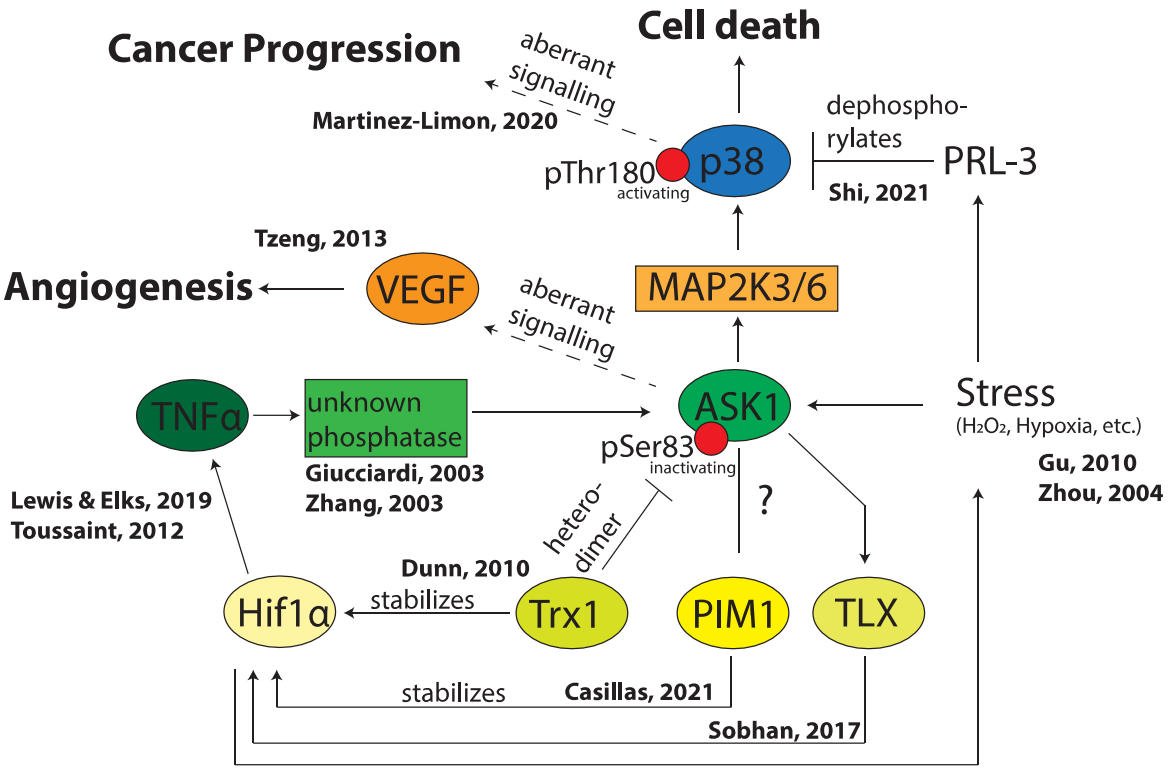


Figure 5G. Non-canonical PIM1 signalling, leading to dephosphorylated ASK1 and dephosphorylated p38. PIM1 signalling could cause a dephosphorylation of ASK1 via HIF1α while the latter also acts as a stress signal leading to simultaneous dephosphorylation of p38.

An additional finding supporting the notion that PIM1 Ser97Asn is more active than the WT however is the number of genes that are differentially expressed in the direct comparison between PIM1 Ser97Asn and PIM1 WT overexpressing cells. The methylase KDM5D is amongst the downregulated genes and is considered a tumour suppressor[381] by transcriptional repression of target genes[382]. Not only are histone modifying genes often perturbed in lymphoma[383] – where PIM1 is also frequently mutated[367] - low expression of KDM5D is also linked to poor prognosis in prostate cancer patients[381] and can lead to increased cell migration in gastric cancer[384].

The cysteine-rich gene LMO2 is also downregulated in PIM1 Ser97Asn overexpressing cells. LMO2 regulates hematopoietic stem cell development[385] and perturbed LMO2 gene expression is associated with formation of leukemia[386,387].

ALX1, a transcription factor, was found to be upregulated in cells overexpressing PIM1 Ser97Asn. This gene is suggested to play an important role in early development and as a regulator of epithelial-to-mesenchymal transition[388,389], with the latter being important for tumour formation and cell invasion. Notably, ALX1 upregulation has been associated with poor prognosis in lung cancer patients[390].

The CARNIVAL derived causal network links PIM1 Ser97Asn overexpression with VHL and HIF1α signalling, presenting a non-canonical means of p38 inhibition. While ELISA results suggest that PIM1 Ser97Asn expressing cells affect p38 through the canonical route via ASK1 downstream signalling, numerous reports do indeed associate PIM1 with HIF1α signalling. For example, it has been reported that there is a feedback loop between HIF1α and PIM2[358,391], and the functional overlap of PIM kinases suggests the same could be true for PIM1. PIM1 was previously found to phosphorylate HIF1α, which protects HIF1α from recognition by VHL and hence from degradation[373]. Additionally, PIM1 has been previously suggested to play a role in HIF signalling of renal-cell carcinomas[392]. PIM1 is also known generally to affect PI3K/AKT signalling pathways[358]. Downstream targets of PI3K/AKT are crucial for regulation of glucose transporters and for activation of key glycolytic enzymes[393]. PIM1 mediated activation of AKT signalling is therefore a way for the cell to produce energy, for instance in hypoxia conditions or when HIF1α is dysregulated, since HIF1α as a key regulator of angiogenesis is often seen to be activated in cancer[373]. These findings also support the general idea that PIM1-HIF1α signalling might be responsible for decreased ASK1 phosphorylation.

Taken together these results suggest that PIM1 Ser97Asn increases enzyme activity by differentially regulating genes. The affected genes show a high significance within the context of cancer and direct hyperactivity could be proven through changes in p38 phosphorylation. Although the results show some ambiguities, there is sufficient evidence likely to suggest that patients with Ser97Asn variants could be given PIM1 inhibitors.

5.5.5 Conclusion and Outlook

The workflow presented here is capable of rapidly and reliably giving insights on whether certain mutations increase activity. My hope is that this knowledge could be used to adjust treatment regimens and thus this workflow presents a novel method and a valuable addition to clinical research. Naturally, new activating mutations would also likely inspire additional downstream research and potentially new therapy development.

Advantages of this approach lie in the reasonable costs and speed. The full gene expression analysis still cost much time and money. I am currently exploring the possibility of using the Sprint Profiler by Nanostring, which would be initially more expensive but generally a faster way to generate gene expression data on a curated set of genes. This technology shares similarities to microarrays, but works without reverse transcription of sample RNA. For instance, this workflow could rely on microarray technology for questions that are primarily of academic interest, but switch to the faster Nanostring technology in clinical contexts, where time might be a more important factor.

In addition, it would also be possible to improve speed and quality of cell-phenotypes by automating cell proliferation assays in combination with imaging. This is possible via *ZenCellOwl*, an incubator-based microscope. Moreover, cell viability data based on cellular ATP could be a fast and more stable alternative to cell counting.

Overall, the results of this chapter show great promise in assessing variant impact for both scientific and potentially clinically applications.

Chapter VI: References

1. Dayhoff, M. O. & Eck, R. V. *Atlas of protein sequence and structure*. (National Biomedical Research Foundation., 1972).

2. Dayhoff, M. O. & Ledley, R. S. Comprotein. in *Proceedings of the December 4-6, 1962, fall joint computer conference on - AFIPS '62 (Fall)* 262–274 (ACM Press, 1962). doi:10.1145/1461518.1461546.

3. BioTools. https://bio.tools/.

4. Rigden, D. J. & Fernández, X. M. The 2021 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Res.* **49**, D1–D9 (2021).

5. Wild, E. J. & Tabrizi, S. J. Therapies targeting DNA and RNA in Huntington's disease. *Lancet Neurol.* **16**, 837–847 (2017).

6. Pereira, S. V.-N., Ribeiro, J. D., Ribeiro, A. F., Bertuzzo, C. S. & Marson, F. A. L. Novel, rare and common pathogenic variants in the CFTR gene screened by high-throughput sequencing technology and predicted by in silico tools. *Sci. Rep.* **9**, 6234 (2019).

7. Fan, C. & Liu, N. Identification of dysregulated microRNAs associated with diagnosis and prognosis in triple-negative breast cancer: An in silico study. *Oncol. Rep.* (2019) doi:10.3892/or.2019.7094.

8. Tang, W. *et al.* The role of upregulated miR-375 expression in breast cancer: An in vitro and in silico study. *Pathol. - Res. Pract.* **216**, 152754 (2020).

9. Sabiha, B., Bhatti, A., Roomi, S., John, P. & Ali, J. In silico analysis of non-synonymous missense SNPs (nsSNPs) in CPE, GNAS genes and experimental validation in type II diabetes mellitus through Next Generation Sequencing. *Genomics* **113**, 2426–2440 (2021).

10. Sur, S. In silico analysis reveals interrelation of enriched pathways and genes in type 1 diabetes. *Immunogenetics* **72**, 399–412 (2020).

11. Campos-Náñez, E., Layne, J. E. & Zisser, H. C. In Silico Modeling of Minimal Effective Insulin Doses Using the UVA/PADOVA Type 1 Diabetes Simulator. *J. Diabetes Sci. Technol.* **12**, 376–380 (2018).

12. Sanger, F. *et al.* Nucleotide sequence of bacteriophage φX174 DNA. *Nature* **265**, 687–695 (1977).

13. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

14. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).

15. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).

16. Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).

17. Zhang, L. *et al.* ADHDgene: a genetic database for attention deficit hyperactivity disorder. *Nucleic Acids Res.* **40**, D1003–D1009 (2012).

18. Ballouz, S., Dobin, A. & Gillis, J. A. Is it time to change the reference genome? *Genome Biol.* **20**, 159 (2019).

19. Gao, G. F. *et al.* Before and After: Comparison of Legacy and Harmonized TCGA Genomic Data Commons' Data. *Cell Syst.* **9**, 24-34.e10 (2019).

20. Anderson-Trocmé, L. *et al.* Legacy Data Confound Genomics Studies. *Mol. Biol. Evol.* **37**, 2–10 (2020).

21. Lazebnik, Y. Can a biologist fix a radio?—Or, what I learned while studying apoptosis. *Cancer Cell* **2**, 179–182 (2002).

22. Orwell, G. *Politics and the English language.* (1946).

23. Raimondi, F. & Russell, R. B. Studying how genetic variants affect mechanism in biological systems. *Essays Biochem.* **62**, 575–582 (2018).

24. Plon, S. E. *et al.* Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum. Mutat.* **29**, 1282–1291 (2008).

25. Loewe, L. & Hill, W. G. The population genetics of mutations: good, bad and indifferent. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 1153–1167 (2010).

26. Chu, D. & Wei, L. Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. *BMC Cancer* **19**, 359 (2019).

27. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.* **12**, 32–42 (2011).

28. No Title.

29. Chiu, W.-C., Chen, S.-H., Lo, M.-C. & Kuo, Y.-T. Classic Ehlers–Dalnos syndrome presenting as atypical chronic haematoma: a case report with novel frameshift mutation in COL5A1. *BMC Pediatr.* **20**, 495 (2020).

30. Imtiaz, A., Kohrman, D. C. & Naz, S. A Frameshift Mutation in GRXCR2 Causes Recessively Inherited Hearing Loss. *Hum. Mutat.* **35**, 618–624 (2014).

31. Dent, K. M. *et al.* Improved molecular diagnosis of dystrophinopathies in an unselected clinical cohort. *Am. J. Med. Genet. Part A* **134A**, 295–298 (2005).

32. Dayhoff, M. O., Schwartz, R. & Orcutt, B. C. *A model of Evolutionary Change in Proteins*. (1978).

33. Henikoff, S. & Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* **89**, 10915–10919 (1992).

34. Petukh, M., Kucukkal, T. G. & Alexov, E. On Human Disease-Causing Amino Acid Variants: Statistical Study of Sequence and Structural Patterns. *Hum. Mutat.* **36**, 524–534 (2015).

35. Vitkup, D., Sander, C. & Church, G. M. The Amino-acid mutational spectrum of human genetic disease. *Genome Biol.* **4**, (2003).

36. Hu, R., Xu, H., Jia, P. & Zhao, Z. KinaseMD: kinase mutations and drug response database. *Nucleic Acids Res.* **49**, D552–D561 (2021).

37. Sousounis, K., Haney, C. E., Cao, J., Sunchu, B. & Tsonis, P. A. Conservation of the three-dimensional structure in non-homologous or unrelated proteins. *Hum. Genomics* **6**, 10 (2012).

38. Hao, Y. *et al.* Gain of Interaction with IRS1 by p110α-Helical Domain Mutants Is Crucial for Their Oncogenic Functions. *Cancer Cell* **23**, 583–593 (2013).

39. Fraser-Pitt, D. & O'Neil, D. Cystic fibrosis – a multiorgan protein misfolding disease. *Futur. Sci. OA* **1**, (2015).

40. Stocker, H. *et al.* Genetic predisposition, Aβ misfolding in blood plasma, and Alzheimer's disease. *Transl. Psychiatry* **11**, 261 (2021).

41. Marinko, J. T. *et al.* Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem. Rev.* **119**, 5537–5606 (2019).

42. Fanen, P., Wohlhuter-Haddad, A. & Hinzpeter, A. Genetics of cystic fibrosis: CFTR mutation classifications toward genotype-based CF therapies. *Int. J. Biochem. Cell Biol.* **52**, 94–102 (2014).

43. D. Amaral, M. & M. Farinha, C. Rescuing Mutant CFTR: A Multi-task Approach to a Better Outcome in Treating Cystic Fibrosis. *Curr. Pharm. Des.* **19**, 3497–3508 (2013).

44. Mitter, D. *et al.* FOXG1 syndrome: genotype–phenotype association in 83 patients with FOXG1 variants. *Genet. Med.* **20**, 98–108 (2018).

45. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–58 (2013).

46. González-Sánchez, J. C., Ibrahim, M. F. R., Leist, I. C., Weise, K. R. & Russell, R. B. Mechnetor: a

web server for exploring protein mechanism and the functional context of genetic variants. *Nucleic Acids Res.* **49**, W366–W374 (2021).

47. Zhong, Q. *et al.* Edgetic perturbation models of human inherited disorders. *Mol. Syst. Biol.* **5**, 321 (2009).

48. Sahni, N. *et al.* Widespread macromolecular interaction perturbations in human genetic disorders. *Cell* **161**, 647–660 (2015).

49. Dihazi, H. *et al.* Integrative omics - from data to biology. *Expert Rev. Proteomics* **15**, 463–466 (2018).

50. Kataka, E., Zaucha, J., Frishman, G., Ruepp, A. & Frishman, D. Edgetic perturbation signatures represent known and novel cancer biomarkers. *Sci. Rep.* **10**, 4350 (2020).

51. Ozturk, K. & Carter, H. Predicting functional consequences of mutations using molecular interaction network features. *Hum. Genet.* (2021) doi:10.1007/s00439-021-02329-5.

52. Cesareni, G., Sacco, F. & Perfetto, L. Assembling Disease Networks From Causal Interaction Resources. *Front. Genet.* **12**, (2021).

53. Tu, J.-J. *et al.* Differential network analysis by simultaneously considering changes in gene interactions and gene expression. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab502.

54. Gan, Y., Xin, Y., Hu, X. & Zou, G. Inferring gene regulatory network from single-cell transcriptomic data by integrating multiple prior networks. *Comput. Biol. Chem.* **93**, 107512 (2021).

55. Betts, M. J. *et al.* Mechismo: predicting the mechanistic impact of mutations and modifications on molecular interactions. *Nucleic Acids Res.* **43**, e10–e10 (2015).

56. Mosca, R. *et al.* dSysMap: exploring the edgetic role of disease mutations. *Nat. Methods* **12**, 167–8 (2015).

57. Gozutok, O., Helmold, B. R. & Ozdinler, P. H. Mutations and Protein Interaction Landscape Reveal Key Cellular Events Perturbed in Upper Motor Neurons with HSP and PLS. *Brain Sci.* **11**, 578 (2021).

58. Chen, Y., Gu, Y., Hu, Z. & Sun, X. Sample-specific perturbation of gene interactions identifies breast cancer subtypes. *Brief. Bioinform.* **22**, (2021).

59. Kennedy, S. A. *et al.* Extensive rewiring of the EGFR network in colorectal cancer cells expressing transforming levels of KRASG13D. *Nat. Commun.* **11**, 499 (2020).

60. Yu, H. *et al.* Rewired Pathways and Disrupted Pathway Crosstalk in Schizophrenia Transcriptomes by Multiple Differential Coexpression Methods. *Genes (Basel).* **12**, 665 (2021).

61. Cordero, P. *et al.* Pathologic gene network rewiring implicates PPP1R3A as a central regulator in pressure overload heart failure. *Nat. Commun.* **10**, 2760 (2019).

62. Deng, Y. *et al.* Identifying mutual exclusivity across cancer genomes: computational approaches to discover genetic interaction and reveal tumor vulnerability. *Brief. Bioinform.* **20**, 254–266 (2019).

63. Uhlen, M. *et al.* Tissue-based map of the human proteome. *Science (80-.).* **347**, 1260419–1260419 (2015).

64. Jay, J. J. & Brouwer, C. Lollipops in the Clinic: Information Dense Mutation Plots for Precision Medicine. *PLoS One* **11**, e0160519 (2016).

65. McGrath, S. P., Benton, M. L., Tavakoli, M. & Tatonetti, N. P. Predictions, Pivots, and a Pandemic: a Review of 2020's Top Translational Bioinformatics Publications. *Yearb. Med. Inform.* **30**, 219–225 (2021).

66. Hufsky, F. *et al.* Computational strategies to combat COVID-19: useful tools to accelerate SARS-CoV-2 and coronavirus research. *Brief. Bioinform.* **22**, 642–663 (2021).

67. Schmenger, T., Diwan, G. D., Singh, G., Apic, G. & Russell, R. B. Never-homozygous genetic variants in healthy populations are potential recessive disease candidates. *npj Genomic Med.* **7**, 54 (2022).

68. Manders, F., van Boxtel, R. & Middelkamp, S. The Dynamics of Somatic Mutagenesis During Life in Humans. *Front. Aging* **2**, (2021).

69. Rozhok, A. I. & DeGregori, J. The Evolution of Lifespan and Age-Dependent Cancer Risk. *Trends in Cancer* **2**, 552–560 (2016).

70. Klose, A. Selective disactivation of neurofibromin GAP activity in neurofibromatosis type 1. *Hum. Mol. Genet.* **7**, 1261–1268 (1998).

71. Colman, S. D., Williams, C. A. & Wallace, M. R. Benign neurofibromas in type 1 neurofibromatosis (NF1) show somatic deletions of the NF1 gene. *Nat. Genet.* **11**, 90–92 (1995).

72. Snajderova, M. *et al.* The importance of advanced parental age in the origin of neurofibromatosis type 1. *Am. J. Med. Genet. Part A* **158A**, 519–523 (2012).

73. Fuchs, C. S. *et al.* A Prospective Study of Family History and the Risk of Colorectal Cancer. *N. Engl. J. Med.* **331**, 1669–1674 (1994).

74. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).

75. Stringer, C. B. & Andrews, P. Genetic and Fossil Evidence for the Origin of Modern Humans. *Science (80-.).* **239**, 1263–1268 (1988).

76. Macaulay, V. *et al.* Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes. *Science (80-.).* **308**, 1034–1036 (2005).

77. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

78. Matsumoto, M. & Nishimura, T. Mersenne twister. *ACM Trans. Model. Comput. Simul.* **8**, 3–30 (1998).

79. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

80. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

81. The PyMOL Molecular Graphics System, Schrödinger, LLC.

82. HMMER. http://hmmer.org/.

83. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

84. Touw, W. G. *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **43**, D364–D368 (2015).

85. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

86. Cheng, H. *et al.* ECOD: An Evolutionary Classification of Protein Domains. *PLoS Comput. Biol.* **10**, e1003926 (2014).

87. U. S. Food and Drug Administration/Center for Drug Evaluation and Researc. https://www.fda.gov/drugs/drug-approvals-and-databases/drugsfda-data-files.

88. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, M. Online Mendelian Inheritance in Man, OMIM®. https://omim.org/.

89. Rehm, H. L. *et al.* ClinGen — The Clinical Genome Resource. *N. Engl. J. Med.* **372**, 2235–2242 (2015).

90. Consortium, T. C. G. R. (ClinGen). ClinGen Dosage Sensitivity Map. https://www.ncbi.nlm.nih.gov/projects/dbvar/clingen/clingen_gene.cgi?sym=RET&subject=.

91. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

92. van der Lee, R. *et al.* Integrative Genomics-Based Discovery of Novel Regulators of the Innate Antiviral Response. *PLOS Comput. Biol.* **11**, e1004553 (2015).

93. Pagliarini, D. J. *et al.* A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. *Cell* **134**, 112–123 (2008).

94. Mendell, J. T. & Dietz, H. C. When the Message Goes Awry. *Cell* **107**, 411–414 (2001).

95. Raimondi, F. *et al.* Genetic variants affecting equivalent protein family positions reflect human diversity. *Sci. Rep.* **7**, 12771 (2017).

96. Sim, N. L. *et al.* SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, (2012).

97. López-Ferrando, V., Gazzo, A., de la Cruz, X., Orozco, M. & Gelpí, J. L. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Res.* **45**, W222–W228 (2017).

98. Yao, J., Subramanian, C., Rock, C. O. & Jackowski, S. Human pantothenate kinase 4 is a pseudo-pantothenate kinase. *Protein Sci.* **28**, 1031–1047 (2019).

99. Zhou, B. *et al.* A novel pantothenate kinase gene (PANK2) is defective in Hallervorden-Spatz syndrome. *Nat. Genet.* **28**, 345–9 (2001).

100. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

101. Hayflick, S. J. *et al.* Genetic, Clinical, and Radiographic Delineation of Hallervorden–Spatz Syndrome. *N. Engl. J. Med.* **348**, 33–40 (2003).

102. Wu, Z., Li, C., Lv, S. & Zhou, B. Pantothenate kinase-associated neurodegeneration: insights from a Drosophila model. *Hum. Mol. Genet.* **18**, 3659–3672 (2009).

103. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

104. Yan, J. *et al.* The 3M Complex Maintains Microtubule and Genome Integrity. *Mol. Cell* **54**, 791–804 (2014).

105. Hanson, D., Murray, P. G., Black, G. C. M. & Clayton, P. E. The Genetics of 3-M Syndrome: Unravelling a Potential New Regulatory Growth Pathway. *Horm. Res. Paediatr.* **76**, 369–378 (2011).

106. Boldt, K. *et al.* An organelle-specific protein landscape identifies novel diseases and molecular mechanisms. *Nat. Commun.* **7**, 11491 (2016).

107. Hildebrandt, F., Benzing, T. & Katsanis, N. Ciliopathies. *N. Engl. J. Med.* **364**, 1533–1543 (2011).

108. Mészáros, B., Erdős, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).

109. Wang, P. *et al.* Impaired plasma membrane localization of ubiquitin ligase complex underlies 3-M syndrome development. *J. Clin. Invest.* **129**, 4393–4407 (2019).

110. Nie, J. *et al.* Ankyrin Repeats of ANKRA2 Recognize a PxLPxL Motif on the 3M Syndrome Protein CCDC8. *Structure* **23**, 700–712 (2015).

111. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).

112. Tuladhar, S. & Kanneganti, T.-D. NLRP12 in innate immunity and inflammation. *Mol. Aspects Med.* **76**, 100887 (2020).

113. Zhang, X., Nan, H., Guo, J. & Liu, J. NLRP12 reduces proliferation and inflammation of rheumatoid arthritis fibroblast-like synoviocytes by regulating the NF-κB and MAPK pathways. *Eur. Cytokine Netw.* **32**, 15–22 (2021).

114. Jeru, I. *et al.* Mutations in NALP12 cause hereditary periodic fever syndromes. *Proc. Natl. Acad. Sci.* **105**, 1614–1619 (2008).

115. Van Kim, C. Le, Colin, Y. & Cartron, J.-P. Rh proteins: Key structural and functional components of the red cell membrane. *Blood Rev.* **20**, 93–110 (2006).

116. Gruswitz, F. *et al.* Function of human Rh based on structure of RhCG at 2.1 A. *Proc. Natl. Acad. Sci.* **107**, 9638–9643 (2010).

117. Betts, M. J., Russell, R. B., Barnes, M. R. & Gray, I. C. Amino acid properties and consequences of subsitutions. in *Bioinformatics for Geneticists* (2003).

118. Wagner, F. F. *et al.* Molecular basis of weak D phenotypes. *Blood* **93**, 385–93 (1999).

119. Urbaniak, S. J. & Greiss, M. A. RhD haemolytic disease of the fetus and the newborn. *Blood Rev.* **14**, 44–61 (2000).

120. Atiakshin, D., Buchwalow, I. & Tiemann, M. Mast cell chymase: morphofunctional characteristics. *Histochem. Cell Biol.* **152**, 253–269 (2019).

121. He, A. & Shi, G.-P. Mast cell chymase and tryptase as targets for cardiovascular and metabolic diseases. *Curr. Pharm. Des.* **19**, 1114–25 (2013).

122. Ahmad, S. & Ferrario, C. M. Chymase inhibitors for the treatment of cardiac diseases: a patent review (2010-2018). *Expert Opin. Ther. Pat.* **28**, 755–764 (2018).

123. Lazaar, A. L. *et al.* Mast Cell Chymase Modifies Cell-Matrix Interactions and Inhibits Mitogen-Induced Proliferation of Human Airway Smooth Muscle Cells. *J. Immunol.* **169**, 1014–1020 (2002).

124. Taipale, J., Lohi, J., Saarinen, J., Kovanen, P. T. & Keski-Oja, J. Human Mast Cell Chymase and Leukocyte Elastase Release Latent Transforming Growth Factor-β1 from the Extracellular Matrix of Cultured Human Epithelial and Endothelial Cells. *J. Biol. Chem.* **270**, 4689–4696 (1995).

125. Saarinen, J., Kalkkinen, N., Welgus, H. G. & Kovanen, P. T. Activation of human interstitial procollagenase through direct cleavage of the Leu83-Thr84 bond by mast cell chymase. *J. Biol. Chem.* **269**, 18134–40 (1994).

126. Heutinck, K. M., ten Berge, I. J. M., Hack, C. E., Hamann, J. & Rowshani, A. T. Serine proteases of the human immune system in health and disease. *Mol. Immunol.* **47**, 1943–1955 (2010).

127. Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* **15**, 786–801 (2014).

128. Rozario, T. & DeSimone, D. W. The extracellular matrix in development and morphogenesis: a dynamic view. *Dev. Biol.* **341**, 126–40 (2010).

129. Lu, P., Takai, K., Weaver, V. M. & Werb, Z. Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).

130. Meyer, N. *et al.* Chymase-producing cells of the innate immune system are required for decidual vascular remodeling and fetal growth. *Sci. Rep.* **7**, 45106 (2017).

131. Ellert, J. Muation des Chymase Gens im aktiven Zentrum und ihre Beziehung zu dermatologischen Erkrankungen. (Technische Universität München, 2008).

132. Pereira, P. J. *et al.* The 2.2 A crystal structure of human chymase in complex with succinyl-Ala-Ala-Pro-Phe-chloromethylketone: structural explanation for its dipeptidyl carboxypeptidase specificity. *J. Mol. Biol.* **286**, 163–73 (1999).

133. Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V. & Wallace, D. C. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**, 223–6 (2004).

134. Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153–7 (2005).

135. Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.* **50**, 746–753 (2018).

136. Kryukov, G. V, Pennacchio, L. A. & Sunyaev, S. R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am. J. Hum. Genet.* **80**, 727–39

(2007).

137. Quintana-Murci, L. & Barreiro, L. B. The role played by natural selection on Mendelian traits in humans. *Ann. N. Y. Acad. Sci.* **1214**, 1–17 (2010).

138. Quintana-Murci, L. Understanding rare and common diseases in the context of human evolution. *Genome Biol.* **17**, 225 (2016).

139. Butchbach, M. E. R. Copy Number Variations in the Survival Motor Neuron Genes: Implications for Spinal Muscular Atrophy and Other Neurodegenerative Diseases. *Front. Mol. Biosci.* **3**, (2016).

140. Wirth, B. *et al.* De Novo Rearrangements Found in 2% of Index Patients with Spinal Muscular Atrophy: Mutational Mechanisms, Parental Origin, Mutation Rate, and Implications for Genetic Counseling. *Am. J. Hum. Genet.* **61**, 1102–1111 (1997).

141. Dupuis, S. *et al.* Impaired response to interferon-alpha/beta and lethal viral disease in human STAT1 deficiency. *Nat. Genet.* **33**, 388–91 (2003).

142. Boisson-Dupuis, S. *et al.* Inborn errors of human STAT1: allelic heterogeneity governs the diversity of immunological and infectious phenotypes. *Curr. Opin. Immunol.* **24**, 364–378 (2012).

143. Vahe, C. *et al.* Diseases associated with calcium-sensing receptor. *Orphanet J. Rare Dis.* **12**, 19 (2017).

144. Herberger, A. L. & Loretz, C. A. Vertebrate extracellular calcium-sensing receptor evolution: selection in relation to life history and habitat. *Comp. Biochem. Physiol. Part D. Genomics Proteomics* **8**, 86–94 (2013).

145. Abouelhoda, M., Faquih, T., El-Kalioby, M. & Alkuraya, F. S. Revisiting the morbid genome of Mendelian disorders. *Genome Biol.* **17**, 235 (2016).

146. Möller, M., Hellberg, Å. & Olsson, M. L. Thorough analysis of unorthodox ABO deletions called by the 1000 Genomes project. *Vox Sang.* **113**, 185–197 (2018).

147. Peng, T., Wang, L. & Li, G. The analysis of APOL1 genetic variation and haplotype diversity provided by 1000 Genomes project. *BMC Nephrol.* **18**, 267 (2017).

148. AlHarthi, F. S., Qari, A., Edress, A. & Abedalthagafi, M. Familial/inherited cancer syndrome: a focus on the highly consanguineous Arab population. *npj Genomic Med.* **5**, 3 (2020).

149. Garber, J. E. & Offit, K. Hereditary cancer predisposition syndromes. *J. Clin. Oncol.* **23**, 276–92 (2005).

150. Rahner, N. & Steinke, V. Hereditary cancer syndromes. *Dtsch. Arztebl. Int.* **105**, 706–14 (2008).

151. von Tilenau, W. G. T. & Ludwig, C. F. *Historia Pathologica Singularis Cutis Turpitudinis Jo. Godofredi Rheinhardi Viri L. Annorum*. (Siegfried Lebrecht Crusius, 1793).

152. MACKLIN, M. T. Inheritance of cancer of the stomach and large intestine in man. *J. Natl. Cancer Inst.* **24**, 551–71 (1960).

153. Knudson, A. G. Hereditary cancer: two hits revisited. *J. Cancer Res. Clin. Oncol.* **122**, 135–40 (1996).

154. Merino, D. & Malkin, D. p53 and hereditary cancer. *Subcell. Biochem.* **85**, 1–16 (2014).

155. Rivlin, N., Brosh, R., Oren, M. & Rotter, V. Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes Cancer* **2**, 466–474 (2011).

156. Lynch, H. T., Shaw, T. G. & Lynch, J. F. Inherited predisposition to cancer: A historical overview. *Am. J. Med. Genet.* **129C**, 5–22 (2004).

157. Daca Alvarez, M. *et al.* The Inherited and Familial Component of Early-Onset Colorectal Cancer. *Cells* **10**, 710 (2021).

158. Kipp, B. R. Differentiating Germline vs Somatic Variants in Cancer Tissue: Are Large-Panel Genetic Tests Helping or Hurting the Cancer Patient? *Clin. Chem.* **61**, 1215–1216 (2015).

159. Moody, E. W. *et al.* Comparison of Somatic and Germline Variant Interpretation in Hereditary Cancer Genes. *JCO Precis. Oncol.* 1–8 (2019) doi:10.1200/PO.19.00144.

160. Rodriguez, J. L. *et al.* CDC Grand Rounds: Family History and Genomics as Tools for Cancer Prevention and Control. *MMWR. Morb. Mortal. Wkly. Rep.* **65**, 1291–1294 (2016).

161. Armaghany, T., Wilson, J. D., Chu, Q. & Mills, G. Genetic alterations in colorectal cancer. *Gastrointest. Cancer Res.* **5**, 19–27 (2012).

162. Macklin, S. K., Kasi, P. M., Jackson, J. L. & Hines, S. L. Incidence of Pathogenic Variants in Those With a Family History of Pancreatic Cancer. *Front. Oncol.* **8**, (2018).

163. Pertea, M. *et al.* CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* **19**, 208 (2018).

164. Burley, S. K. *et al.* RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **49**, D437–D451 (2021).

165. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* **18**, 581–586 (2008).

166. Mitternacht, S. FreeSASA: An open source C library for solvent accessible surface area calculations. *F1000Research* **5**, 189 (2016).

167. Miller, S., Janin, J., Lesk, A. M. & Chothia, C. Interior and surface of monomeric proteins. *J. Mol. Biol.* **196**, 641–656 (1987).

168. Masuda, N., Porter, M. A. & Lambiotte, R. Random walks and diffusion on networks. *Phys. Rep.* **716**–**717**, 1–58 (2017).

169. Thusberg, J., Olatubosun, A. & Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum. Mutat.* **32**, 358–368 (2011).

170. Joerger, A. C. & Fersht, A. R. Structure–function–rescue: the diverse nature of common p53 cancer mutants. *Oncogene* **26**, 2226–2242 (2007).

171. Olivier, M., Hollstein, M. & Hainaut, P. TP53 Mutations in Human Cancers: Origins, Consequences, and Clinical Use. *Cold Spring Harb. Perspect. Biol.* **2**, a001008–a001008 (2010).

172. Kim, M. P. & Lozano, G. Mutant p53 partners in crime. *Cell Death Differ.* **25**, 161–168 (2018).

173. Soussi, T. & Wiman, K. G. TP53: an oncogene in disguise. *Cell Death Differ.* **22**, 1239–49 (2015).

174. Ho, W. C., Fitzgerald, M. X. & Marmorstein, R. Structure of the p53 Core Domain Dimer Bound to DNA. *J. Biol. Chem.* **281**, 20494–20502 (2006).

175. Gencel-Augusto, J. & Lozano, G. p53 tetramerization: at the center of the dominant-negative effect of mutant p53. *Genes Dev.* **34**, 1128–1146 (2020).

176. Zhang, B. *et al.* Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res.* **23**, 1522–1540 (2013).

177. Saraon, P. *et al.* Receptor tyrosine kinases and cancer: oncogenic mechanisms and therapeutic approaches. *Oncogene* **40**, 4079–4093 (2021).

178. Spry, M., Scott, T., Pierce, H. & D'Orazio, J. A. DNA repair pathways and hereditary cancer susceptibility syndromes. *Front. Biosci.* **12**, 4191–207 (2007).

179. Sharma, R., Lewis, S. & Wlodarski, M. W. DNA Repair Syndromes and Cancer: Insights Into Genetics and Phenotype Patterns. *Front. Pediatr.* **8**, (2020).

180. Lüönd, F., Tiede, S. & Christofori, G. Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. *Br. J. Cancer* **125**, 164–175 (2021).

181. Shulman, M., Shi, R. & Zhang, Q. Von Hippel-Lindau tumor suppressor pathways & corresponding therapeutics in kidney cancer. *J. Genet. Genomics* **48**, 552–559 (2021).

182. Elias, R., Zhang, Q. & Brugarolas, J. The von Hippel-Lindau Tumor Suppressor Gene: Implications and Therapeutic Opportunities. *Cancer J.* **26**, 390–398.

183. Kim, H. *et al.* Loss of Von Hippel-Lindau (VHL) Tumor Suppressor Gene Function: VHL-HIF Pathway and Advances in Treatments for Metastatic Renal Cell Carcinoma (RCC). *Int. J. Mol. Sci.* **22**, (2021).

184. Tanimoto, K. Mechanism of regulation of the hypoxia-inducible factor-1alpha by the von Hippel-Lindau tumor suppressor protein. *EMBO J.* **19**, 4298–4309 (2000).

185. Gläsker, S., Vergauwen, E., Koch, C. A., Kutikov, A. & Vortmeyer, A. O. Von Hippel-Lindau Disease: Current Challenges and Future Prospects. *Onco. Targets. Ther.* **13**, 5669–5690 (2020).

186. Semenza, G. L. Heritable disorders of oxygen sensing. *Am. J. Med. Genet. A* **185**, 3334–3339 (2021).

187. Zhang, C. *et al.* The Interplay Between Tumor Suppressor p53 and Hypoxia Signaling Pathways in Cancer. *Front. Cell Dev. Biol.* **9**, (2021).

188. Chung, C. From oxygen sensing to angiogenesis: Targeting the hypoxia signaling pathway in metastatic kidney cancer. *Am. J. Health. Syst. Pharm.* **77**, 2064–2073 (2020).

189. Roe, J.-S. *et al.* Phosphorylation of von Hippel-Lindau protein by checkpoint kinase 2 regulates p53 transactivation. *Cell Cycle* **10**, 3920–3928 (2011).

190. Whyte, M. P. Hypophosphatasia: An overview For 2017. *Bone* **102**, 15–25 (2017).

191. Ferianec, V. & Linhartová, L. Extreme elevation of placental alkaline phosphatase as a marker of preterm delivery, placental insufficiency and low birth weight. *Neuro Endocrinol. Lett.* **32**, 154–7 (2011).

192. Vatin, M. *et al.* Polymorphisms of human placental alkaline phosphatase are associated with in vitro fertilization success and recurrent pregnancy loss. *Am. J. Pathol.* **184**, 362–8 (2014).

193. Salazar, L. *et al.* A novel interaction between fibroblast growth factor receptor 3 and the p85 subunit of phosphoinositide 3-kinase: activation-dependent regulation of ERK by p85 in multiple myeloma cells. *Hum. Mol. Genet.* **18**, 1951–61 (2009).

194. Khalid, S. *et al.* Fibroblast Growth Factor Receptor 3 Mutation as a Prognostic Indicator in Patients with Urothelial Carcinoma: A Systematic Review and Meta-analysis. *Eur. Urol. open Sci.* **21**, 61–68 (2020).

195. Zengin, Z. B. *et al.* Targeted therapies: Expanding the role of FGFR3 inhibition in urothelial carcinoma. *Urol. Oncol.* **40**, 25–36 (2022).

196. French, T. & Savarirayan, R. *Thanatophoric Dysplasia*. *GeneReviews*® (1993).

197. Stembalska, A., Dudarewicz, L. & Śmigiel, R. Lethal and life-limiting skeletal dysplasias: Selected prenatal issues. *Adv. Clin. Exp. Med.* **30**, 641–647 (2021).

198. Krejci, P. The paradox of FGFR3 signaling in skeletal dysplasia: why chondrocytes growth arrest while other cells over proliferate. *Mutat. Res. Rev. Mutat. Res.* **759**, 40–8.

199. Tavormina, P. L. *et al.* Another mutation that results in the substitution of an unpaired cysteine residue in the extracellular domain of FGFR3 in thanatophoric dysplasia type I. *Hum. Mol. Genet.* **4**, 2175–7 (1995).

200. González, B. *et al.* Crystal Structures of Methionine Adenosyltransferase Complexed with Substrates and Products Reveal the Methionine-ATP Recognition and Give Insights into the Catalytic Mechanism. *J. Mol. Biol.* **331**, 407–416 (2003).

201. Dudev, T., Grauffel, C. & Lim, C. How Native and Alien Metal Cations Bind ATP: Implications for Lithium as a Therapeutic Agent. *Sci. Rep.* **7**, 42377 (2017).

202. Wei, Y. *et al.* Crystal structure of RhoA–GDP and its functional implications. *Nat. Struct. Biol.* **4**, 699–703 (1997).

203. Ihara, K. *et al.* Crystal Structure of Human RhoA in a Dominantly Active Form Complexed with a GTP Analogue. *J. Biol. Chem.* **273**, 9656–9666 (1998).

204. Dvorsky, R. & Ahmadian, M. R. Always look on the bright site of Rho: structural implications for a conserved intermolecular interface. *EMBO Rep.* **5**, 1130–1136 (2004).

205. Mukhopadhyay, S. & Ross, E. M. Rapid GTP binding and hydrolysis by G q promoted by receptor and GTPase-activating proteins. *Proc. Natl. Acad. Sci.* **96**, 9539–9544 (1999).

206. Kim, J. *et al.* Regulation of RhoA GTPase and various transcription factors in the RhoA pathway. *J. Cell. Physiol.* **233**, 6381–6392 (2018).

207. Gheyouche, E., Bagueneau, M., Loirand, G., Offmann, B. & Téletchéa, S. Structural Design and Analysis of the RHOA-ARHGEF1 Binding Mode: Challenges and Applications for Protein-Protein Interface Prediction. *Front. Mol. Biosci.* **8**, (2021).

208. Mosaddeghzadeh, N. & Ahmadian, M. R. The RHO Family GTPases: Mechanisms of Regulation and Signaling. *Cells* **10**, 1831 (2021).

209. Aspenström, P. Fast-cycling Rho GTPases. *Small GTPases* **11**, 248–255 (2020).

210. Müller, P. M. *et al.* Systems analysis of RhoGEF and RhoGAP regulatory proteins reveals spatially organized RAC1 signalling from integrin adhesions. *Nat. Cell Biol.* **22**, 498–511 (2020).

211. Svensmark, J. H. & Brakebusch, C. Rho GTPases in cancer: friend or foe? *Oncogene* **38**, 7447–7456 (2019).

212. Nam, S., Kim, J. H. & Lee, D. H. RHOA in Gastric Cancer: Functional Roles and Therapeutic Potential. *Front. Genet.* **10**, (2019).

213. Kilian, L. S., Frank, D. & Rangrez, A. Y. RhoA Signaling in Immune Cell Response and Cardiac Disease. *Cells* **10**, 1681 (2021).

214. Kilian, L. S., Voran, J., Frank, D. & Rangrez, A. Y. RhoA: a dubious molecule in cardiac pathophysiology. *J. Biomed. Sci.* **28**, 33 (2021).

215. Shimokawa, H., Sunamura, S. & Satoh, K. RhoA/Rho-Kinase in the Cardiovascular System. *Circ. Res.* **118**, 352–366 (2016).

216. Fujisawa, M. *et al.* Activation of RHOA–VAV1 signaling in angioimmunoblastic T-cell lymphoma. *Leukemia* **32**, 694–702 (2018).

217. Voena & Chiarle. RHO Family GTPases in the Biology of Lymphoma. *Cells* **8**, 646 (2019).

218. García-Mariscal, A. *et al.* Loss of RhoA promotes skin tumor formation and invasion by upregulation of RhoB. *Oncogene* **37**, 847–860 (2018).

219. WHEELER, A. & RIDLEY, A. Why three Rho proteins? RhoA, RhoB, RhoC, and cell motility. *Exp. Cell Res.* **301**, 43–49 (2004).

220. Kim, J.-G. *et al.* RhoA GTPase phosphorylated at tyrosine 42 by src kinase binds to β-catenin and contributes transcriptional regulation of vimentin upon Wnt3A. *Redox Biol.* **40**, 101842 (2021).

221. Liguori, I. *et al.* Oxidative stress, aging, and diseases. *Clin. Interv. Aging* **Volume 13**, 757–772 (2018).

222. Bros, Haas, Moll & Grabbe. RhoA as a Key Regulator of Innate and Adaptive Immunity. *Cells* **8**, 733 (2019).

223. Gilbert-Ross, M., Marcus, A. I. & Zhou, W. RhoA, a novel tumor suppressor or oncogene as a therapeutic target? *Genes Dis.* **2**, 2–3 (2015).

224. Brakebusch, C. Rho GTPase Signaling in Health and Disease: A Complex Signaling Network. *Cells* **10**, 401 (2021).

225. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

226. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).

227. Sales, A. *et al.* Cell Type-Dependent Integrin Distribution in Adhesion and Migration Responses on Protein-Coated Microgrooved Substrates. *ACS Omega* **4**, 1791–1800 (2019).

228. Mayor, R. & Carmona-Fontaine, C. Keeping in touch with contact inhibition of locomotion. *Trends Cell Biol.* **20**, 319–328 (2010).

229. Alkasalias, T. *et al.* RhoA knockout fibroblasts lose tumor-inhibitory capacity in vitro and promote tumor growth in vivo. *Proc. Natl. Acad. Sci.* **114**, E1413–E1421 (2017).

230. Ciobanasu, C., Faivre, B. & Le Clainche, C. Actin Dynamics Associated with Focal Adhesions. *Int. J. Cell Biol.* **2012**, 1–9 (2012).

231. Pasapera, A. M., Schneider, I. C., Rericha, E., Schlaepfer, D. D. & Waterman, C. M. Myosin II activity regulates vinculin recruitment to focal adhesions through FAK-mediated paxillin phosphorylation. *J. Cell Biol.* **188**, 877–890 (2010).

232. Mullin, B. H., Mamotte, C., Prince, R. L. & Wilson, S. G. Influence of ARHGEF3 and RHOA Knockdown on ACTA2 and Other Genes in Osteoblasts and Osteoclasts. *PLoS One* **9**, e98116 (2014).

233. Rajnicek, A. M., Foubister, L. E. & McCaig, C. D. Temporally and spatially coordinated roles for Rho, Rac, Cdc42 and their effectors in growth cone guidance by a physiological electric field. *J. Cell Sci.* **119**, 1723–1735 (2006).

234. Worseck, J. M., Grossmann, A., Weimann, M., Hegele, A. & Stelzl, U. A Stringent Yeast Two-Hybrid Matrix Screening Approach for Protein–Protein Interaction Discovery. in 63–87 (2012). doi:10.1007/978-1-61779-455-1_4.

235. Kolde, R. Pheatmap. (2019).

236. Haga, R. B. & Ridley, A. J. Rho GTPases: Regulation and roles in cancer cell biology. *Small GTPases* **7**, 207–221 (2016).

237. Ushiku, T. *et al.* RHOA mutation in diffuse-type gastric cancer: a comparative clinicopathology analysis of 87 cases. *Gastric Cancer* **19**, 403–411 (2016).

238. Nishizawa, T. *et al.* DGC-specific RHOA mutations maintained cancer cell survival and promoted cell migration via ROCK inactivation. *Oncotarget* **9**, 23198–23207 (2018).

239. Sakata-Yanagimoto, M. *et al.* Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat. Genet.* **46**, 171–175 (2014).

240. JEONG, D. *et al.* RhoA is associated with invasion and poor prognosis in colorectal cancer. *Int. J. Oncol.* **48**, 714–722 (2016).

241. Kalpana, G., Figy, C., Yeung, M. & Yeung, K. C. Reduced RhoA expression enhances breast cancer metastasis with a concomitant increase in CCR5 and CXCR4 chemokines signaling. *Sci. Rep.* **9**, 16351 (2019).

242. Dosztányi, Z., Chen, J., Dunker, A. K., Simon, I. & Tompa, P. Disorder and Sequence Repeats in Hub Proteins and Their Implications for Network Evolution. *J. Proteome Res.* **5**, 2985–2995 (2006).

243. Hao, T. *et al.* RhoA mutations in diffuse-type gastric cancer. *Dig. Med. Res.* **3**, 4–4 (2020).

244. Wang, K. *et al.* Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46**, 573–582 (2014).

245. Hartmann, S., Ridley, A. J. & Lutz, S. The Function of Rho-Associated Kinases ROCK1 and ROCK2 in the Pathogenesis of Cardiovascular Disease. *Front. Pharmacol.* **6**, (2015).

246. Liu, C.-A., Wang, M.-J., Chi, C.-W., Wu, C.-W. & Chen, J.-Y. Rho/Rhotekin-mediated NF-κB activation confers resistance to apoptosis. *Oncogene* **23**, 8731–8742 (2004).

247. Ito, H., Morishita, R. & Nagata, K. Functions of Rhotekin, an Effector of Rho GTPase, and Its Binding

Partners in Mammals. *Int. J. Mol. Sci.* **19**, 2121 (2018).

248. Toledo, C. M. *et al.* BuGZ Is Required for Bub3 Stability, Bub1 Kinetochore Function, and Chromosome Alignment. *Dev. Cell* **28**, 282–294 (2014).

249. Boulter, E. *et al.* Regulation of Rho GTPase crosstalk, degradation and activity by RhoGDI1. *Nat. Cell Biol.* **12**, 477–483 (2010).

250. Getsios, S. *et al.* Coordinated expression of desmoglein 1 and desmocollin 1 regulates intercellular adhesion. *Differentiation* **72**, 419–433 (2004).

251. Ercan-Sencicek, A. G. *et al.* Homozygous loss of DIAPH1 is a novel cause of microcephaly in humans. *Eur. J. Hum. Genet.* **23**, 165–172 (2015).

252. Palander, O. & Trimble, W. S. DIAPH1 regulates ciliogenesis and trafficking in primary cilia. *FASEB J.* **34**, 16516–16535 (2020).

253. Bai, S. W. *et al.* Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biol.* **9**, 54 (2011).

254. Poelmans, G., Franke, B., Pauls, D. L., Glennon, J. C. & Buitelaar, J. K. AKAPs integrate genetic findings for autism spectrum disorders. *Transl. Psychiatry* **3**, e270–e270 (2013).

255. Russo, L., Farias, J., Ferruzo, P., Monteiro, L. & Forti, F. Revisiting the roles of VHR/DUSP3 phosphatase in human diseases. *Clinics* **73**, (2018).

256. Worbs, T., Hammerschmidt, S. I. & Förster, R. Dendritic cell migration in health and disease. *Nat. Rev. Immunol.* **17**, 30–48 (2017).

257. Vega, F. M., Fruhwirth, G., Ng, T. & Ridley, A. J. RhoA and RhoC have distinct roles in migration and invasion by acting through different targets. *J. Cell Biol.* **193**, 655–665 (2011).

258. El-Sibai, M. *et al.* Dysregulation of Rho GTPases in orofacial cleft patients-derived primary cells leads to impaired cell migration, a potential cause of cleft/lip palate development. *Cells Dev.* **165**, 203656 (2021).

259. Tkach, V., Bock, E. & Berezin, V. The role of RhoA in the regulation of cell morphology and motility. *Cell Motil. Cytoskeleton* **61**, 21–33 (2005).

260. Gajadhar, A. S. *et al.* Phosphotyrosine Signaling Analysis in Human Tumors Is Confounded by Systemic Ischemia-Driven Artifacts and Intra-Specimen Heterogeneity. *Cancer Res.* **75**, 1495–1503 (2015).

261. Takata, K. & Singer, S. J. Localization of high concentrations of phosphotyrosine-modified proteins in mouse megakaryocytes. *Blood* **71**, 818–21 (1988).

262. Shen, Y. & Schaller, M. D. Focal Adhesion Targeting: The Critical Determinant of FAK Regulation and Substrate Phosphorylation. *Mol. Biol. Cell* **10**, 2507–2518 (1999).

263. Schaller, M. D. Paxillin: a focal adhesion-associated adaptor protein. *Oncogene* **20**, 6459–6472 (2001).

264. Barry, S. T. & Critchley, D. R. The RhoA-dependent assembly of focal adhesions in Swiss 3T3 cells is associated with increased tyrosine phosphorylation and the recruitment of both pp125FAK and protein kinase C-delta to focal adhesions. *J. Cell Sci.* **107 (Pt 7**, 2033–45 (1994).

265. Khalil, B. D. *et al.* The regulation of RhoA at focal adhesions by StarD13 is important for astrocytoma cell motility. *Exp. Cell Res.* **321**, 109–122 (2014).

266. Tsubouchi, A. *et al.* Localized suppression of RhoA activity by Tyr31/118-phosphorylated paxillin in cell adhesion and migration. *J. Cell Biol.* **159**, 673–683 (2002).

267. Holinstat, M. *et al.* Suppression of RhoA Activity by Focal Adhesion Kinase-induced Activation of p190RhoGAP. *J. Biol. Chem.* **281**, 2296–2305 (2006).

268. LIANG, L. *et al.* Loss of ARHGDIA expression is associated with poor prognosis in HCC and promotes invasion and metastasis of HCC cells. *Int. J. Oncol.* **45**, 659–666 (2014).

269. Lu, W. *et al.* Downregulation of ARHGDIA contributes to human glioma progression through activation of Rho GTPase signaling pathway. *Tumor Biol.* **37**, 15783–15793 (2016).

270. Katoh, M. Dysregulation of stem cell signaling network due to germline mutation, SNP, helicobacter pylori infection, epigenetic change, and genetic alteration in gastric cancer. *Cancer Biol. Ther.* **6**, 832–839 (2007).

271. Katoh, M. & Katoh, M. Conserved POU/OCT- and GATA-binding sites in 5'-flanking promoter region of mammalian WNT8B orthologs. *Int. J. Oncol.* **30**, 1273–7 (2007).

272. Wang, F. *et al.* RhoA promotes epidermal stem cell proliferation via PKN1-cyclin D1 signaling. *PLoS One* **12**, e0172613 (2017).

273. Guo, H., Ray, R. M. & Johnson, L. R. RhoA stimulates IEC-6 cell proliferation by increasing polyamine-dependent Cdk2 activity. *Am. J. Physiol. Gastrointest. Liver Physiol.* **285**, G704-13 (2003).

274. Richter, L., Oberländer, V. & Schmidt, G. RhoA/C inhibits proliferation by inducing the synthesis of GPRC5A. *Sci. Rep.* **10**, 12532 (2020).

275. Wang, J. Q., Derges, J. D., Bodepudi, A., Pokala, N. & Mao, L.-M. Roles of non-receptor tyrosine kinases in pathogenesis and treatment of depression. *J. Integr. Neurosci.* **21**, 025 (2022).

276. Pomella, S. *et al.* New Insights on the Nuclear Functions and Targeting of FAK in Cancer. *Int. J. Mol. Sci.* **23**, 1998 (2022).

277. Roskoski Jr., R. Properties of FDA-approved small molecule protein kinase inhibitors: A 2022 update. *Pharmacol. Res.* **175**, 106037 (2022).

278. Jonasch, E., Atkins, M. B., Chowdhury, S. & Mainwaring, P. Combination of Anti-Angiogenics and Checkpoint Inhibitors for Renal Cell Carcinoma: Is the Whole Greater Than the Sum of Its Parts? *Cancers (Basel).* **14**, 644 (2022).

279. Amelia, T., Kartasasmita, R. E., Ohwada, T. & Tjahjono, D. H. Structural Insight and Development of EGFR Tyrosine Kinase Inhibitors. *Molecules* **27**, 819 (2022).

280. Trenker, R. & Jura, N. Receptor tyrosine kinase activation: From the ligand perspective. *Curr. Opin. Cell Biol.* **63**, 174–185 (2020).

281. Kovacs, T., Zakany, F. & Nagy, P. It Takes More than Two to Tango: Complex, Hierarchal, and Membrane-Modulated Interactions in the Regulation of Receptor Tyrosine Kinases. *Cancers (Basel).* **14**, 944 (2022).

282. Du, Z. & Lovly, C. M. Mechanisms of receptor tyrosine kinase activation in cancer. *Mol. Cancer* **17**, 58 (2018).

283. Porta, C., Paglino, C. & Mosca, A. Targeting PI3K/Akt/mTOR Signaling in Cancer. *Front. Oncol.* **4**, 64 (2014).

284. Ersahin, T., Tuncbag, N. & Cetin-Atalay, R. The PI3K/AKT/mTOR interactive pathway. *Mol. Biosyst.* **11**, 1946–1954 (2015).

285. Barbosa, R., Acevedo, L. A. & Marmorstein, R. The MEK/ERK Network as a Therapeutic Target in Human Cancer. *Mol. Cancer Res.* **19**, 361–374 (2021).

286. Dillon, M. *et al.* Progress on Ras/MAPK Signaling Research and Targeting in Blood and Solid Cancers. *Cancers (Basel).* **13**, 5059 (2021).

287. Sugiura, R., Satoh, R. & Takasaki, T. ERK: A Double-Edged Sword in Cancer. ERK-Dependent Apoptosis as a Potential Therapeutic Strategy for Cancer. *Cells* **10**, 2509 (2021).

288. Ros, J. *et al.* BRAF, MEK and EGFR inhibition as treatment strategies in BRAF V600E metastatic colorectal cancer. *Ther. Adv. Med. Oncol.* **13**, 175883592199297 (2021).

289. Ramadan, F., Fahs, A., Ghayad, S. E. & Saab, R. Signaling pathways in Rhabdomyosarcoma invasion and metastasis. *Cancer Metastasis Rev.* **39**, 287–301 (2020).

290. Suzuki, S. *et al.* KRAS Inhibitor Resistance in MET -Amplified KRAS G12C Non–Small Cell Lung Cancer Induced By RAS- and Non–RAS-Mediated Cell Signaling Mechanisms. *Clin. Cancer Res.* **27**, 5697–5707 (2021).

291. Choi, Y. L. *et al.* Oncogenic MAP2K1 mutations in human epithelial tumors. *Carcinogenesis* **33**, 956–61 (2012).

292. Arcila, M. E. *et al.* MAP2K1 (MEK1) Mutations Define a Distinct Subset of Lung Adenocarcinoma Associated with Smoking. *Clin. Cancer Res.* **21**, 1935–43 (2015).

293. Fischmann, T. O. *et al.* Crystal Structures of MEK1 Binary and Ternary Complexes with Nucleotides and Inhibitors. *Biochemistry* **48**, 2661–2674 (2009).

294. Baron, U. & Bujard, H. Tet repressor-based system for regulated gene expression in eukaryotic cells: Principles and advances. in 401–421 (2000). doi:10.1016/S0076-6879(00)27292-3.

295. T. Das, A., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Curr. Gene Ther.* **16**, 156–167 (2016).

296. Bernd, K. & Reisenauer, S. An end to end workflow for differential gene expression using Affymetrix microarrays. (2018) doi:10.18129/B9.bioc.maEndToEnd.

297. Holland, C. H., Szalai, B. & Saez-Rodriguez, J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochim. Biophys. acta. Gene Regul. Mech.* **1863**, 194431 (2020).

298. Dugourd, A. *et al.* Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. *Mol. Syst. Biol.* **17**, e9730 (2021).

299. Liu, A. *et al.* From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. *NPJ Syst. Biol. Appl.* **5**, 40 (2019).

300. IBM. Cplex, I. I. (2009).

301. Gjerga, E. Modelling and Analysis of Large-Scale Models of Signalling Networks. (Rheinisch-Westfälischen Technischen Hochschule Aachen, 2020).

302. Hanna, T. P. *et al.* Mortality due to cancer treatment delay: systematic review and meta-analysis. *BMJ* m4087 (2020) doi:10.1136/bmj.m4087.

303. Martínez, M. T. *et al.* Ten-year assessment of a cancer fast-track programme to connect primary care with oncology: reducing time from initial symptoms to diagnosis and treatment initiation. *ESMO Open* **6**, 100148 (2021).

304. Murchie, P. *et al.* Time from first presentation in primary care to treatment of symptomatic colorectal cancer: effect on disease stage and survival. *Br. J. Cancer* **111**, 461–469 (2014).

305. Thomas, P. & Smart, T. G. HEK293 cell line: a vehicle for the expression of recombinant proteins. *J. Pharmacol. Toxicol. Methods* **51**, 187–200.

306. Tan, E., Chin, C. S. H., Lim, Z. F. S. & Ng, S. K. HEK293 Cell Line as a Platform to Produce Recombinant Proteins and Viral Vectors. *Front. Bioeng. Biotechnol.* **9**, (2021).

307. Shin, S., Kim, S. H., Lee, J. S. & Lee, G. M. Streamlined Human Cell-Based Recombinase-Mediated Cassette Exchange Platform Enables Multigene Expression for the Production of Therapeutic Proteins. *ACS Synth. Biol.* **10**, 1715–1727 (2021).

308. Sandvik, A. K. *et al.* Gene expression analysis and clinical diagnosis. *Clin. Chim. Acta* **363**, 157–164 (2006).

309. Ravkin, H. D., Givton, O., Geffen, D. B. & Rubin, E. Direct comparison shows that mRNA-based diagnostics incorporate information which cannot be learned directly from genomic mutations. *BMC Bioinformatics* **21**, 196 (2020).

310. Ong, E., Smidt, P. & McGrew, J. T. Limiting the metabolic burden of recombinant protein expression during selection yields pools with higher expression levels. *Biotechnol. Prog.* **35**, (2019).

311. Moullan, N. *et al.* Tetracyclines Disturb Mitochondrial Function across Eukaryotic Models: A Call for Caution in Biomedical Research. *Cell Rep.* **10**, 1681–1691 (2015).

312. Brittain, H. K., Scott, R. & Thomas, E. The rise of the genome and personalised medicine. *Clin. Med.* **17**, 545–551 (2017).

313. Bu, R. *et al.* Recurrent Somatic MAP2K1 Mutations in Papillary Thyroid Cancer and Colorectal Cancer. *Front. Oncol.* **11**, 670423 (2021).

314. Houde, N. *et al.* Fine-tuning of MEK signaling is pivotal for limiting B and T cell activation. *Cell Rep.* **38**, 110223 (2022).

315. Marks, J. L. *et al.* Novel MEK1 mutation identified by mutational analysis of epidermal growth factor receptor signaling pathway genes in lung adenocarcinoma. *Cancer Res.* **68**, 5524–8 (2008).

316. Ordan, M. *et al.* Intrinsically active MEK variants are differentially regulated by proteinases and phosphatases. *Sci. Rep.* **8**, 11830 (2018).

317. Holter, M. C. *et al.* Hyperactive MEK1 Signaling in Cortical GABAergic Neurons Promotes Embryonic Parvalbumin Neuron Loss and Defects in Behavioral Inhibition. *Cereb. Cortex* **31**, 3064–3081 (2021).

318. Kang, H. *et al.* Somatic activating mutations in MAP2K1 cause melorheostosis. *Nat. Commun.* **9**, 1390 (2018).

319. Tanaka, N. *et al.* Clinical Acquired Resistance to KRASG12C Inhibition through a Novel KRAS Switch-II Pocket Mutation and Polyclonal Alterations Converging on RAS-MAPK Reactivation. *Cancer Discov.* **11**, 1913–1922 (2021).

320. Nussinov, R., Tsai, C.-J. & Jang, H. Anticancer drug resistance: An update and perspective. *Drug Resist. Updat.* **59**, 100796 (2021).

321. Jing, C. *et al.* MEK inhibitor enhanced the antitumor effect of oxaliplatin and 5-fluorouracil in MEK1 Q56P-mutant colorectal cancer cells. *Mol. Med. Rep.* **19**, 1092–1100 (2019).

322. Shan, J. *et al.* A Mitogen-activated Protein Kinase/Extracellular Signal-regulated Kinase Kinase (MEK)-dependent Transcriptional Program Controls Activation of the Early Growth Response 1 (EGR1) Gene during Amino Acid Limitation. *J. Biol. Chem.* **289**, 24665–24679 (2014).

323. Rockel, J. S., Bernier, S. M. & Leask, A. Egr-1 inhibits the expression of extracellular matrix genes in chondrocytes by TNFα-induced MEK/ERK signalling. *Arthritis Res. Ther.* **11**, R8 (2009).

324. Silva, P. N. G. *et al.* Differential role played by the MEK/ERK/EGR-1 pathway in orthopoxviruses vaccinia and cowpox biology. *Biochem. J.* **398**, 83–95 (2006).

325. Gokce, O., Runne, H., Kuhn, A. & Luthi-Carter, R. Short-Term Striatal Gene Expression Responses to Brain-Derived Neurotrophic Factor Are Dependent on MEK and ERK Activation. *PLoS One* **4**, e5292 (2009).

326. To, S. Q., Knower, K. C. & Clyne, C. D. NFκB and MAPK signalling pathways mediate TNFα-induced Early Growth Response gene transcription leading to aromatase expression. *Biochem. Biophys. Res. Commun.* **433**, 96–101 (2013).

327. Kim, J. H. *et al.* Brain-derived neurotrophic factor uses CREB and Egr3 to regulate NMDA receptor levels in cortical neurons. *J. Neurochem.* **120**, 210–9 (2012).

328. Castleman, V. H. *et al.* Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities. *Am. J. Hum. Genet.* **84**, 197–209 (2009).

329. Jivan, A., Earnest, S., Juang, Y.-C. & Cobb, M. H. Radial Spoke Protein 3 Is a Mammalian Protein Kinase A-anchoring Protein That Binds ERK1/2. *J. Biol. Chem.* **284**, 29437–29445 (2009).

330. Maschler, S. *et al.* Annexin A1 attenuates EMT and metastatic potential in breast cancer. *EMBO Mol. Med.* **2**, 401–14 (2010).

331. Barbosa, C. M. V. *et al.* Extracellular annexin-A1 promotes myeloid/granulocytic differentiation of

hematopoietic stem/progenitor cells via the Ca2+/MAPK signalling transduction pathway. *Cell Death Discov.* **5**, 135 (2019).

332. Kremer, V. *et al.* MEG8 regulates Tissue Factor Pathway Inhibitor 2 (TFPI2) expression in the endothelium. *Sci. Rep.* **12**, 843 (2022).

333. Neaud, V., Duplantier, J. G., Mazzocco, C., Kisiel, W. & Rosenbaum, J. Thrombin Up-regulates Tissue Factor Pathway Inhibitor-2 Synthesis through a Cyclooxygenase-2-dependent, Epidermal Growth Factor Receptor-independent Mechanism. *J. Biol. Chem.* **279**, 5200–5206 (2004).

334. Pou, J. *et al.* Tissue factor pathway inhibitor 2 is induced by thrombin in human macrophages. *Biochim. Biophys. Acta* **1813**, 1254–60 (2011).

335. Muñoz-Maldonado, C., Zimmer, Y. & Medová, M. A Comparative Analysis of Individual RAS Mutations in Cancer Biology. *Front. Oncol.* **9**, 1088 (2019).

336. Pantsar, T. The current understanding of KRAS protein structure and dynamics. *Comput. Struct. Biotechnol. J.* **18**, 189–198 (2020).

337. Der, C. J., Finkel, T. & Cooper, G. M. Biological and biochemical properties of human rasH genes mutated at codon 61. *Cell* **44**, 167–176 (1986).

338. Khan, A. Q. *et al.* RAS-mediated oncogenic signaling pathways in human malignancies. *Semin. Cancer Biol.* **54**, 1–13 (2019).

339. Nakaguro, M. *et al.* The Diagnostic Utility of RAS Q61R Mutation-specific Immunohistochemistry in Epithelial-Myoepithelial Carcinoma. *Am. J. Surg. Pathol.* **Publish Ah**, (2021).

340. Pareja, F. *et al.* Immunohistochemical assessment of HRAS Q61R mutations in breast adenomyoepitheliomas. *Histopathology* **76**, 865–874 (2020).

341. Kiessling, M. K. *et al.* Mutant HRAS as novel target for MEK and mTOR inhibitors. *Oncotarget* **6**, 42183–42196 (2015).

342. Yu, Y. *et al.* Epigenetic silencing of tumor suppressor gene CDKN1A by oncogenic long non-coding RNA SNHG1 in cholangiocarcinoma. *Cell Death Dis.* **9**, 746 (2018).

343. Zamagni, A. *et al.* CDKN1A upregulation and cisplatin-pemetrexed resistance in non-small cell lung cancer cells. *Int. J. Oncol.* **56**, 1574–1584 (2020).

344. Kreis, Louwen & Yuan. The Multifaceted p21 (Cip1/Waf1/CDKN1A) in Cell Differentiation, Migration and Cancer Therapy. *Cancers (Basel).* **11**, 1220 (2019).

345. Leeksma, O. C. *et al.* Human sprouty 4, a new ras antagonist on 5q31, interacts with the dual specificity kinase TESK1. *Eur. J. Biochem.* **269**, 2546–56 (2002).

346. Marques, I. J. *et al.* Identification of SPRY4 as a Novel Candidate Susceptibility Gene for Familial Nonmedullary Thyroid Cancer. *Thyroid* **31**, 1366–1375 (2021).

347. Kumar, R. *et al.* Growth suppression by dual BRAF(V600E) and NRAS(Q61) oncogene expression is mediated by SPRY4 in melanoma. *Oncogene* **38**, 3504–3520 (2019).

348. Shan, J., Dudenhausen, E. & Kilberg, M. S. Induction of early growth response gene 1 (EGR1) by endoplasmic reticulum stress is mediated by the extracellular regulated kinase (ERK) arm of the MAPK pathways. *Biochim. Biophys. Acta - Mol. Cell Res.* **1866**, 371–381 (2019).

349. Shin, S. Y. *et al.* Suppression of Egr-1 transcription through targeting of the serum response factor by oncogenic H-Ras. *EMBO J.* **25**, 1093–103 (2006).

350. Wang, H. *et al.* Low expression of CDHR1 is an independent unfavorable prognostic factor in glioma. *J. Cancer* **12**, 5193–5205 (2021).

351. Mathas, S. *et al.* Aberrantly expressed c-Jun and JunB are a hallmark of Hodgkin lymphoma cells, stimulate proliferation and synergize with NF-kappa B. *EMBO J.* **21**, 4104–13 (2002).

352. Cobellis, G., Missero, C., Simionati, B., Valle, G. & Di Lauro, R. Immediate early genes induced by H-Ras in thyroid cells. *Oncogene* **20**, 2281–2290 (2001).

353. Brenner, S. & Miller, J. H. *Encyclopedia of Genetics*. (Elsevier Science Inc., 2001).

354. Giambernardi, T. A. *et al.* Overview of matrix metalloproteinase expression in cultured human cells. *Matrix Biol.* **16**, 483–96 (1998).

355. Tauro, B. J. *et al.* Oncogenic H-Ras Reprograms Madin-Darby Canine Kidney (MDCK) Cell-derived Exosomal Proteins Following Epithelial-Mesenchymal Transition. *Mol. Cell. Proteomics* **12**, 2148–2159 (2013).

356. Lund, P. *et al.* Oncogenic HRAS suppresses clusterin expression through promoter hypermethylation. *Oncogene* **25**, 4890–4903 (2006).

357. Regala, R. P. *et al.* Matrix metalloproteinase-10 promotes Kras-mediated bronchio-alveolar stem cell expansion and lung cancer formation. *PLoS One* **6**, e26439 (2011).

358. Warfel, N. A. & Kraft, A. S. PIM kinase (and Akt) biology and signaling in tumors. *Pharmacol. Ther.* **151**, 41–49 (2015).

359. Brault, L. *et al.* PIM serine/threonine kinases in the pathogenesis and therapy of hematologic malignancies and solid cancers. *Haematologica* **95**, 1004–1015 (2010).

360. Magnuson, N. S., Wang, Z., Ding, G. & Reeves, R. Why target PIM1 for cancer diagnosis and treatment? *Futur. Oncol.* **6**, 1461–1478 (2010).

361. TURSYNBAY, Y. *et al.* Pim-1 kinase as cancer drug target: An update. *Biomed. Reports* **4**, 140–146 (2016).

362. Bachmann, M., Hennemann, H., Xing, P. X., Hoffmann, I. & Möröy, T. The Oncogenic Serine/Threonine Kinase Pim-1 Phosphorylates and Inhibits the Activity of Cdc25C-associated Kinase 1 (C-TAK1). *J. Biol. Chem.* **279**, 48319–48328 (2004).

363. Santio, N. M. *et al.* PIM1 accelerates prostate cancer cell motility by phosphorylating actin capping proteins. *Cell Commun. Signal.* **18**, 121 (2020).

364. Weirauch, U. *et al.* Functional Role and Therapeutic Potential of the Pim-1 Kinase in Colon Carcinoma. *Neoplasia* **15**, 783-IN28 (2013).

365. Brault, L. *et al.* PIM kinases are progression markers and emerging therapeutic targets in diffuse large B-cell lymphoma. *Br. J. Cancer* **107**, 491–500 (2012).

366. Bruno, A. *et al.* Mutational analysis of primary central nervous system lymphoma. *Oncotarget* **5**, 5065–75 (2014).

367. Mareschal, S. *et al.* Identification of Somatic Mutations in Primary Cutaneous Diffuse Large B-Cell Lymphoma, Leg Type by Massive Parallel Sequencing. *J. Invest. Dermatol.* **137**, 1984–1994 (2017).

368. Jacobs, M. D. *et al.* Pim-1 ligand-bound structures reveal the mechanism of serine/threonine kinase inhibition by LY294002. *J. Biol. Chem.* **280**, 13728–34 (2005).

369. Lee, S. J. *et al.* Crystal Structure of Pim1 Kinase in Complex with a Pyrido[4,3-D]Pyrimidine Derivative Suggests a Unique Binding Mode. *PLoS One* **8**, e70358 (2013).

370. Manikanta, K., Naveen Kumar, S. K., Hemshekhar, M., Kemparaju, K. & Girish, K. S. ASK1 inhibition triggers platelet apoptosis via p38-MAPK-mediated mitochondrial dysfunction. *Haematologica* **105**, e419–e423 (2020).

371. Gu, J. J., Wang, Z., Reeves, R. & Magnuson, N. S. PIM1 phosphorylates and negatively regulates ASK1-mediated apoptosis. *Oncogene* **28**, 4261–4271 (2009).

372. Hattori, K., Naguro, I., Runchel, C. & Ichijo, H. The roles of ASK family proteins in stress responses and diseases. *Cell Commun. Signal.* **7**, 9 (2009).

373. Casillas, A. L. *et al.* Direct phosphorylation and stabilization of HIF-1α by PIM1 kinase drives angiogenesis in solid tumors. *Oncogene* **40**, 5142–5152 (2021).

374. Lewis, A. & Elks, P. M. Hypoxia Induces Macrophage tnfa Expression via Cyclooxygenase and Prostaglandin E2 in vivo. *Front. Immunol.* **10**, (2019).

375. Toussaint, M. *et al.* Increased hypoxia-inducible factor 1α expression in lung cells of horses with recurrent airway obstruction. *BMC Vet. Res.* **8**, 64 (2012).

376. Zhang, R. *et al.* AIP1 mediates TNF-α–induced ASK1 activation by facilitating dissociation of ASK1 from its inhibitor 14-3-3. *J. Clin. Invest.* **111**, 1933–1943 (2003).

377. Guicciardi, M. E. & Gores, G. J. AIP1: a new player in TNF signaling. *J. Clin. Invest.* **111**, 1813–1815 (2003).

378. Tzeng, H.-E. *et al.* Interleukin-6 induces vascular endothelial growth factor expression and promotes angiogenesis through apoptosis signal-regulating kinase 1 in human osteosarcoma. *Biochem. Pharmacol.* **85**, 531–540 (2013).

379. Shi, Y., Xu, S., Ngoi, N. Y. L., Zeng, Q. & Ye, Z. PRL-3 dephosphorylates p38 MAPK to promote cell survival under stress. *Free Radic. Biol. Med.* **177**, 72–87 (2021).

380. Martínez-Limón, A., Joaquin, M., Caballero, M., Posas, F. & de Nadal, E. The p38 Pathway: From Biology to Cancer Therapy. *Int. J. Mol. Sci.* **21**, 1913 (2020).

381. Plch, J., Hrabeta, J. & Eckschlager, T. KDM5 demethylases and their role in cancer cell chemoresistance. *Int. J. Cancer* **144**, 221–231 (2019).

382. Geutjes, E.-J., Bajpe, P. K. & Bernards, R. Targeting the epigenome for treatment of cancer. *Oncogene* **31**, 3827–3844 (2012).

383. Morin, R. D. *et al.* Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298–303 (2011).

384. Shen, X. *et al.* KDM5D inhibit epithelial-mesenchymal transition of gastric cancer through demethylation in the promoter of Cul4A in male. *J. Cell. Biochem.* **120**, 12247–12258 (2019).

385. El Omari, K. *et al.* Structure of the leukemia oncogene LMO2: implications for the assembly of a hematopoietic transcription factor complex. *Blood* **117**, 2146–56 (2011).

386. de Bock, C. E. *et al.* HOXA9 Cooperates with Activated JAK/STAT Signaling to Drive Leukemia Development. *Cancer Discov.* **8**, 616–631 (2018).

387. De Smedt, R. *et al.* Pre-clinical evaluation of second generation PIM inhibitors for the treatment of T-cell acute lymphoblastic leukemia and lymphoma. *Haematologica* **104**, e17–e20 (2019).

388. Uz, E. *et al.* Disruption of ALX1 causes extreme microphthalmia and severe facial clefting: expanding the spectrum of autosomal-recessive ALX-related frontonasal dysplasia. *Am. J. Hum. Genet.* **86**, 789–96 (2010).

389. Yuan, H. *et al.* ALX1 Induces Snail Expression to Promote Epithelial-to-Mesenchymal Transition and Invasion of Ovarian Cancer Cells. *Cancer Res.* **73**, 1581–1590 (2013).

390. Yao, W. *et al.* ALX1 promotes migration and invasion of lung cancer cells through increasing snail expression. *Int. J. Clin. Exp. Pathol.* **8**, 12129–39 (2015).

391. Yu, Z. *et al.* A Regulatory Feedback Loop between HIF-1α and PIM2 in HepG2 Cells. *PLoS One* **9**, e88301 (2014).

392. Zhao, B. *et al.* PIM1 mediates epithelial-mesenchymal transition by targeting Smads and c-Myc in the nucleus and potentiates clear-cell renal-cell carcinoma oncogenesis. *Cell Death Dis.* **9**, 307 (2018).

393. Leung, C. O. *et al.* PIM1 regulates glycolysis and promotes tumor progression in hepatocellular carcinoma. *Oncotarget* **6**, 10880–10892 (2015).

394. Bruford, E. A. *et al.* Guidelines for human gene nomenclature. *Nat. Genet.* **52**, 754–758 (2020).

395. IUPAC: Nomenclature and Symbolism for Amino Acids and Peptides. https://iupac.qmul.ac.uk/AminoAcid/A2021.html.

Chapter VII: List of abbreviations

1kG	The 1000 Genomes Project
AF	Allele Frequency
AFR	African Ancestry
AMR	American Ancestry
APS	Ammoniumperoxodisulfat
ATP	Adenosine triphosphate
AUC	Area under the curve
BSA	Bovine Serum Albumin
CMPF	Corrected Mean Particle Fluorescence
COSMIC	Catalogue of Somatic Mutations in Cancer
DMSO	Dimethyl Sulfoxide
DNA	Desoxyribonucleidacid
e.Coli	Escherischia Coli
EAS	East Asian Ancestry
ECM	Extracellular Matrix
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
EUR	European Ancestry
FBS	Fetal Bovine Serum
FDA	U.S. Food and Drug Administration
fs	Frameshift
GAP	GTPase-activiating Protein
GDF	GDI Displacement Factor
GDI	Guanine Nucleotide Dissociation Inhibitor
GEF	Guanine Nucleotide Exchange Factor
gnomAD	Gnome Aggregation Database
GO (term)	Gene Ontology Term
GoF	Gain of Function
HMMer	Biosequence Analysis Using Profile Hidden Markov Models
kDa	Kilodalton
LB	Lysogeni Broth
LoF	Loss of Function

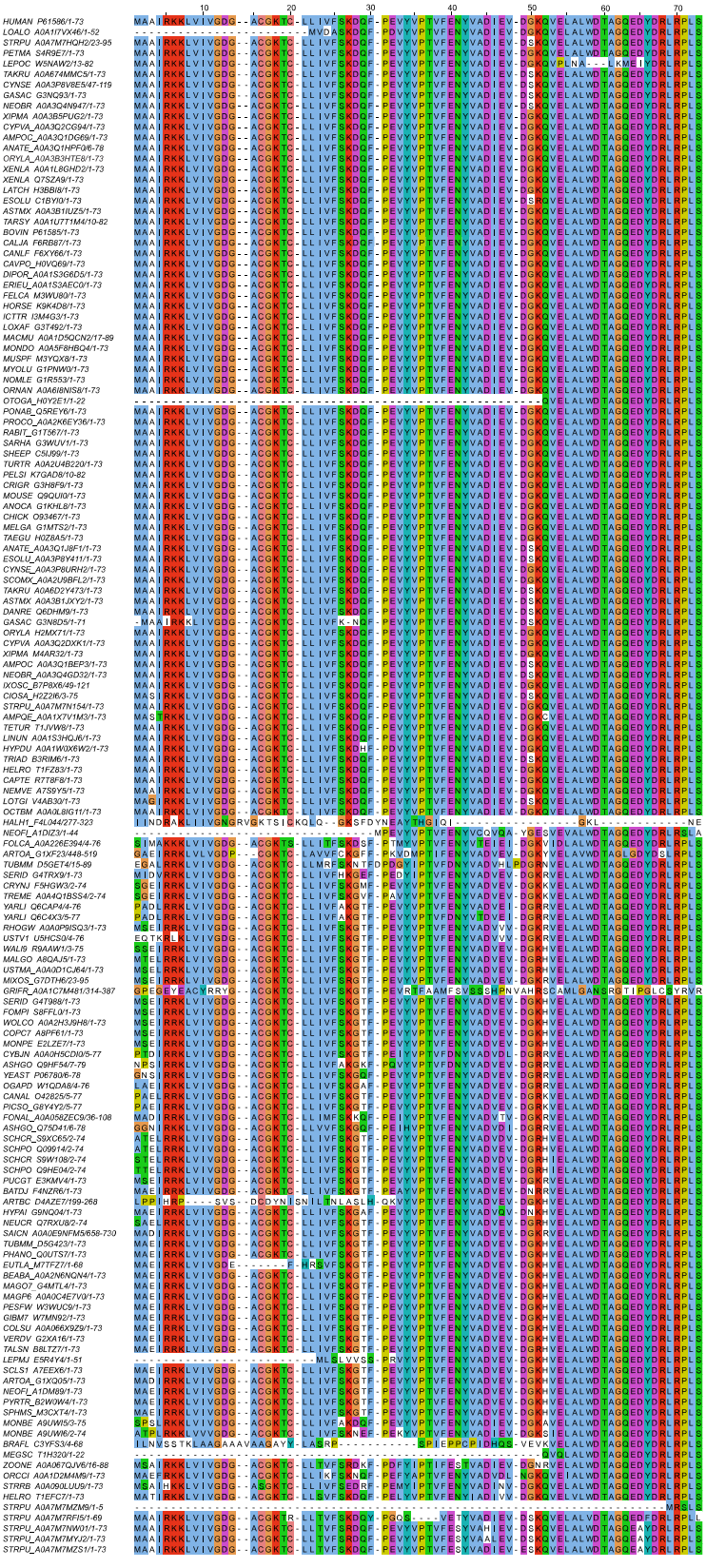
MAF	Major Allele Frequency
Mechismo	Mechanistic Interpretations of Structural Modifications
MeOH	Methanol
MSA	Mulple Sequence Alignment
OMIM	Online Mendelian Inheritance in Man
PBS	Phosphate-buffered Saline
PCR	Polyerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
PFAM	Protein Family Database
PPIs	Protein-Protein Interactions
PTMs	Post-Translational Modifications
RNA	Ribonucleidacid
ROC	Receiver operating characteristic
rpm	Rotations per Minute
RTK	Receptor Tyrosine Kinase
SAS	South Asian Ancestry
SASA	Solvent Accessible Surface Area
SDS	Sodium Dodecyl Sulfate
siRNA	small interfering RNA
SNVs	Single Nucleotide Variants
TBS	Tris-buffered Saline
TBST	Tris-buffered Saline-Triton-X
Temed	N-Tetramethylethylenediamine B
Tris	Tris(hydroxymethyl)aminomethane
VUS	Variant of Unknown Significance
WGS	Whole Genome Shotgun
WT	Wildtype
Y2H	Yeast-Two-Hybrid
YPD	Yeast Extract–Peptone–Dextrose

Proteins are referred to in the text by either their gene name or by their HUGO standard gene symbols[394]. Amino acids are referred to in the text by their IUPAC 3-letter code or by their 1-letter code in figures[395].

Chapter VIII: Appendix

8.1 RHOA MSA with Orthologs

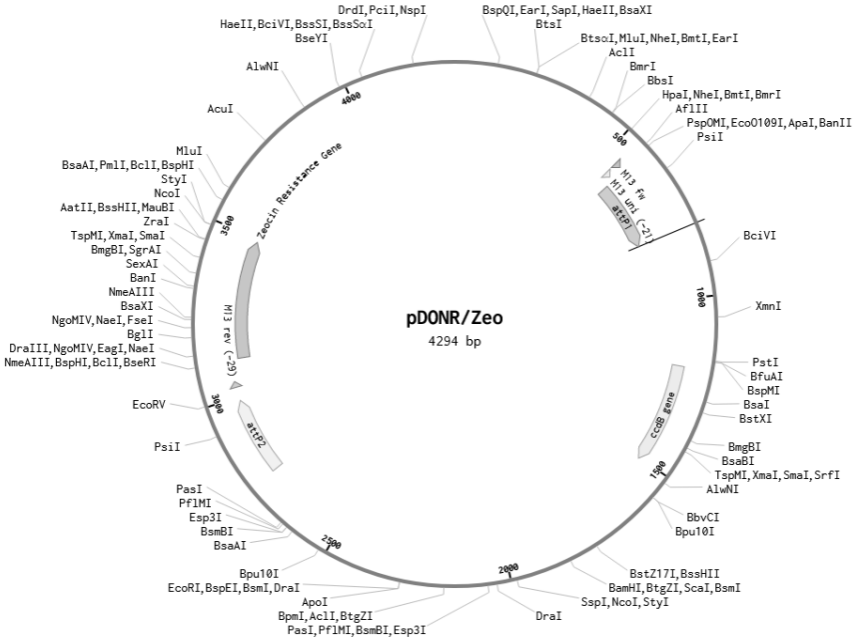
Alignment was performed as described in Schmenger et al. 2022.



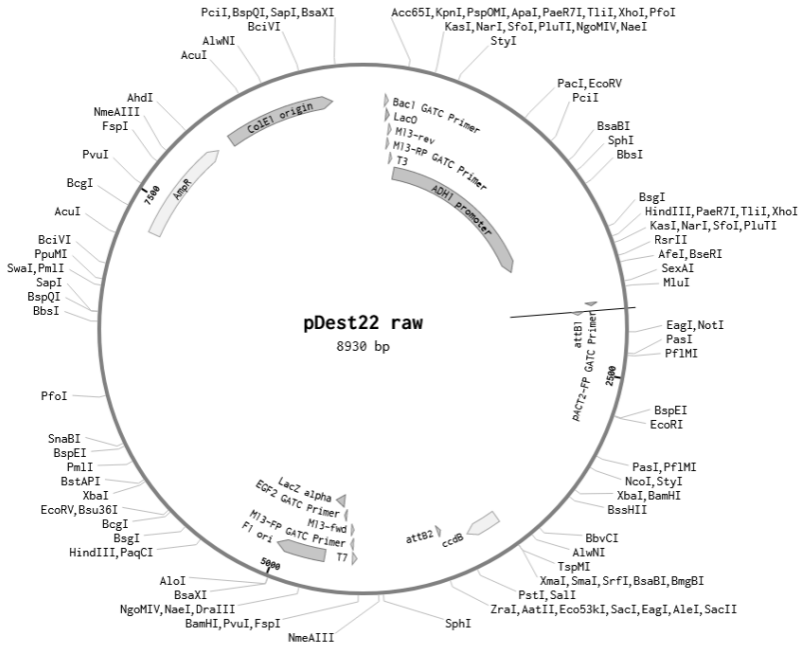
8.2 Plasmid maps

Plasmid sequences were imported into Benchling (benchling.com) to create the following plasmid maps.

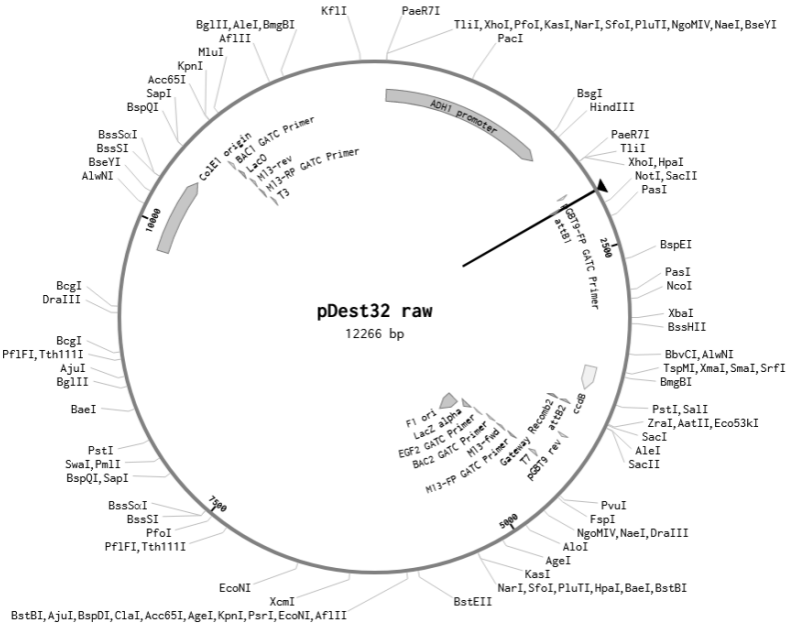
8.2.1 pDONR/Zeo



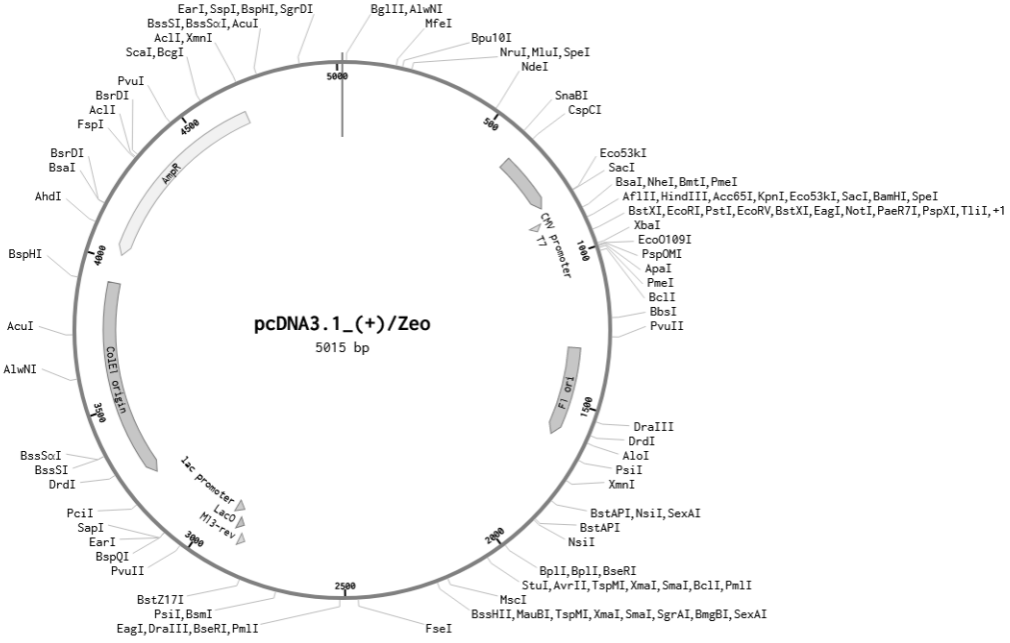
8.2.2 pDest22



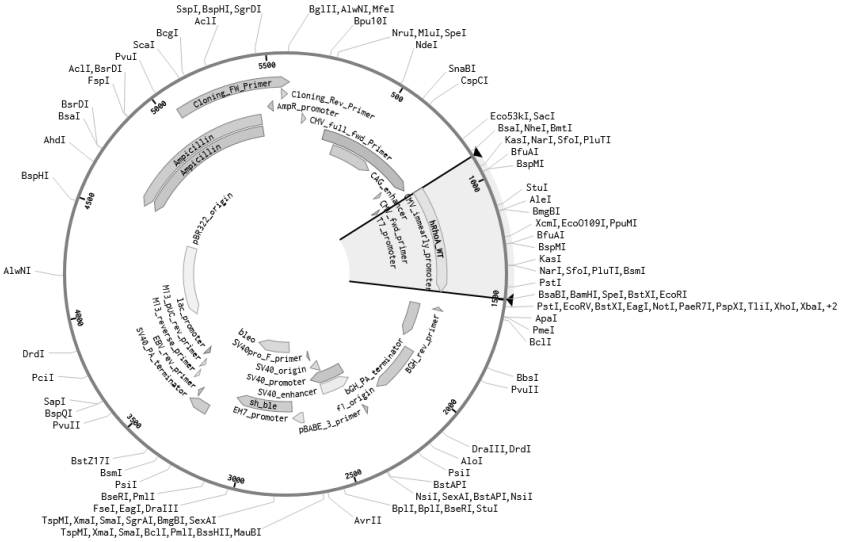
8.2.3 pDest32



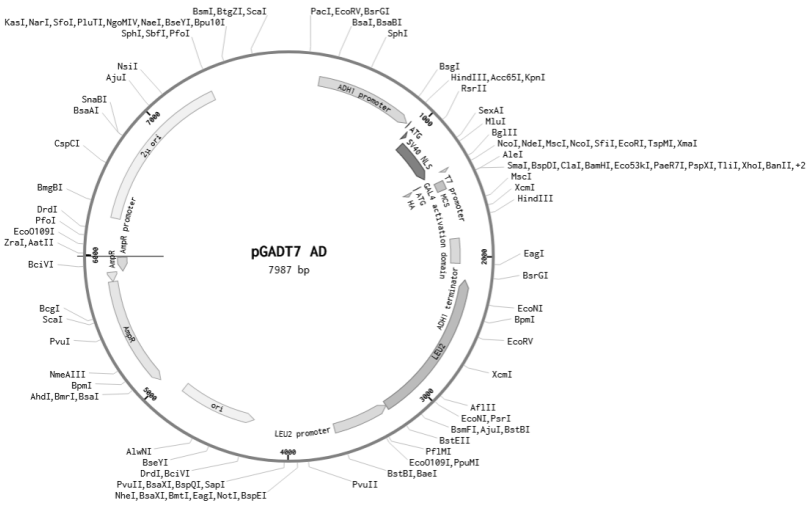
8.2.4 pcDNA3.1(+) /Zeo



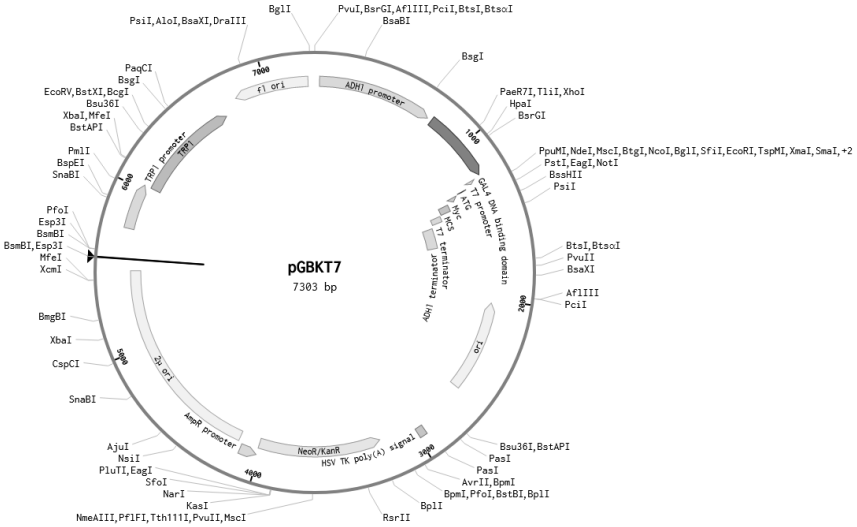
8.2.5 pcDNA3.1(+)/Zeo_hRhoA



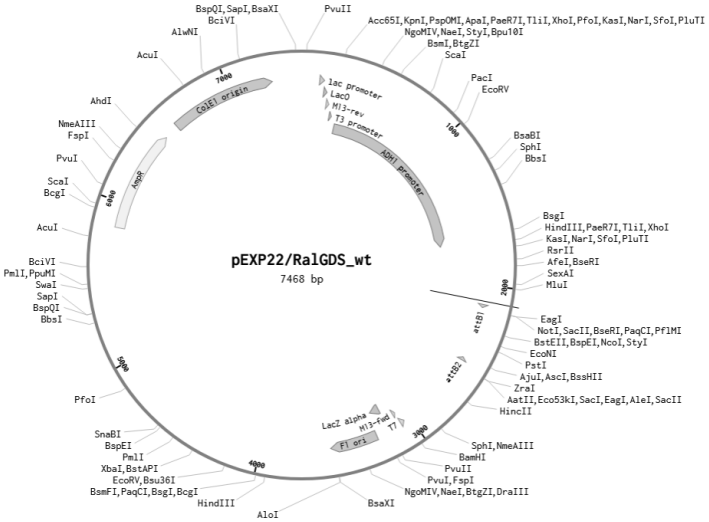
8.2.6 pGADT7 AD



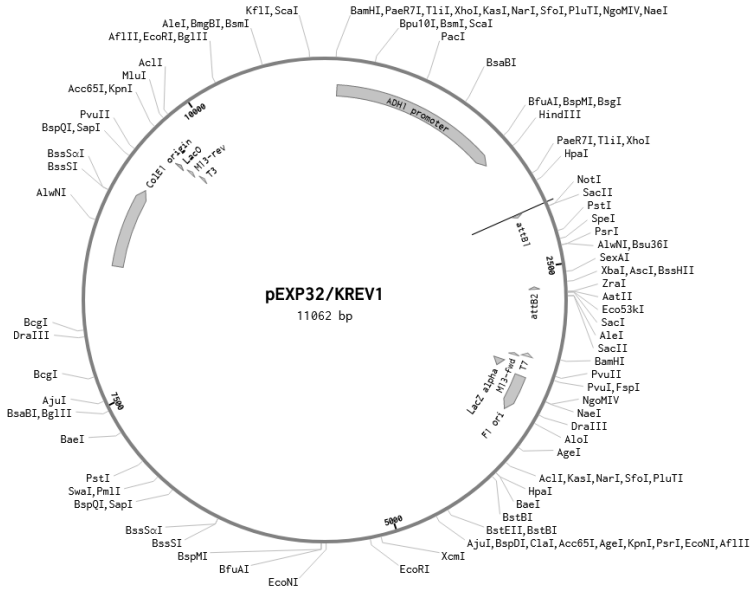
8.2.7 pGBKT7



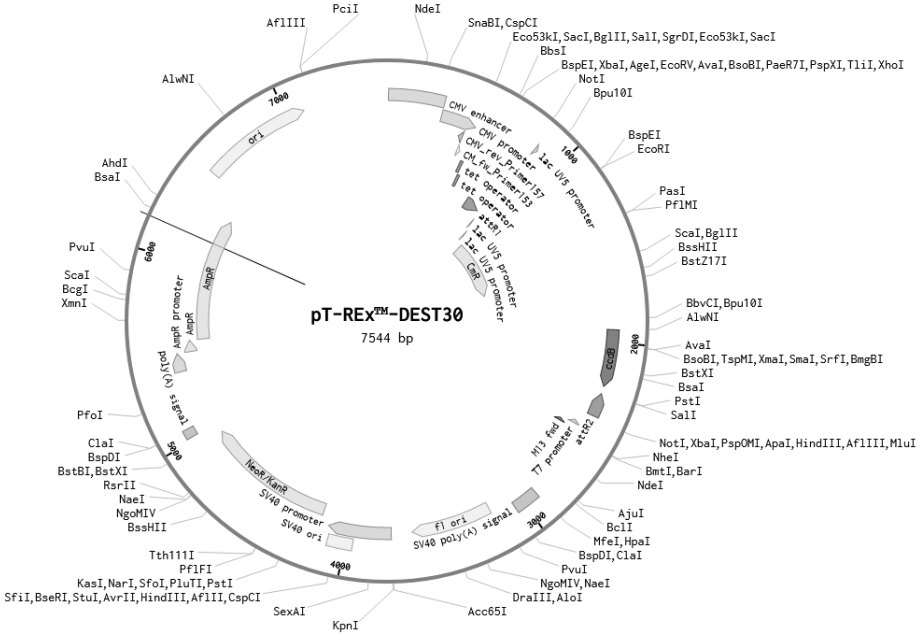
8.2.8 pEXP22/RalGDS wt



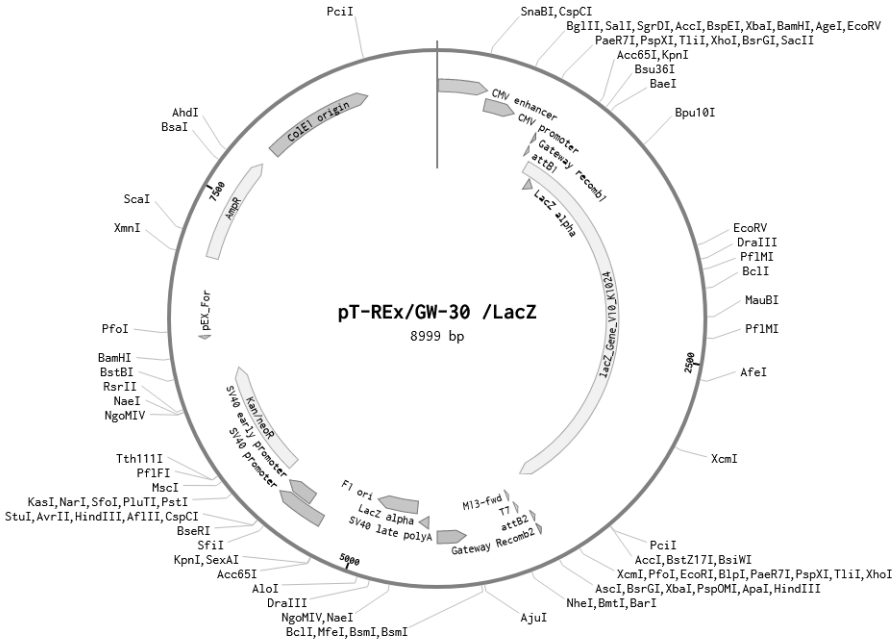
8.2.9 pEXP32/KREV1



8.2.10 pT-REx-DEST30



8.2.11 pT-REx/GW-30 /LacZ



8.3 Complete DNA Sequences

The following fasta sequences were used as given below. Changes of the sequence compared to the WT sequence are **highlighted**.

> human RHOA wildtype

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA A61D

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAG**A**
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA V24F

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATC**TTC**TTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA P75R

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACC**G**CGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human NRAS wildtype

ATGACTGAGTACAAACTGGTGGTGGTTGGAGCAGGTGGTGTTGGGAAAAGCGCACTGACAATCCAGCTAATCCAGAACCACTTTGTAGATGA
ATATGATCCCACCATAGAGGATTCTTACAGAAAACAAGTGGTTATAGATGGTGAAACCTGTTTGTTGGACATACTGGATACAGCTGGACAAGA
AGAGTACAGTGCCATGAGAGACCAATACATGAGGACAGGCGAAGGCTTCCTCTGTGTATTTGCCATCAATAATAGCAAGTCATTTGCGGATAT
TAACCTCTACAGGGAGCAGATTAAGCGAGTAAAAGACTCGGATGATGTACCTATGGTGCTAGTGGGAAACAAGTGTGATTTGCCAACAAGGA
CAGTTGATACAAAACAAGCCCACGAACTGGCCAAGAGTTACGGGATTCCATTCATTGAAACCTCAGCCAAGACCAGACAGGGTGTTGAAGAT
GCTTTTTACACACTGGTAAGAGAAATACGCCAGTACCGAATGAAAAAACTCAACAGCAGTGATGATGGGACTCAGGGTTGTATGGGATTGCCA
TGTGTGGTGATG

> human GNAQ wildtype

ATGACTCTGGAGTCCATCATGGCGTGCTGCCTGAGCGAGGAGGCCAAGGAAGCCCGGCGGATCAACGACGAGATCGAGCGGCACGTCCGC
AGGGACAAGCGGGACGCCCGCCGGGAGCTCAAGCTGCTGCTGCTCGGGACAGGAGAGAGTGGCAAGAGTACGTTTATCAAGCAGATGAGA
ATCATCCATGGGTCAGGATACTCTGATGAAGATAAAAGGGGCTTCACCAAGCTGGTGTATCAGAACATCTTCACGGCCATGCAGGCCATGAT

CAGAGCCATGGACACACTCAAGATCCCATACAAGTATGAGCACAATAAGGCTCATGCACAATTAGTTCGAGAAGTTGATGTGGAGAAGGTGT
CTGCTTTTGAGAATCCATATGTAGATGCAATAAAGAGTTTATGGAATGATCCTGGAATCCAGGAATGCTATGATAGACGACGAGAATATCAATT
ATCTGACTCTACCAAATACTATCTTAATGACTTGGACCGCGTAGCTGACCCTGCCTACCTGCCTACGCAACAAGATGTGCTTAGAGTTCGAGT
CCCCACCACAGGGATCATCGAATACCCCTTTGACTTACAAAGTGTCATTTTCAGAATGGTCGATGTAGGGGGCCAAAGGTCAGAGAGAAGAA
AATGGATACACTGCTTTGAAAATGTCACCTCTATCATGTTTCTAGTAGCGCTTAGTGAATATGATCAAGTTCTCGTGGAGTCAGACAATGAGAA
CCGAATGGAGGAAAGCAAGGCTCTCTTTAGAACAATTATCACATACCCCTGGTTCCAGAACTCCTCGGTTATTCTGTTCTTAAACAAGAAAGAT
CTTCTAGAGGAGAAAATCATGTATTCCCATCTAGTCGACTACTTCCCAGAATATGATGGACCCCAGAGAGATGCCCAGGCAGCCCGAGAATT
CATTCTGAAGATGTTCGTGGACCTGAACCCAGACAGTGACAAAATTATCTACTCCCACTTCACGTGCGCCACAGACACCGAGAATATCCGCTT
TGTCTTTGCTGCCGTCAAGGACACCATCCTCCAGTTGAACCTGAAGGAGTACAATCTGGTC

> human GNAQ Q209L

ATGACTCTGGAGTCCATCATGGCGTGCTGCCTGAGCGAGGAGGCCAAGGAAGCCCGGCGGATCAACGACGAGATCGAGCGGCACGTCCGC
AGGGACAAGCGGGACGCCCGCCGGGAGCTCAAGCTGCTGCTGCTCGGGACAGGAGAGAGTGGCAAGAGTACGTTTATCAAGCAGATGAGA
ATCATCCATGGGTCAGGATACTCTGATGAAGATAAAAGGGGCTTCACCAAGCTGGTGTATCAGAACATCTTCACGGCCATGCAGGCCATGAT
CAGAGCCATGGACACACTCAAGATCCCATACAAGTATGAGCACAATAAGGCTCATGCACAATTAGTTCGAGAAGTTGATGTGGAGAAGGTGT
CTGCTTTTGAGAATCCATATGTAGATGCAATAAAGAGTTTATGGAATGATCCTGGAATCCAGGAATGCTATGATAGACGACGAGAATATCAATT
ATCTGACTCTACCAAATACTATCTTAATGACTTGGACCGCGTAGCTGACCCTGCCTACCTGCCTACGCAACAAGATGTGCTTAGAGTTCGAGT
CCCCACCACAGGGATCATCGAATACCCCTTTGACTTACAAAGTGTCATTTTCAGAATGGTCGATGTAGGGGGCC**T**AAGGTCAGAGAGAAGAA
AATGGATACACTGCTTTGAAAATGTCACCTCTATCATGTTTCTAGTAGCGCTTAGTGAATATGATCAAGTTCTCGTGGAGTCAGACAATGAGAA
CCGAATGGAGGAAAGCAAGGCTCTCTTTAGAACAATTATCACATACCCCTGGTTCCAGAACTCCTCGGTTATTCTGTTCTTAAACAAGAAAGAT
CTTCTAGAGGAGAAAATCATGTATTCCCATCTAGTCGACTACTTCCCAGAATATGATGGACCCCAGAGAGATGCCCAGGCAGCCCGAGAATT
CATTCTGAAGATGTTCGTGGACCTGAACCCAGACAGTGACAAAATTATCTACTCCCACTTCACGTGCGCCACAGACACCGAGAATATCCGCTT
TGTCTTTGCTGCCGTCAAGGACACCATCCTCCAGTTGAACCTGAAGGAGTACAATCTGGTC

> human GNAQ Q209R

ATGACTCTGGAGTCCATCATGGCGTGCTGCCTGAGCGAGGAGGCCAAGGAAGCCCGGCGGATCAACGACGAGATCGAGCGGCACGTCCGC
AGGGACAAGCGGGACGCCCGCCGGGAGCTCAAGCTGCTGCTGCTCGGGACAGGAGAGAGTGGCAAGAGTACGTTTATCAAGCAGATGAGA
ATCATCCATGGGTCAGGATACTCTGATGAAGATAAAAGGGGCTTCACCAAGCTGGTGTATCAGAACATCTTCACGGCCATGCAGGCCATGAT
CAGAGCCATGGACACACTCAAGATCCCATACAAGTATGAGCACAATAAGGCTCATGCACAATTAGTTCGAGAAGTTGATGTGGAGAAGGTGT
CTGCTTTTGAGAATCCATATGTAGATGCAATAAAGAGTTTATGGAATGATCCTGGAATCCAGGAATGCTATGATAGACGACGAGAATATCAATT
ATCTGACTCTACCAAATACTATCTTAATGACTTGGACCGCGTAGCTGACCCTGCCTACCTGCCTACGCAACAAGATGTGCTTAGAGTTCGAGT
CCCCACCACAGGGATCATCGAATACCCCTTTGACTTACAAAGTGTCATTTTCAGAATGGTCGATGTAGGGGGCC**G**AAGGTCAGAGAGAAGAA
AATGGATACACTGCTTTGAAAATGTCACCTCTATCATGTTTCTAGTAGCGCTTAGTGAATATGATCAAGTTCTCGTGGAGTCAGACAATGAGAA
CCGAATGGAGGAAAGCAAGGCTCTCTTTAGAACAATTATCACATACCCCTGGTTCCAGAACTCCTCGGTTATTCTGTTCTTAAACAAGAAAGAT
CTTCTAGAGGAGAAAATCATGTATTCCCATCTAGTCGACTACTTCCCAGAATATGATGGACCCCAGAGAGATGCCCAGGCAGCCCGAGAATT
CATTCTGAAGATGTTCGTGGACCTGAACCCAGACAGTGACAAAATTATCTACTCCCACTTCACGTGCGCCACAGACACCGAGAATATCCGCTT
TGTCTTTGCTGCCGTCAAGGACACCATCCTCCAGTTGAACCTGAAGGAGTACAATCTGGTC

> human ARHGEF25 wildtype

ATGCGGGGGGGGCACAAAGGGGGTCGCTGTGCCTGTCCCCGTGTGATCCGAAAAGTGCTGGCAAAATGCGGCTGCTGCTTCGCCCGGGG
GGGACGTGAATCCTATTCCATTGCGGGCAGTGAGGGGAGTATATCGGCTTCTGCTGCCTCCGGTCTGGCTGCCCCCTCTGGCCCCAGCTCT
GGCCTCAGCTCTGGCCCCTGTTCCCCAGGCCCCCCAGGGCCCGTCAGTGGCCTGAGGAGATGGTTGGATCATTCCAAACATTGTCTCAGTG
TGGAAACTGAGGCAGACAGTGGTCAGGCAGGACCATATGAGAACTGGATGTTGGAGCCAGCTCTAGCCACAGGAGAGGAGCTGCCGGAAC
TGACCTTGCTGACCACACTGTTGGAGGGCCCTGGAGATAAGACGCAGCCACCTGAAGAGGAGACTTTGTCCCAAGCCCCTGAGAGTGAGGA
GGAACAGAAGAAGAAGGCTCTGGAAAGGAGTATGTATGTCCTGAGTGAACTGGTAGAAACAGAGAAAATGTACGTGGACGACTTGGGGCAG
ATTGTGGAGGGTTATATGGCCACCATGGCTGCTCAGGGGGTCCCCGAGAGTCTTCGAGGCCGTGACAGGATTGTGTTTGGGAATATCCAGC
AAATCTATGAGTGGCACCGAGACTATTTCTTGCAAGAGCTACAACGGTGTCTGAAAGATCCTGATTGGCTGGCTCAGCTATTCATCAAACACG
AGCGCCGGCTGCATATGTATGTGGTGTACTATCAGAATAAGCCCAAGTCAGAGCATGTGGTGTCAGAGTTTGGGGACAGCTACTTTGAGGAG
CTCCGGCAGCAGCTGGGGCACCGCCTGCAGCTGAACGACCTCCTCATCAAACCTGTGCAGCGGATCATGAAATACCAGCTGCTGCTCAAGG
ATTTTCTCAAGTATTACAATAGAGCTGGGATGGATACTGCAGACCTAGAGCAAGCTGTGGAGGTCATGTGCTTTGTGCCCAAGCGCTGCAAC
GATATGATGACGCTGGGGAGATTGCGGGGATTTGAGGGCAAACTGACTGCTCAGGGGAAGCTCTTGGGCCAGGACACTTTCTGGGTCACCG
AGCCTGAGGCTGGAGGGCTGCTGTCTTCCCGAGGTCGAGAGAGGCGCGTCTTCCTCTTTGAGCAAATCATCATCTTCAGTGAAGCCCTGGG
AGGAAGAGTGAGAGGTGGAACACAGCCTGGATATGTATACAAGAACAGCATTAAGGTGAGCTGCCTGGGACTGGAGGGGAACCTCCAAGGT
GACCCTTGCCGCTTTGCACTGACCTCCAGAGGGCCAGAGGGTGGGATCCAGCGCTATGTCCTGCAGGCTGCAGACCCTGCTATCAGTCAGG
CCTGGATCAAGCATGTGGCTCAGATCTTGGAGAGCCAACGGGACTTCCTCAACGCATTGCAGTCACCCATTGAGTACCAGAGACGGGAGAG
CCAGACCAACAGCCTGGGGCGGCCAAGAGGGCCTGGAGTGGGGAGCCCTGGAAGAATTCGGCTTGGAGATCAGGCCCAGGGCAGCACAC
ACACACCCATCAATGGCTCTCTCCCCTCTCTGCTGCTGTCACCCAAAGGGGAGGTGGCCAGAGCCCTCTTGCCACTGGATAAACAGGCCCTT

GGTGACATCCCCCAGGCTCCCCATGACTCTCCTCCAGTCTCTCCAACTCCAAAAACCCCTCCCTGCCAAGCCAGACTTGCCAAGCTGGATGA
AGATGAGCTG

> human DIAPH1 wildtype

ATGGAGCCGCCCGGCGGGAGCCTGGGGCCCGGCCGCGAGACCCGGGACAAGAAGAAGGGCCGGAGCCCAGATGAGCTGCCCTCGGCGG
GCGGCGACGGCGGCAAATCTAAGAAATTTCTGGAGAGATTTACCAGCATGAGAATTAAGAAGGAGAAGGAAAAGCCCAATTCTGCTCATAGA
AATTCTTCTGCATCATATGGGGATGATCCCACAGCACAGTCATTGCAAGATGTTTCAGATGAACAAGTGCTGGTTCTCTTTGAACAGATGCTG
CTGGATATGAACCTGAATGAGGAGAAACAGCAACCTTTGAGGGAGAAGGACATCATCATCAAGAGGGAGATGGTGTCCCAATACTTGTACAC
CTCCAAGGCTGGCATGAGCCAGAAGGAGAGCTCTAAGTCTGCCATGATGTATATTCAGGAGTTGAGGTCAGGCTTGCGGGATATGCCTCTGC
TCAGCTGCCTGGAGTCCCTTCGTGTGTCTCTCAACAACAACCCTGTCAGTTGGGTGCAAACATTTGGTGCTGAAGGCTTGGCCTCCTTATTG
GACATTCTTAAACGACTTCATGATGAGAAAGAAGAGACTGCTGGGAGTTACGATAGCCGGAACAAGCATGAGATCATTCGCTGCTTGAAAGCT
TTTATGAACAACAAGTTTGGAATCAAGACCATGTTGGAGACAGAAGAAGGAATCCTACTGCTGGTCAGAGCCATGGATCCTGCTGTTCCCAAC
ATGATGATTGATGCAGCTAAGCTGCTTTCTGCTCTTTGTATTCTACCGCAGCCAGAGGACATGAATGAAAGGGTTTTGGAGGCAATGACAGAA
AGAGCTGAGATGGATGAAGTGGAACGTTTCCAGCCGCTGCTGGATGGATTAAAAAGTGGAACCACTATTGCACTGAAGGTTGGATGCCTACA
GCTGATCAATGCTCTCATCACACCAGCGGAGGAACTTGACTTCCGAGTTCACATCAGAAGTGAACTGATGCGTTTGGGGCTACATCAGGTGT
TGCAGGACCTTCGAGAGATTGAAAATGAAGATATGAGAGTGCAACTAAATGTGTTTGATGAACAAGGGGAAGAGGATTCCTATGACCTGAAG
GGACGGCTGGATGACATTCGCATGGAGATGGATGACTTTAATGAAGTCTTTCAGATTCTCTTAAACACAGTGAAGGATTCAAAGGCAGAGCCA
CACTTCCTTTCCATCCTGCAGCACTTACTCTTGGTCCGAAATGACTATGAGGCCAGACCTCAGTACTATAAGTTGATTGAAGAATGTATTTCCC
AGATAGTTCTGCACAAGAACGGGGCTGATCCTGACTTCAAGTGCCGGCACCTCCAGATTGAGATTGAGGGATTAATTGATCAAATGATTGATA
AGACAAAGGTGGAGAAATCTGAAGCCAAAGCTGCAGAGCTGGAAAAGAAGTTGGACTCAGAGTTAACAGCCCGACATGAGCTACAGGTGGA
AATGAAAAAGATGGAAAGTGACTTTGAGCAGAAGCTTCAAGATCTTCAGGGAGAAAAAGATGCACTGCATTCTGAAAAGCAGCAAATTGCCAC
AGAGAAACAGGACCTGGAAGCAGAGGTGTCCCAGCTCACAGGAGAGGTTGCCAAGCTGACAAAGGAACTGGAAGATGCCAAGAAAGAAATG
GCTTCCCTCTCTGCGGCAGCTATTACTGTACCTCCTTCTGTTCCTAGTCGTGCTCCTGTTCCCCCTGCCCCTCCTTTACCTGGTGACTCTGGC
ACTATTATTCCACCACCACCTGCTCCTGGGGATAGTACCACTCCTCCTCCTCCTCCACCACCACCACCTCCACCACCTCCTTTACCTGGAGGT
ACTGCTATCTCTCCACCCCCTCCTTTGTCTGGGGATGCTACCATCCCTCCACCCCCTCCTTTGCCTGAGGGTGTTGGCATCCCTTCACCCTCT
TCTTTGCCTGGAGGTACTGCCATCCCCCCACCTCCTCCTTTGCCTGGGAGTGCTAGAATCCCCCCACCACCACCTCCTTTGCCTGGGAGTGC
TGGAATTCCCCCCCCACCTCCTCCCTTGCCTGGAGAAGCAGGAATGCCACCTCCTCCTCCCCCTCTTCCTGGTGGTCCTGGAATCCCTCCAC
CTCCTCCATTTCCCGGAGGCCCTGGCATTCCTCCACCTCCACCCGGAATGGGTATGCCTCCACCTCCCCCATTTGGATTTGGAGTTCCTGCA
GCCCCAGTTCTGCCATTTGGATTAACCCCCAAAAAGCTTTATAAGCCAGAGGTGCAGCTCCGGAGGCCAAACTGGTCCAAGCTTGTGGCTGA
GGACCTCTCCCAGGACTGCTTCTGGACAAAGGTGAAGGAGGACCGCTTTGAGAACAATGAACTTTTCGCCAAACTTACCCTTACCTTCTCTGC
CCAGACCAAGACCAAGAAGGATCAAGAAGGTGGAGAAGAAAAGAAATCTGTGCAAAAGAAAAAAGTAAAAGAGTTAAAGGTGTTGGATTCAA
AGACAGCCCAGAATCTCTCAATCTTTTTGGGTTCCTTCCGCATGCCCTATCAAGAGATTAAGAATGTCATCCTGGAGGTGAATGAGGCTGTTC
TGACTGAGTCTATGATCCAGAACCTCATTAAGCAAATGCCAGAGCCAGAGCAGTTAAAAATGCTTTCTGAACTGAAGGATGAATATGATGACC
TGGCTGAGTCAGAGCAGTTTGGCGTGGTGATGGGCACTGTGCCCCGACTGCGGCCTCGCCTCAATGCCATTCTCTTCAAGCTACAATTCAGC
GAGCAAGTGGAGAATATCAAGCCAGAGATTGTGTCTGTCACTGCTGCATGTGAGGAGTTACGTAAGAGTGAGAGCTTTTCCAATCTCCTAGA
GATTACCTTGCTTGTTGGAAATTACATGAATGCTGGCTCCAGAAATGCTGGTGCTTTTGGCTTCAATATCAGCTTCCTCTGTAAGCTTCGAGAC
ACCAAGTCCACAGATCAGAAGATGACGTTGTTACACTTCTTGGCTGAGTTGTGTGAGAATGACTATCCCGATGTCCTCAAGTTTCCAGACGAG
CTTGCCCATGTGGAGAAAGCCAGCCGAGTTTCTGCTGAAAACTTGCAAAAGAACCTAGATCAGATGAAGAAACAAATTTCTGATGTGGAACGT
GATGTTCAGAATTTCCCAGCTGCCACAGATGAAAAAGACAAGTTTGTTGAAAAAATGACCAGCTTTGTGAAGGATGCACAGGAACAGTATAAC
AAGCTGCGGATGATGCATTCTAACATGGAGACCCTCTATAAGGAGCTGGGCGAGTACTTCCTCTTTGACCCCAAGAAGTTGTCTGTTGAAGAA
TTTTTCATGGATCTTCACAATTTTCGGAATATGTTTTTGCAAGCAGTCAAGGAGAACCAGAAGCGGCGGAAGACAGAAGAAAAGATGAGGCGA
GCAAAACTAGCCAAGGAGAAGGCAGAGAAGGAGCGGCTAGAGAAGCAGCAGAAGAGAGAGCAACTCATAGACATGAATGCAGAGGGCGAT
GAGACAGGTGTGATGGACAGTCTTCTAGAAGCCCTGCAGTCAGGGGCAGCATTCCGACGGAAGAGAGGGCCCCGTCAAGCCAACAGGAAG
GCCGGGTGTGCAGTCACATCTCTGCTAGCTTCGGAGCTGACCAAGGATGATGCCATGGCTGCTGTTCCTGCCAAGGTGTCCAAGAACAGTG
AGACATTCCCCACAATCCTTGAGGAAGCCAAGGAGTTGGTTGGCCGTGCAAGC

> human ARHGDIA wildtype

ATGGCTGAGCAGGAGCCCACAGCCGAGCAGCTGGCCCAGATTGCAGCGGAGAACGAGGAGGATGAGCACTCGGTCAACTACAAGCCCCCG
GCCCAGAAGAGCATCCAGGAGATCCAGGAGCTGGACAAGGACGACGAGAGCCTGCGAAAGTACAAGGAGGCCCTGCTGGGCCGCGTGGC
CGTTTCCGCAGACCCCAACGTCCCCAACGTCGTGGTGACTGGCCTGACCCTGGTGTGCAGCTCGGCCCCGGGCCCCCTGGAGCTGGACCT
GACGGGCGACCTGGAGAGCTTCAAGAAGCAGTCGTTTGTGCTGAAGGAGGGTGTGGAGTACCGGATAAAAATCTCTTTCCGGGTTAACCGA
GAGATAGTGTCCGGCATGAAGTACATCCAGCATACGTACAGGAAAGGCGTCAAGATTGACAAGACTGACTACATGGTAGGCAGCTATGGGCC
CCGGGCCGAGGAGTACGAGTTCCTGACCCCCGTGGAGGAGGCACCCAAGGGTATGCTGGCCCGGGGCAGCTACAGCATCAAGTCCCGCTT
CACAGACGACGACAAGACCGACCACCTGTCCTGGGAGTGGAATCTCACCATCAAGAAGGACTGGAAGGAC

> human RHOA Tyr34Cys

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGT**G**CGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAG
CCGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAG
CCTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTG
CGGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATC
GGCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCC
AGACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA E40K

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTC**AG**GAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAG
CCGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAG
CCTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTG
CGGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATC
GGCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCC
AGACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA E40Q

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTC**C**AGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA Y42C

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATT**G**CGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAG
CCGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAG
CCTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTG
CGGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATC
GGCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCC
AGACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA Tyr42Ser

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATT**C**CGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAG
CCGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAG
CCTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTG
CGGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATC
GGCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCC
AGACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA R5Q

ATGGCCGCCATC**CA**AAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA G14E

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATG**AA**GCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC

CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA G14V

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATG**T**CGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA G17E

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCG**A**AAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA G17V

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCG**T**AAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA D67N

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTAC**A**ACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human RHOA Leu69Arg

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGAC**G**GAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAG
CCTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTG
CGGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATC
GGCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCC
AGACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG

> human ARHGAP20 wildtype, 383-551

ATGTTGTTCTTCTTGAATCAAAAGGGTCCATTGACCAAGGGTATCTTTAGACAATCTGCTAACGTTAAGTCCTGCAGAGAATTGAAAGAAAAGT
TGAACTCCGGTGTTGAAGTTCACTTGGATTGCGAATCTATTTTCGTTATCGCCTCCGTTTTGAAGGACTTCTTGAGAAATATTCCAGGTTCTAT
CTTCTCCTCCGACTTGTATGATCATTGGGTTTCTGTTATGGATCAAGGTAACGACGAAGAAAAGATCAACACCGTCCAAAGATTATTGGACCAA
TTGCCAAGAGCTAACGTTGTCTTGTTGAGATATTTGTTCGGTGTCTTGCACAATATCGAACAACACTCCTCTTCTAATCAAATGACCGCTTTTAA

CTTGGCTGTTTGTGTTGCTCCATCTATTTTGTGGCCACCAGCTTCATCTTCACCAGAATTGGAAAATGAATTCACCAAGAAGGTTTCTTTGTTG
ATCCAATTCTTGATCGAAAACTGCTTGAGAATTTTT

> human PAK1 wildtype

ATGTCCAACAACGGTTTGGATATTCAAGATAAGCCACCAGCTCCACCAATGAGAAATACTTCTACTATGATTGGTGCCGGTTCTAAAGATGCT
GGTACTTTGAATCATGGTTCTAAACCATTGCCACCAAATCCAGAAGAAAAGAAGAAGGACAGATTCTACAGGTCTATTTTGCCAGGTGAT
AAGACCAACAAGAAGAAAGAGAAAGAAAGGCCCGAAATCTCTTTGCCATCTGATTTTGAACATACCATCCACGTTGGTTTCGATGCTGTTACT
GGTGAGTTTACTGGTATGCCAGAACAATGGGCTAGATTATTGCAAACCTCTAACATCACCAAGTCCGAGCAAAAAAAGAATCCACAAGCAGTT
TTGGACGTCTTGGAATTCTACAACTCTAAAAAGACCTCCAACAGCCAAAAGTACATGTCTTTCACTGATAAGTCCGCCGAAGATTACAATTCTT
CTAATGCCTTGAATGTGAAGGCCGTTTCTGAAACTCCAGCTGTTCCACCAGTTTCTGAAGATGAAGATGATGACGACGATGATGCTACTCCAC
CACCAGTTATTGCTCCAAGACCAGAACATACAAAGTCTGTTTACACCAGATCCGTTATTGAACCTTTGCCAGTTACTCCAACTAGAGATGTTGC
TACTTCTCCAATTTCTCCAACTGAAAACAATACCACTCCACCAGATGCTTTGACTAGAAACACTGAAAAGCAAAAGAAAAAGCCCAAGATGTCC
GACGAAGAAATCTTGGAAAAACTGAGGTCTATCGTTTCTGTTGGTGATCCTAAAAAGAAGTACACCAGGTTCGAAAAGATTGGTCAAGGTGCT
TCTGGTACAGTTTACACTGCTATGGATGTTGCAACTGGTCAAGAAGTTGCTATCAAGCAAATGAACTTGCAACAGCAGCCAAAGAAAGAGTTG
ATCATCAACGAAATTCTGGTCATGAGGGAAAACAAGAACCCAAACATCGTTAACTACCTGGATTCTTACTTGGTTGGTGATGAATTGTGGGTT
GTCATGGAATATTTGGCTGGTGGTTCTTTGACTGATGTTGTTACTGAAACCTGTATGGACGAAGGTCAAATTGCTGCTGTATGTAGAGAATGC
TTACAGGCCTTGGAATTTCTGCATTCCAATCAAGTTATCCACAGGGATATCAAGTCCGACAACATTTTGTTAGGTATGGATGGTTCTGTTAAGT
TGACCGATTTTGGTTTCTGTGCTCAAATTACCCCTGAACAGTCTAAGAGATCTACAATGGTTGGTACTCCATATTGGATGGCTCCAGAAGTTGT
TACAAGAAAAGCTTACGGTCCAAAGGTTGATATTTGGTCCTTGGGTATTATGGCCATCGAAATGATTGAAGGTGAACCACCATACTTGAACGA
AAATCCATTGAGAGCCTTGTACTTGATTGCTACTAATGGTACACCAGAATTGCAGAACCCAGAAAAGTTGTCTGCTATCTTCAGAGACTTCTTG
AACAGATGCTTGGAAATGGATGTCGAAAAAAGAGGTTCTGCCAAAGAATTGCTGCAACACCAATTTTTGAAGATCGCTAAGCCATTGTCATCTT
TGACTCCATTGATTGCTGCAGCTAAAGAAGCTACTAAGAACAACCAT

> human ROCK1 wildtype, 948 - 1323

TTGACCAAGGACATCGAAATTTTGAGAAGGGAAAACGAAGAACTGACCGAGAAAATGAAGAAAGCTGAAGAAGAGTACAAGTTGGAGAAAGA
GGAAGAGATCTCCAATTTGAAAGCTGCCTTTGAGAAGAACATCAACACTGAAAGAACCTTGAAAACCCAAGCCGTTAACAAGTTGGCTGAAAT
CATGAACAGAAAGGACTTCAAGATCGATAGAAAGAAGGCTAACACCCAAGACTTGAGGAAGAAAGAAAAAGAGAATAGGAAGCTGCAGTTGG
AGTTGAATCAAGAAAGGGAAAAGTTCAACCAGATGGTTGTCAAGCACCAGAAAGAATTGAATGATATGCAAGCCCAATTGGTTGAAGAATGCG
CTCATAGAAATGAATTGCAAATGCAGTTGGCCTCCAAAGAATCCGATATTGAACAATTGAGAGCCAAGTTGTTGGACTTGTCTGATTCTACTTC
TGTTGCTTCTTTTCCATCTGCTGACGAAACTGATGGTAATTTGCCAGAATCTAGAATCGAAGGTTGGTTGTCTGTTCCAAATAGAGGTAACATT
AAGAGGTACGGTTGGAAGAAGCAATACGTTGTTGTTTCCTCCAAAAAGATCCTGTTCTACAACGACGAACAAGACAAAGAACAAAGCAACCCA
TCTATGGTTTTGGACATCGATAAGTTGTTCCACGTTAGACCAGTTACTCAAGGTGATGTTTACAGAGCTGAAACCGAAGAAATTCCAAAGATCT
TCCAAATCTTGTACGCCAACGAAGGTGAATGTAGAAAGGATGTTGAAATGGAACCAGTTCAACAAGCTGAAAAGACCAACTTTCAAAACCACA
AGGGTCACGAATTCATTCCAACCTTGTATCATTTTCCAGCTAACTGTGATGCTTGTGCTAAACCATTGTGGCATGTTTTTAAACCACCACCAGC
TTTGGAGTGTAGAAGATGTCATGTTAAGTGCCATAGAGATCACTTGGACAAAAAAGAGGATTTGATTTGCCCATGCAAGGTTTCTTACGATGTT
ACTTCTGCTAGAGACATGTTGTTGTTGGCATGTTCTCAAGATGAACAAAAGAAGTGGGTTACCCATTTGGTTAAGAAGATCCCAAAAAATCCAC
CATCCGGT

> human ITSN1 wildtype, 1237 - 1571

AAAAGGCAAGGTTACATCCATGAATTGATCCAGACCGAAGAAAGATATATGGCTGACTTGCAATTGGTCGTCGAAGTTTTTCAAAAGAGAATG
GCTGAATCTGGTTTCTTGACTGAAGGTGAAATGGCCTTGATTTTCGTCAACTGGAAAGAACTGATTATGTCCAACACCAAGTTGTTGAAGGCTT
TGAGAGTTAGAAAAAAGACCGGTGGTGAAAAGATGCCAGTTCAAATGATTGGTGATATTTTGGCTGCCGAATTGTCTCATATGCAAGCCTACA
TTAGATTCTGCTCTTGTCAATTGAATGGTGCTGCCTTGTTACAACAAAAGACTGATGAAGATACCGACTTCAAAGAGTTCTTGAAGAAGTTGGC
TTCTGATCCAAGATGTAAGGGTATGCCATTGTCATCTTTTCTGTTGAAGCCAATGCAAAGGATCACTAGATACCCTTTGTTGATCAGATCCATC
TTGGAAAACACTCCAGAATCTCATGCTGATCACTCTTCATTGAAATTGGCTTTGGAAAGGGCTGAAGAATTGTGCTCTCAAGTTAATGAAGGTG
TCAGGGAAAAAGAAAACTCCGATAGATTGGAATGGATTCAAGCTCATGTTCAATGTGAAGGTTTGGCCGAACAATTGATCTTCAATTCTTTGAC
CAACTGCTTGGGTCCAAGAAAGTTGTTGCATTCTGGTAAGTTGTACAAGACCAAGTCCAACAAAGAACTGCACGGTTTTTTGTTCAACGACTT
CTTGTTGTTGACCTACATGGTTAAGCAATTCGCTGTTTCTTCTGGTTCCGAAAAGTTGTTCTCCTCTAAATCTAACGCCCAGTTCAAGATGTAC
AAAACCCCAATTTTCTTGAACGAGGTCTTGGTTAAGTTGCCAACTGATCCATCATCTGATGAACCAGTTTTCCATATCTCCCATATCGATAGAG
TTTACACCTTGAGAACCGACAACATTAACGAAAGAACAGCTTGGGTTCAAAAGATTAAGGCTGCTTCTGAA

> human ARHGEF6 wildtype, 238 - 550

ACCAAGAACTACTACACTGTTGTCTTGCAAAACATCTTGGACACCGAAAAAGAATACGCCAAAGAATTGCAGTCTTTGTTGGTTACCTACTTGA
GGCCATTGCAATCTAACAACAACTTGTCTACTGTTGAGGTCACCTCTTTGTTAGGTAATTTCGAAGAAGTCTGCACCTTCCAACAAACTTTGTG
TCAAGCTTTGGAAGAGTGTTCTAAGTTTCCAGAAAACCAGCATAAGGTTGGTGGTTGTTTGTTGTCTTTGATGCCACATTTCAAGTCCATGTAC
TTGGCTTACTGTGCTAATCATCCATCTGCTGTTAACGTTTTGACCCAACATTCTGATGAATTGGAGCAGTTTATGGAAAATCAAGGTGCTTCTT

CACCAGGCATTTTGATTTTGACTACCAACTTGTCCAAGCCATTCATGAGATTGGAAAAGTATGTCACCTTGTTGCAAGAATTGGAAAGGCATAT
GGAAGATACCCATCCAGATCACCAAGATATTTTGAAAGCTATCGTTGCCTTCAAGACCTTGATGGGTCAATGTCAAGATTTGAGAAAGAGGAA
GCAATTGGAGTTGCAAATCTTGTCCGAACCTATTCAAGCCTGGGAAGGTGAAGATATTAAGAATTTGGGTAACGTCATCTTCATGTCCCAAGT
TATGGTTCAATATGGTGCCTGCGAAGAAAAAGAAGAAAGGTACTTGATGCTGTTCTCCAACGTCTTGATTATGTTGTCTGCTTCTCCAAGAATG
TCCGGTTTTATCTACCAAGGTAAAATTCCAATTGCCGGTACTGTTGTTACTAGGTTGGACGAAATTGAAGGTAACGACTGCACTTTTGAAATCA
CTGGTAACACCGTTGAAAGAATCGTTGTTCACTGCAACAACAACCAGGATTTTCAAGAATGGTTGGAACAGTTGAACAGATTGATTAGAGGTC
CC

> human DOCK7 wildtype, 1373 - 2140

AAGGGTTACCAAACTTCTCCAGATTTGAGATTGACTTGGTTGCAAAACATGGCCGGTAAACATTCTGAAAGATCTAATCATGCTGAAGCTGCC
CAATGTTTGGTTCATTCTGCTGCTTTGGTTGCTGAATACTTGTCTATGTTGGAAGATAGGAAGTACTTGCCAGTTGGTTGTGTTACCTTCCAAA
ACATCTCTTCCAATGTCTTGGAAGAATCCGCTGTTTCTGATGATGTTGTTTCACCAGATGAAGAAGGTATTTGCTCCGGTAAGTACTTTACTGA
ATCTGGTTTGGTTGGTTTGTTGGAACAAGCTGCTGCTTCTTTTTCTATGGCTGGTATGTATGAAGCTGTCAACGAAGTTTACAAGGTCTTGATT
CCAATCCATGAAGCTAACAGAGATGCCAAAAAGTTGTCTACCATCCACGGTAAATTGCAAGAAGCTTTCTCTAAGATCGTTCACCAATCTACTG
GTTGGGAAAGAATGTTCGGTACTTACTTTAGAGTTGGTTTCTACGGTACTAAGTTCGGTGATTTGGATGAACAAGAGTTCGTCTACAAAGAAC
CAGCTATTACTAAGTTGGCCGAAATCTCTCATAGATTGGAAGGTTTTTACGGTGAAAGGTTCGGTGAAGATGTTGTTGAAGTTATCAAGGATTC
TAACCCAGTTGACAAGTGCAAATTGGATCCTAACAAGGCCTACATTCAAATCACTTACGTTGAACCATACTTCGACACCTACGAAATGAAGGAT
AGAATTACCTACTTCGATAAGAACTACAACCTGAGGCGTTTCATGTACTGTACTCCATTCACTTTAGATGGTAGAGCACATGGTGAATTGCACG
AACAATTCAAGAGAAAGACTATCTTGACTACCTCTCATGCTTTCCCATATATCAAGACCAGAGTTAACGTTACCCACAAAGAAGAAATTATCTT
GACCCCAATTGAAGTTGCCATCGAAGATATGCAAAAAAAGACCCAAGAATTGGCTTTCGCTACTCATCAAGATCCAGCTGATCCAAAAATGTT
GCAAATGGTATTGCAAGGTTCTGTTGGTACTACCGTTAATCAAGGTCCATTGGAAGTTGCTCAAGTGTTCTTGTCTGAAATTCCATCAGATCCA
AAGTTGTTCAGACACCACAACAAGTTGAGATTGTGCTTTAAGGATTTCACCAAGAGATGCGAAGATGCCTTGAGAAAAAACAAGTCTTTGATC
GGTCCAGACCAGAAAGAATACCAAAGAGAATTGGAAAGAAACTACCACAGGTTGAAAGAAGCCTTGCAACCATTGATTAACAGAAAGATTCCA
CAGTTGTACAAGGCCGTTTTGCCAGTTACTTGTCATAGAGATTCATTCTCCAGGATGAGCTTGAGAAAGATGGATTTG

> human RHOA wildtype + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA Y34C + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGT**G**CGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAG
CCGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAG
CCTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTG
CGGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATC
GGCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCC
AGACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA E40K + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTC**AA**GAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA E40Q + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTC**C**AGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC

CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA Y42C + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATT**G**CGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAG
CCGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAG
CCTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTG
CGGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATC
GGCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCC
AGACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA Y42S + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATT**C**CGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAG
CCGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAG
CCTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTG
CGGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATC
GGCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCC
AGACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA R5Q + FLAG tag

ATGGCCGCCATC**CA**AAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA G14E + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATG**AA**GCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA G14V + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATG**T**CGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA G17E + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCG**A**AAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG

GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA G17V + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCG**T**AAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA D67N + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTAC**A**ACAGACTGAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAGC
CTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTGC
GGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATCG
GCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCCA
GACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human RHOA L69R + FLAG tag

ATGGCCGCCATCAGAAAGAAGCTGGTCATCGTCGGAGATGGCGCCTGCGGAAAAACCTGCCTGCTGATCGTGTTCAGCAAGGATCAGTTCC
CCGAGGTGTACGTGCCCACCGTGTTCGAGAATTACGTGGCCGACATCGAGGTGGACGGCAAACAGGTTGAACTGGCCCTGTGGGATACAGC
CGGCCAAGAGGACTACGACAGAC**G**GAGGCCTCTGAGCTACCCCGACACCGACGTGATCCTGATGTGCTTCAGCATCGACAGCCCCGACAG
CCTGGAAAACATCCCCGAGAAGTGGACCCCTGAAGTGAAGCACTTCTGCCCCAACGTGCCCATCATCCTCGTGGGCAACAAGAAGGACCTG
CGGAACGACGAGCACACTCGGAGAGAACTGGCCAAGATGAAGCAAGAGCCCGTGAAGCCCGAAGAGGGCAGAGACATGGCCAATAGAATC
GGCGCCTTCGGCTACATGGAATGCAGCGCCAAGACCAAGGATGGCGTGCGGGAAGTGTTTGAGATGGCCACAAGAGCCGCTCTGCAGGCC
AGACGGGGCAAGAAGAAATCTGGCTGTCTGGTGCTG**GATTATAAAGATGATGATGATAAA**

> human MAP2K1 WT

ATGCCCAAGAAGAAGCCGACGCCCATCCAGCTGAACCCGGCCCCCGACGGCTCTGCAGTTAACGGGACCAGCTCTGCGGAGACCAACTTG
GAGGCCTTGCAGAAGAAGCTGGAGGAGCTAGAGCTTGATGAGCAGCAGCGAAAGCGCCTTGAGGCCTTTCTTACCCAGAAGCAGAAGGTGG
GAGAACTGAAGGATGACGACTTTGAGAAGATCAGTGAGCTGGGGGCTGGCAATGGCGGTGTGGTGTTCAAGGTCTCCCACAAGCCTTCTGG
CCTGGTCATGGCCAGAAAGCTAATTCATCTGGAGATCAAACCCGCAATCCGGAACCAGATCATAAGGGAGCTGCAGGTTCTGCATGAGTGCA
ACTCTCCGTACATCGTGGGCTTCTATGGTGCGTTCTACAGCGATGGCGAGATCAGTATCTGCATGGAGCACATGGATGGAGGTTCTCTGGAT
CAAGTCCTGAAGAAAGCTGGAAGAATTCCTGAACAAATTTTAGGAAAAGTTAGCATTGCTGTAATAAAAGGCCTGACATATCTGAGGGAGAAG
CACAAGATCATGCACAGAGATGTCAAGCCCTCCAACATCCTAGTCAACTCCCGTGGGGAGATCAAGCTCTGTGACTTTGGGGTCAGCGGGCA
GCTCATCGACTCCATGGCCAACTCCTTCGTGGGCACAAGGTCCTACATGTCGCCAGAAAGACTCCAGGGGACTCATTACTCTGTGCAGTCAG
ACATCTGGAGCATGGGACTGTCTCTGGTAGAGATGGCGGTTGGGAGGTATCCCATCCCTCCTCCAGATGCCAAGGAGCTGGAGCTGATGTT
TGGGTGCCAGGTGGAAGGAGATGCGGCTGAGACCCCACCCAGGCCAAGGACCCCCGGGAGGCCCCTTAGCTCATACGGAATGGACAGCC
GACCTCCCATGGCAATTTTTGAGTTGTTGGATTACATAGTCAACGAGCCTCCTCCAAAACTGCCCAGTGGAGTGTTCAGTCTGGAATTTCAAG
ATTTTGTGAATAAATGCTTAATAAAAAACCCCGCAGAGAGAGCAGATTTGAAGCAACTCATGGTTCATGCTTTTATCAAGAGATCTGATGCTGA
GGAAGTGGATTTTGCAGGTTGGCTCTGCTCCACCATCGGCCTTAACCAGCCCAGCACACCAACCCATGCTGCTGGCGTC

> human MAP2K1 Q56P

ATGCCCAAGAAGAAGCCGACGCCCATCCAGCTGAACCCGGCCCCCGACGGCTCTGCAGTTAACGGGACCAGCTCTGCGGAGACCAACTTG
GAGGCCTTGCAGAAGAAGCTGGAGGAGCTAGAGCTTGATGAGCAGCAGCGAAAGCGCCTTGAGGCCTTTCTTACCC**C**GAAGCAGAAGGTG
GGAGAACTGAAGGATGACGACTTTGAGAAGATCAGTGAGCTGGGGGCTGGCAATGGCGGTGTGGTGTTCAAGGTCTCCCACAAGCCTTCTG
GCCTGGTCATGGCCAGAAAGCTAATTCATCTGGAGATCAAACCCGCAATCCGGAACCAGATCATAAGGGAGCTGCAGGTTCTGCATGAGTGC
AACTCTCCGTACATCGTGGGCTTCTATGGTGCGTTCTACAGCGATGGCGAGATCAGTATCTGCATGGAGCACATGGATGGAGGTTCTCTGGA
TCAAGTCCTGAAGAAAGCTGGAAGAATTCCTGAACAAATTTTAGGAAAAGTTAGCATTGCTGTAATAAAAGGCCTGACATATCTGAGGGAGAA
GCACAAGATCATGCACAGAGATGTCAAGCCCTCCAACATCCTAGTCAACTCCCGTGGGGAGATCAAGCTCTGTGACTTTGGGGTCAGCGGG
CAGCTCATCGACTCCATGGCCAACTCCTTCGTGGGCACAAGGTCCTACATGTCGCCAGAAAGACTCCAGGGGACTCATTACTCTGTGCAGTC
AGACATCTGGAGCATGGGACTGTCTCTGGTAGAGATGGCGGTTGGGAGGTATCCCATCCCTCCTCCAGATGCCAAGGAGCTGGAGCTGATG
TTTGGGTGCCAGGTGGAAGGAGATGCGGCTGAGACCCCACCCAGGCCAAGGACCCCCGGGAGGCCCCTTAGCTCATACGGAATGGACAGC

CGACCTCCCATGGCAATTTTTGAGTTGTTGGATTACATAGTCAACGAGCCTCCTCCAAAACTGCCCAGTGGAGTGTTCAGTCTGGAATTTCAA
GATTTTGTGAATAAATGCTTAATAAAAAACCCCGCAGAGAGAGCAGATTTGAAGCAACTCATGGTTCATGCTTTTATCAAGAGATCTGATGCTG
AGGAAGTGGATTTTGCAGGTTGGCTCTGCTCCACCATCGGCCTTAACCAGCCCAGCACACCAACCCATGCTGCTGGCGTC

> human HRAS WT

ATGACGGAATATAAGCTGGTGGTGGTGGGCGCCGGCGGTGTGGGCAAGAGTGCGCTGACCATCCAGCTGATCCAGAACCATTTTGTGGACG
AATACGACCCCACTATAGAGGATTCCTACCGGAAGCAGGTGGTCATTGATGGGGAGACGTGCCTGTTGGACATCCTGGATACCGCCGGCCA
GGAGGAGTACAGCGCCATGCGGGACCAGTACATGCGCACCGGGGAGGGCTTCCTGTGTGTGTTTGCCATCAACAACACCAAGTCTTTTGAG
GACATCCACCAGTACAGGGAGCAGATCAAACGGGTGAAGGACTCGGATGACGTGCCCATGGTGCTGGTGGGGAACAAGTGTGACCTGGCT
GCACGCACTGTGGAATCTCGGCAGGCTCAGGACCTCGCCCGAAGCTACGGCATCCCCTACATCGAGACCTCGGCCAAGACCCGGCAGGGA
GTGGAGGATGCCTTCTACACGTTGGTGCGTGAGATCCGGCAGCACAAGCTGCGGAAGCTGAACCCTCCTGATGAGAGTGGCCCCGGCTGC
ATGAGCTGCAAGTGTGTGCTCTCC

> human HRAS Q61R

ATGACGGAATATAAGCTGGTGGTGGTGGGCGCCGGCGGTGTGGGCAAGAGTGCGCTGACCATCCAGCTGATCCAGAACCATTTTGTGGACG
AATACGACCCCACTATAGAGGATTCCTACCGGAAGCAGGTGGTCATTGATGGGGAGACGTGCCTGTTGGACATCCTGGATACCGCCGGCC**G**
GGAGGAGTACAGCGCCATGCGGGACCAGTACATGCGCACCGGGGAGGGCTTCCTGTGTGTGTTTGCCATCAACAACACCAAGTCTTTTGAG
GACATCCACCAGTACAGGGAGCAGATCAAACGGGTGAAGGACTCGGATGACGTGCCCATGGTGCTGGTGGGGAACAAGTGTGACCTGGCT
GCACGCACTGTGGAATCTCGGCAGGCTCAGGACCTCGCCCGAAGCTACGGCATCCCCTACATCGAGACCTCGGCCAAGACCCGGCAGGGA
GTGGAGGATGCCTTCTACACGTTGGTGCGTGAGATCCGGCAGCACAAGCTGCGGAAGCTGAACCCTCCTGATGAGAGTGGCCCCGGCTGC
ATGAGCTGCAAGTGTGTGCTCTCC

> human PIM1 WT

TGCTCTTGTCCAAAATCAACTCGCTTGCCCACCTGCGCGCCGCGCCCTGCAACGACCTGCACGCCACCAAGCTGGCGCCCGGCAAGGAGAA
GGAGCCCCTGGAGTCGCAGTACCAGGTGGGCCCGCTACTGGGCAGCGGCGGCTTCGGCTCGGTCTACTCAGGCATCCGCGTCTCCGACAA
CTTGCCGGTGGCCATCAAACACGTGGAGAAGGACCGGATTTCCGACTGGGGAGAGCTGCCTAATGGCACTCGAGTGCCCATGGAAGTGGTC
CTGCTGAAGAAGGTGAGCTCGGGTTTCTCCGGCGTCATTAGGCTCCTGGACTGGTTCGAGAGGCCCGACAGTTTCGTCCTGATCCTGGAGA
GGCCCGAGCCGGTGCAAGATCTCTTCGACTTCATCACGGAAAGGGGAGCCCTGCAAGAGGAGCTGGCCCGCAGCTTCTTCTGGCAGGTGC
TGGAGGCCGTGCGGCACTGCCACAACTGCGGGGTGCTCCACCGCGACATCAAGGACGAAAACATCCTTATCGACCTCAATCGCGGCGAGCT
CAAGCTCATCGACTTCGGGTCGGGGGCGCTGCTCAAGGACACCGTCTACACGGACTTCGATGGGACCCGAGTGTATAGCCCTCCAGAGTGG
ATCCGCTACCATCGCTACCATGGCAGGTCGGCGGCAGTCTGGTCCCTGGGGATCCTGCTGTATGATATGGTGTGTGGAGATATTCCTTTCGA
GCATGACGAAGAGATCATCAGGGGCCAGGTTTTCTTCAGGCAGAGGGTCTCTTCAGAATGTCAGCATCTCATTAGATGGTGCTTGGCCCTGA
GACCATCAGATAGGCCAACCTTCGAAGAAATCCAGAACCATCCATGGATGCAAGATGTTCTCCTGCCCCAGGAAACTGCTGAGATCCACCTC
CACAGCCTGTCGCCGGGGCCCAGCAAA

> human PIM1 T23I

ATGCTCTTGTCCAAAATCAACTCGCTTGCCCACCTGCGCGCCGCGCCCTGCAACGACCTGCACGCCA**T**CAAGCTGGCGCCCGGCAAGGAGA
AGGAGCCCCTGGAGTCGCAGTACCAGGTGGGCCCGCTACTGGGCAGCGGCGGCTTCGGCTCGGTCTACTCAGGCATCCGCGTCTCCGACA
ACTTGCCGGTGGCCATCAAACACGTGGAGAAGGACCGGATTTCCGACTGGGGAGAGCTGCCTAATGGCACTCGAGTGCCCATGGAAGTGGT
CCTGCTGAAGAAGGTGAGCTCGGGTTTCTCCGGCGTCATTAGGCTCCTGGACTGGTTCGAGAGGCCCGACAGTTTCGTCCTGATCCTGGAG
AGGCCCGAGCCGGTGCAAGATCTCTTCGACTTCATCACGGAAAGGGGAGCCCTGCAAGAGGAGCTGGCCCGCAGCTTCTTCTGGCAGGTG
CTGGAGGCCGTGCGGCACTGCCACAACTGCGGGGTGCTCCACCGCGACATCAAGGACGAAAACATCCTTATCGACCTCAATCGCGGCGAG
CTCAAGCTCATCGACTTCGGGTCGGGGGCGCTGCTCAAGGACACCGTCTACACGGACTTCGATGGGACCCGAGTGTATAGCCCTCCAGAGT
GGATCCGCTACCATCGCTACCATGGCAGGTCGGCGGCAGTCTGGTCCCTGGGGATCCTGCTGTATGATATGGTGTGTGGAGATATTCCTTTC
GAGCATGACGAAGAGATCATCAGGGGCCAGGTTTTCTTCAGGCAGAGGGTCTCTTCAGAATGTCAGCATCTCATTAGATGGTGCTTGGCCCT
GAGACCATCAGATAGGCCAACCTTCGAAGAAATCCAGAACCATCCATGGATGCAAGATGTTCTCCTGCCCCAGGAAACTGCTGAGATCCACC
TCCACAGCCTGTCGCCGGGGCCCAGCAAA

> human PIM1 S97N

ATGCTCTTGTCCAAAATCAACTCGCTTGCCCACCTGCGCGCCGCGCCCTGCAACGACCTGCACGCCACCAAGCTGGCGCCCGGCAAGGAGA
AGGAGCCCCTGGAGTCGCAGTACCAGGTGGGCCCGCTACTGGGCAGCGGCGGCTTCGGCTCGGTCTACTCAGGCATCCGCGTCTCCGACA
ACTTGCCGGTGGCCATCAAACACGTGGAGAAGGACCGGATTTCCGACTGGGGAGAGCTGCCTAATGGCACTCGAGTGCCCATGGAAGTGGT
CCTGCTGAAGAAGGTGA**A**CTCGGGTTTCTCCGGCGTCATTAGGCTCCTGGACTGGTTCGAGAGGCCCGACAGTTTCGTCCTGATCCTGGAG
AGGCCCGAGCCGGTGCAAGATCTCTTCGACTTCATCACGGAAAGGGGAGCCCTGCAAGAGGAGCTGGCCCGCAGCTTCTTCTGGCAGGTG
CTGGAGGCCGTGCGGCACTGCCACAACTGCGGGGTGCTCCACCGCGACATCAAGGACGAAAACATCCTTATCGACCTCAATCGCGGCGAG
CTCAAGCTCATCGACTTCGGGTCGGGGGCGCTGCTCAAGGACACCGTCTACACGGACTTCGATGGGACCCGAGTGTATAGCCCTCCAGAGT

GGATCCGCTACCATCGCTACCATGGCAGGTCGGCGGCAGTCTGGTCCCTGGGGATCCTGCTGTATGATATGGTGTGTGGAGATATTCCTTTC
GAGCATGACGAAGAGATCATCAGGGGCCAGGTTTTCTTCAGGCAGAGGGTCTCTTCAGAATGTCAGCATCTCATTAGATGGTGCTTGGCCCT
GAGACCATCAGATAGGCCAACCTTCGAAGAAATCCAGAACCATCCATGGATGCAAGATGTTCTCCTGCCCCAGGAAACTGCTGAGATCCACC
TCCACAGCCTGTCGCCGGGGCCCAGCAAA

> human PIM1 Q127E

ATGCTCTTGTCCAAAATCAACTCGCTTGCCCACCTGCGCGCCGCGCCCTGCAACGACCTGCACGCCACCAAGCTGGCGCCCGGCAAGGAGA
AGGAGCCCCTGGAGTCGCAGTACCAGGTGGGCCCGCTACTGGGCAGCGGCGGCTTCGGCTCGGTCTACTCAGGCATCCGCGTCTCCGACA
ACTTGCCGGTGGCCATCAAACACGTGGAGAAGGACCGGATTTCCGACTGGGGAGAGCTGCCTAATGGCACTCGAGTGCCCATGGAAGTGGT
CCTGCTGAAGAAGGTGAGCTCGGGTTTCTCCGGCGTCATTAGGCTCCTGGACTGGTTCGAGAGGCCCGACAGTTTCGTCCTGATCCTGGAG
AGGCCCGAGCCGGTG**G**AAGATCTCTTCGACTTCATCACGGAAAGGGGAGCCCTGCAAGAGGAGCTGGCCCGCAGCTTCTTCTGGCAGGTG
CTGGAGGCCGTGCGGCACTGCCACAACTGCGGGGTGCTCCACCGCGACATCAAGGACGAAAACATCCTTATCGACCTCAATCGCGGCGAG
CTCAAGCTCATCGACTTCGGGTCGGGGGCGCTGCTCAAGGACACCGTCTACACGGACTTCGATGGGACCCGAGTGTATAGCCCTCCAGAGT
GGATCCGCTACCATCGCTACCATGGCAGGTCGGCGGCAGTCTGGTCCCTGGGGATCCTGCTGTATGATATGGTGTGTGGAGATATTCCTTTC
GAGCATGACGAAGAGATCATCAGGGGCCAGGTTTTCTTCAGGCAGAGGGTCTCTTCAGAATGTCAGCATCTCATTAGATGGTGCTTGGCCCT
GAGACCATCAGATAGGCCAACCTTCGAAGAAATCCAGAACCATCCATGGATGCAAGATGTTCTCCTGCCCCAGGAAACTGCTGAGATCCACC
TCCACAGCCTGTCGCCGGGGCCCAGCAAA

8.4 Western Blot Membranes