Mapiranje malih GTPaza HRAS, NRAS i MRAS korištenjem metode ligacije proteina posredovane pocijepanim inteinom

Mateševac, Josipa

University of Zagreb

Faculty of Science

Department of Biology

Josipa Mateševac

Mapping of small GTPase proteins HRAS, NRAS and MRAS using the Split Intein-Mediated Protein Ligation assay

Master thesis

Zagreb, 2020.

This thesis was performed in Stagljar Lab, Department of Biochemistry and Department of Molecular Genetics, Faculty of Medicine, University of Toronto, Toronto, Canada under supervision of Dr. Igor Štagljar, Prof. The thesis is submitted for evaluation to Division of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia in order to acquire the academic title of Master of Molecular Biology.

Acknowledgements

I wish to thank Dr. Igor Štagljar for giving me the amazing opportunity to make my master thesis and gain experience in a world renowned laboratory, Dr. Zhong Yao for his patience and guidance in training me, as well as all the Staglab members for answering my numerous questions and being a very welcoming group. A special thanks to Shivanthy Pathmanathan for all the conversations and support in times of panic and Nigel Henriques for making me laugh. It helped a lot.

A big thanks to Dr. Renata Matoničkin Kepčija for helping around matters regarding the scholarship and Dr. Petra Korać for being my supervisor at my home University.

I would like to thank my Canadian relatives for giving me a taste of home in foreign surroundings. I'll be sure to treasure all the laughs and good times.

To my parents, sisters, and friends, for their steady support throughout the years, I remain forever grateful.

TEMELJNA DOKUMENTACIJSKA KARTICA

Sveučilište u Zagrebu

Prirodoslovno-matematički fakultet

Biološki odsjek Diplomski rad

Mapiranje malih GTPaza HRAS, NRAS i MRAS korištenjem metode ligacije proteina posredovane pocijepanim inteinom

Josipa Mateševac

Rooseveltov trg 6, 10 000 Zagreb, Hrvatska

Nadporodica malih GTPaza Ras ima ulogu u prijenosu signala i modulaciji staničnih procesa. Njeni članovi djeluju kao binarni molekularni prekidači koji izmjenjuju inaktivni oblik vezan za GDP i aktivni oblik vezan za GTP koji veže svoje nizvodne efektore s velikim afinitetom. Porodica Ras uključena je u regulaciju unutarstanične signalizacije i utječe na procese kao što su stanični rast, preživljavanje, morfologija i diferencijacija. Proteini RAS post-translacijski se modificiraju adicijom lipida koja omogućava njihovo vezanje za membrane i biokemijsku aktivnost u raznim signalnim putovima. Nepravilnosti u normalnoj funkciji proteina obično rezultiraju malignim promjenama. Određene mutacije različitih proteina RAS često se pronalaze u različitim vrstama tumora, a također mogu voditi i do kongenitalnih razvojnih poremećaja. Budući da međuproteinske interakcije imaju ulogu u patofiziologiji bolesti, njihovo je razumijevanje ključno u otkrivanju potencijalnih meta terapija. Cilj ovog istraživanja bio je mapirati interakcije između tri proteina porodice Ras, HRAS, NRAS i MRAS, i proteina koji su s njima potencijalno u interakciji koristeći novorazvijen pristup, SIMPL (Split Intein-Mediated Protein Ligation), koji omogućava kvantifikaciju međuproteinskih interakcija metodom ELISA. Osam interaktora identificirano je za HRAS, trideset i četiri za NRAS te osamnaest za MRAS.

(56 stranica, 32 slike, 1 tablica, 49 literaturnih navoda, jezik izvornika: engleski)

Rad je pohranjen u Središnjoj biološkoj knjižnici

Ključne riječi: GTPaze, RAS, PPI, inteini, proteinski splicing, SIMPL

Voditelj: Dr.sc. Igor Štagljar, red. prof.

Suvoditeljica: Dr. sc. Petra Korać, izv. prof.

Ocjenitelji: Dr. sc. Petra Korać, izv. prof., Dr.sc. Jasna Lajtner, izv.prof., Dr.sc. Sofia Ana Blažević, doc.

Rad prihvaćen: 18.6.2020.

BASIC DOCUMENTATION CARD

University of Zagreb

Faculty of Science

Division of Biology

Master Thesis

Mapping of small GTPase proteins HRAS, NRAS and MRAS using the Split Intein-Mediated Protein Ligation assay

Josipa Mateševac

Rooseveltov trg 6, 10 000 Zagreb, Hrvatska

The Ras superfamily of small GTPases plays a role in signal transduction and modulation of cellular processes. Its members act as binary molecular switches that cycle between the inactive, GDP-bound form, and the active, GTP-bound form, in which they bind their downstream effectors with high affinity. The Ras family is involved in regulation of intracellular signaling and affects processes such as cell growth, survival, morphology and differentiation. They are post-translationally modified by lipid addition which enables their membrane recruitment and biochemical activity in various signaling pathways. Irregularities in the normal function of RAS proteins usually result in malignant changes. Certain mutations of various RAS proteins are commonly found in distinct cancer types and can also lead to different congenital developmental disorders. As protein-protein interactions play a role in disease pathophysiology, understanding of PPIs is crucial in discovering potential drug therapy targets. The aim of this study was to map interactions between three members of the Ras family, HRAS, NRAS and MRAS and their potential interactors using the newly developed SIMPL assay which allows PPI quantification using ELISA. Eight interactors were identified for HRAS, thirty-four for NRAS and eighteen for MRAS.

(56 pages, 32 figures, 1 table, 49 references, original in: English)

Thesis deposited in the Central Biological Library

Key words: GTPase, RAS, PPI, inteins, protein splicing, SIMPL

Supervisor: Dr. Igor Štagljar, Prof.

Co-supervisor: Dr. Petra Korać, Assoc. Prof.

Reviewers: Dr. Petra Korać, Assoc. Prof., Dr. Jasna Lajtner, Assoc. Prof., Dr. Sofia Ana Blažević, Asst. Prof.

Thesis accepted: 18.6.2020.

Table of Contents

1. Introduction

1.1. Ras superfamily of small GTPase proteins

Comprised of over 150 human members, the Ras superfamily of small guanosine trisphosphatases (GTPases) plays a crucial role in signal transduction and modulation of various cellular processes. The superfamily members act as binary molecular switches that bind and hydrolise guanosine tryphosphate (GTP) (Wennerberg, Rossman, and Der 2005). The basic biochemical activity is reflected in a shared set of conserved ″G box″ GDP/GTP binding motifs: G1 – (aaaaGxxxxGK(S/T), a = L/I/V/M, x = any amino acid) involved in binding the purine nucleotide, G2 located in one of the segments that reorient upon nucleotide binding, G3 – (blbbDxxGl, l = hydrophilic, b = hydrophobic) involved in binding a Mg^{2+} ion, G4 – (bbbb(N/T)(K/Q)xD) that confers specificity to GTP by making a hydrogen bond with the guanine ring and interacts with G1 box residues and G5 – (bbE(A/C/S/T)SA(K/L)) that is less well conserved in the superfamily and makes indirect associations with the guanine nucleotide (Colicelli 2004).

The GTPases cycle between GTP and GDP bound forms (Figure 1) which have similar conformations (Wennerberg, Rossman, and Der 2005). Structural changes associated with distinct nucleotide binding correspond to two dynamic loop regions called switch 1 and switch 2 (Colicelli 2004; Karnoub and Weinberg 2014). When bound to GTP, they display a high affinity for their downstream effector targets. This conformation is transient, and downstream signaling is attentuated upon GTP hydrolysis (Brunsveld, Waldmann, and Huster 2009; Colicelli 2004). Since they exibit low intrinsic GTP hydrolysis and exchange activities, they are regulated by two classes of proteins. Guanine-nucleotide-exchange factors (GEFs) promote GTP loading and protein activation while GTPase-activating proteins (GAPs) enhance the intrinsic GTPase activity favouring protein inactivation (Wennerberg, Rossman, and Der 2005). A particular RAS protein may have several GEFs depending on the upstream stimuli. In addition to GEFs and GAPs, guanine nucleotide dissociation inhibitors (GDIs) may bind to GDP-bound GTPases to prolong the inactive state by inhibiting GDP release (Colicelli 2004).

Ras superfamily GTPases can be divided into five major (sub)families based on protein sequence and functional similarities: Ras, Rho, Arf, Rab, and Ran (Wennerberg, Rossman, and Der 2005). Ras (Rat sarcoma) family members, which the superfamily was

named after, regulate intracellular signaling and processes such as cell growth, survival, morphology and differentiation; Rho (Ras homologous) family is involed in cytoskeletal remodeling (lamellipodia and stress fiber formation) and cell polarity; Arf (ADP ribosylation factor) and Rab (Ras-like in brain) family proteins are crucial for vesicle trafficking, with Rab proteins regulating the processes from vesicle budding to cargo release. Ran (Ras-like nuclear) family comprises of only one protein, the most abundant cellular small GTPase involved in nuclear transport of proteins and RNA (Colicelli 2004; Rojas et al. 2012; Wennerberg, Rossman, and Der 2005).

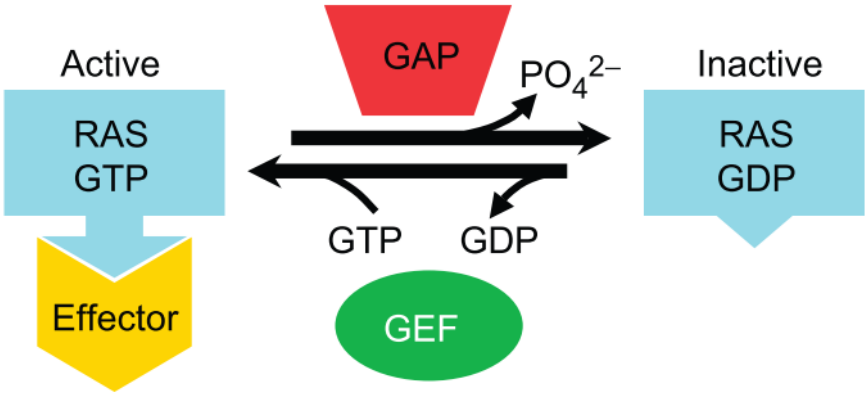


Figure 1. The balance between GTP-bound ″on″/active and GDP-bound ″off″/inactive form. GEFs promote GTPase activation and interaction with effector proteins, while GAPs promote GTPase inactivation by GTP hydrolysis (adapted from Colicelli 2004).

1.2. Post-translational modifications of Ras family proteins

Most Ras family proteins are relatively small (20-29 kDa), and their proper function is determined by adequate protein processing (Colicelli 2004; Karnoub and Weinberg 2014). They are post-translationally modified by lipid addition which is a requirement for membrane recruitment and biochemical activity (Brunsveld, Waldmann, and Huster 2009). Most Ras family proteins terminate with a 'Caax' (a = aliphatic, x = any amino acid) motif in the hypervariable C-terminal domain (Khan et al. 2019) that is recognized by farnesyltransferase and geranylgeranyltransferase I which catalyze the isoprenoid (farnesyl or geranylgeranyl) lipid addition. The motif and the upstream residues make up the membrane-targeting sequences that determine subcellular localization by targeting specific membrane

compartments (Wennerberg, Rossman, and Der 2005). The 'aax' sequence is further proteolytically removed by RAS-converting enzyme 1 (RCE1) and the lipid-modified cysteine residue is methylated by isoprenylcystein-carboxyl-methyltransferase (ICMT) (Khan et al. 2019). Some Ras proteins require an additional palmitoylation step on the upstream cysteine residues to stabilize their membrane anchoring (Karnoub and Weinberg 2014), while some are not modified by lipids (Colicelli 2004; Wennerberg, Rossman, and Der 2005).

1.3. Ras signaling pathways

Ras family proteins regulate intracellular signaling by functioning as signaling nodes and can be activated by various extracellular stimuli (Rojas et al. 2012). The growth factor or mitogen stimulated receptor tyrosine kinase activates a RasGEF, which in turn activates RAS by GTP loading. Once activated, RAS binds to downstream effectors in diverse signaling pathways (Johnson and Chen 2012; Khan et al. 2019).

The RAF/MEK/ERK signaling cascade is the most studied RAS effector pathway and is associated with proliferation, migration and differentiation (Khan et al. 2019). Serine/threonine kinase RAF activates the dual-specificity protein kinase MEK1/2, which, in turn, activates the mitogen-activated protein kinase ERK1/2. Activation is carried out by phosphorylation and results in target gene promotor activation by transcription factors (Wennerberg, Rossman, and Der 2005).

Activation of the PI3K pathway leads to increase in membrane-associated phosphatidylinositol 3,4,5-triphosphate (PIP_3) and recruitment of proteins with a PH domain, such as AKT1 and PDK1 (Colicelli 2004). PDK1 is involved in protein kinase activation, while AKT promotes cell survival by regulating survival and apoptotic genes (Khan et al. 2019).

RAS proteins affect RalGEFs which activate Ral signaling pathway, involved in processes vital for cell homeostasis and survival; they activate phospholipase C involved in calcium signaling and NORE1 implicated in growth inhibition and apoptosis (Khan et al. 2019).

1.4. Ras proteins and their role in disease development

Due to their role in signal transduction from the cell suroundings, such as growth factors and mitogens, which results in transcription factor activation and ultimately a cellular response (differentiation, proliferation, cell growth, survival) (Maffeis, Nicolè, and Cappellesso 2019), irregularities in the normal function of RAS proteins usually result in malignant changes. Years of research have shown that RAS proteins are involved in tumorigenesis, as well as developmental disorders (Karnoub and Weinberg 2014).

There are several RAS isoforms that act as cellular proto-oncogenes and have been extensively studied due to their role in human cancer development (Karnoub and Weinberg 2014). Proteins KRAS, HRAS and NRAS are known for their mitogenic properties, and their mutationally activated forms which are able to transform cells (*in vivo* and *in vitro*) (Colicelli 2004) arise mainly due to missense gain-of-function mutations in hotspot codons (codons 12, 13 and 61) which leads to constitutive RAS activation (Khan et al. 2019; Prior, Lewis, and Mattos 2012). The isoforms have a nearly 85% sequence identity (Engin et al. 2017) and they share activator and effector proteins, however certain oncogenic isoforms have been linked to distinct cancer types (Nussinov, Tsai, and Jang 2018). *KRAS* mutations are commonly found in pancreatic, lung and colon cancer, *HRAS* in dermatological and head and neck carcinomas, while *NRAS* are most frequent in melanomas and hematological malignancies (Alcantara et al. 2019; Khan et al. 2019). While oncogenic mutations predominantly affect *KRAS* and are detected in 25-30% of all tumors, mutations in *NRAS* and *HRAS* occur with lower frequency (8% and 3%, respectively). Such a discrepancy can be explained by a higher physiological relevance of KRAS, sugested by knock-out mice phenotypes, as it is essential for normal mouse development, while NRAS and HRAS are dispensable (Fernández-Medarde and Santos 2011).

The reason for isoform differential signaling can be attributed to characteristic combinations of post-translational modifications in the hypervariable C-terminal domain. RAS isoforms prefer plasma membrane microdomains of distinct organization and composition. This can affect RAS orientation and isoform accessibility, and consequently the association with effector proteins, however the underlying cause of different isoform prevalence in certain cancers still remains unknown (Nussinov, Tsai, and Jang 2018).

The three proteins are part of the RAS oncoprotein branch of the Ras family. A member of the RRAS branch, MRAS (RRAS3) has been less extensively studied. Its expression has been reported in the central nervous system and muscle cells, and in its constitutively active form, it promotes cytoskelet reorganization, cell transformation, survival and differentiation, however the physiological functions are largely unknown (Nunez Rodriguez et al. 2006).

Expansion of cancer cells is mainly caused by uncontrolled cell proliferation. As RAS has a central role on deciding the fate of mitogenic stimuli, constitutive activation enables the proliferative potential of cancer cells. Oncogenic RAS was shown to be involved in bypassing death signals, upregulation of hypoxia-inducible factor 1α (HIF1α), and glucose transporter GLUT1 involved in glycolytic pathways. It is linked to angiogenesis and cancer cell evasion of the immune system, as well as cancer cell invasion and migration (Khan et al. 2019).

Aberrant RAS regulation due to mutations of RAS proteins or their regulators can also lead to congenital developmental disorders (Karnoub and Weinberg 2014). RASopathies are predominantly caused by mutations of the RAS/MAPK signaling pathway genes. They are a set of syndromes that are inherited in an autosomal dominant manner. The syndromes share some phenotypic features, such as a higher risk of cancer, craniofacial dysmorphisms, congenital heart disease and short stature (Chinton et al. 2019).

Neurofibromatosis type-1 is caused by a mutation in the *NF1* gene which encodes a RasGAP neurofibromin-1. It is a familial cancer syndrome characterized by build-up of pigmented lesions in the skin and eye and the tendency to develop sporadic malignant outgrowths (Karnoub and Weinberg 2014). A milder form of neurofibromatosis type-1 is Legius syndrome where loss-of-function mutations occur in *SPRED1*, a supressor of RAS/MAPK signaling essential for interaction of NF1 and RAS in the plasma membrane (Shimanshu, Nissley, and McCormick 2017).

Noonan Syndrome and Noonan Syndrome with Multiple Lentigines (previously Leopard) are mainly caused by mutations of the *PTPN11* gene (50% and 90%, respectively) (Chinton et al. 2019) that encodes a receptor tyrosine kinase phosphatase. Its product, SHP2, acts as a mediator in conveying upstream signals to downstream targets, mostly the RAS/MAPK pathway. Apart from *PTPN11,* gene mutation connected to Noonan syndrome include *SOS1*, *KRAS, NRAS, BRAF*, *CRAF* etc. Mutations in RasGEF *SOS1* and RAS protein genes lead to higher RAS-GTP levels and stronger pathway activation (Fernández-Medarde

and Santos 2011). Costello Syndrome is predominatnly caused by *HRAS* gene mutations, leading to higher pathway activation and increase in proliferation, while *BRAF* mutations are the most common in Cardiofaciocutaneous Syndrome (Fernández-Medarde and Santos 2011).

Due to RAS importance in disease development, continuous effort has been invested in finding effective drug therapies. As association with the plasma membrane is crucial for oncogenic activity, RAS membrane localization was initially targeted by using farnesyltranferase inhibitors (FTIs), and later other enzymes that process RAS. Different activation stages of RAS, RAS effector binding and nucleotide exchange have also been targeted. While some approaches have been proved effective in model cell systems, direct RAS inhibitors for clinical use have yet to be developed (O'Bryan 2019).

Studying protein-protein interactions (PPIs) of RAS proteins, and other proteins in general is vital in understanding disease pathophysiology (Petschnigg, Snider, and Stagljar 2011). Discovering interactions between RAS and its effectors is therefore crucial for determining new potential targets for theurapeutic purposes (Gysin et al. 2011).

1.5. Protein-protein interactions and methods for their analysis

Protein interactions are essential in maintaining cellular functions. Disruptions in the formation or crosstalk of protein complexes can affect fundamental processes, such as cell cycle control, gene transription and signal transduction, often leading to cellular dysfunction and ultimately disease (Petschnigg, Snider, and Stagljar 2011). Protein-protein interactions are essential for elucidating protein function in the context of cellular physiology and have become attractive as drug targets (Machleidt et al. 2015).

PPIs can be discovered using biochemical methods, where complex composition is determined by directly working with proteins, or using genetic methods based on reporter reconstruction upon protein interaction (Petschnigg, Snider, and Stagljar 2011).

Y2H (Yeast two-hybrid) is a genetic approach that utilizes bait and prey proteins fused to the DNA binding (BD) and activation domain (AD) of a transription factor, respectively. Physical interaction leads to transcription factor reconstitution and reporter gene activation which can be measured (Suter, Kittanakom, and Stagljar 2008). Membrane yeast two-hybrid (MYTH) allowed the use of full length membrane proteins in their natural environment. It is a

Y2H variant based on split-ubiquitin where the membrane bait protein is fused to the ubiquitin C-terminal fragment (Cub) and an artificial transcription factor at its cytosolic terminus while the prey is fused to the N-terminal fragment. Protein interaction leads to pseudo-ubiquitin reconstitution, recognition and cleavage of the transcription factor by deubiquitinating enzymes and reporter gene expression (Petschnigg, Snider, and Stagljar 2011). Mammalian-membrane two-hybrid (MaMTH) was further derived from MYTH, allowing protein interaction testing in virtually any mammalian cell line (Petschnigg et al. 2014). Y2H and MYTH became popular for binary interactions detection due to their simplicity, low cost, and the ability to test interactions *in vivo*. As the tested proteins are not native to yeast cells, disadvantages include problems associated with the protein expression, modification and interaction (Snider et al. 2015).

Resonance-energy transfer techniques use fluorescent (FRET) or bioluminescent (BRET) proteins fused to interacting proteins and detect non-radiative energy transfer between an excited donor and an acceptor protein (Petschnigg, Snider, and Stagljar 2011). Both methods allow instantaneous, real-time interaction detection in live cells. They also require fusion protein generation and close spatial proximity between donor and acceptor. BRET has lower background and greater sensitivity than FRET, but also tends to give a weaker signal (Snider et al. 2015).

Co-immunoprecipitation is a biochemical method that enables isolation of proteins bound in a complex. Upon cell lysate generation, an antigen is precipitated using a specific antibody. After washing, the bound proteins are eluted and analyzed (Phizicky and Fields 1995). LUMIER (Luminescence-based mammalian interactome mapping) is based on co-immunoprecipitation. It features one protein fused to Renilla luciferase and another to an affinity tag. The tagged proteins are co-expressed, the cells are lysed and the proteins are immunoprecipitated using an antibody against the affinity tag. The measured luciferase activity is used to assess protein interaction. LUMIER can be used in different cell lines and in high-throughout screens, but is not appropriate for measuring weak and transient PPIs, as they can be disrupted during cell lysis and can introduce potential artefacts (Snider et al. 2015).

Computational methods can be used for the prediction of PPIs. FpClass is an *in silico* method used to predict proteome-wide high confidence PPIs, albeit with a 60% percent false discovery rate. Prediction for low-degree proteins (with few known partners) and protein

orphans (with no known partners) remains challenging due to their poor annotation (Kotlyar et al. 2014).

All PPI detection methods have their advantages and limitations. The choice of method depends on the class of proteins, and the scale, time and cost of the study. Some PPIs are stable while others are transient, and can only be detected by a small number of methods. The cellular location the interaction takes place must also be taken into consideration (Snider et al. 2015).

A novel PPI detection method, Split Intein-Mediated Protein Ligation (SIMPL) was recently developed in Igor Stagljar's laboratory. This method utilizes a split intein as a protein interaction sensor. The method's advantages are high sensitivity and specificity and it can be used for PPI detection in various cellular compartments, as well as tracking of kinetic interactions (Yao et al. 2020).

1.5.1. Inteins and Split Intein-Mediated Protein Ligation

An intein (internal protein) is an intervening protein sequence, and through the process called protein splicing, is able to autocatalitically excise itself from the parental peptide while simultaniously ligating the flanking protein sequences called exteins (external proteins). Upon petide bond formation, the parental protein proceeds with folding and assumes its normal function (Aranko, Wlodawer, and Iwaï 2014; Gogarten et al. 2002). Most inteins consist of a self-splicing and an endonuclease domain, deletion of which has no effect on protein splicing. Considered to be parasitic genetic elements, as some inteins can spread through homing events that result in intein duplication (Gogarten et al. 2002), inteins are found in proteins involved in DNA metabolism (replication, transcription and maintenance) as well as other housekeeping genes of unicellular organisms of all three domains of life (Shah and Muir 2014).

Intein-mediated protein splicing occurs spontaneously, the only requisite being the folding of the intein domain, and has a variety of applications in biotechnology, including tagless protein purification, *in vitro* and *in vivo* protein semi-synthesis, peptide and protein cyclization etc. (reviewed in: Shah and Muir 2014; Wood and Camarero 2014).

Split inteins occur naturally, or can be split artificially, and allow protein trans-splicing (Wood and Camarero 2014) (Figure 2). SIMPL utilizes a re-engineered version of GP41-1, a naturally occurring split intein, with minimal fragment association. In this method, a bait protein is C-terminally fused to a V5 tag and an 88 amino acids long intein N-terminal fragment (IN), while a prey protein is N-terminally fused to a FLAG tag and a 37 amino acids long intein C-terminal fragment (IC) (Figure 3). IN and IC are brought into close proximity upon bait and prey association, resulting in the reconstitution of a fully functional intein, its excision and ligation of tagged bait and prey peptides into an intact protein that can be detected, visualized and purified (Yao et al. 2020).

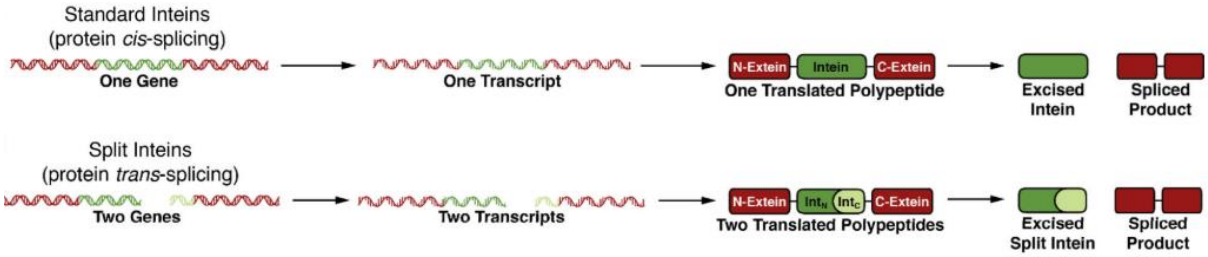


Figure 2. Protein *cis*-splicing by contiguous inteins and protein *trans*-splicing by split inteins (adapted from Shah and Muir 2014).

As the presence of termini-specific tags may impair protein function, or they may be reciprocally inaccessible due to their spacial conformation, two alternative arrangements have been designed (Figure 4). One features a prey with a C-terminally fused IC moiety and a downstream FLAG tag (CIC design), while the other features a bait with an N-terminally fused IN moiety and an upstram V5 tag (NIN design). In the first case, interaction with bait-IN results in the formation of a bait-V5-FLAG peptide, while in the other, interaction with IC-prey results in a V5-FLAG-prey peptide (Yao et al. 2020).

All three formats are complatible with ELISA which allows high-throughput analyses and protein-protein interaction quantification. To utilize the ELISA format, hemaglutinin and c-myc tags were introduced into bait and prey constructs, respectively. This allows protein capture using α-V5 and α-FLAG antibody and detection using α-HA and α-c-myc coupled horseradish peroxidase (HRP). SIMPL signal can similarly be measured by bait tag capture and prey tag detection and the measured interaction can be normalized to bait and/or prey expression (Yao et al. 2020).

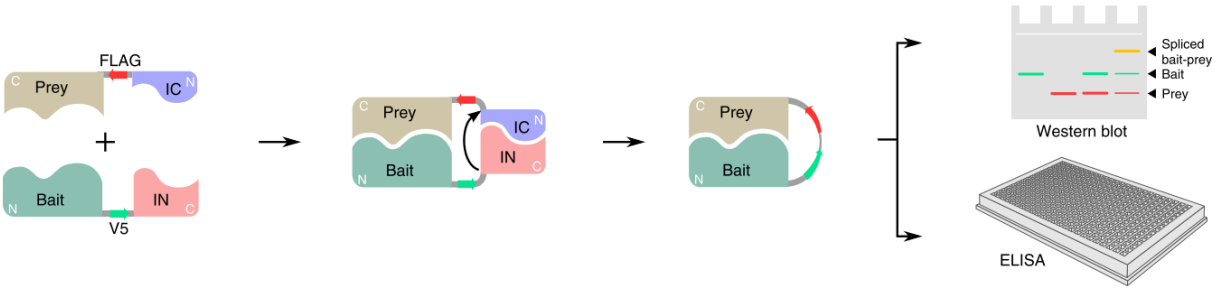


Figure 3. Schematic representation of SIMPL. Interaction of bait and prey results in intein reconstitution and protein splicing. The intein is excised and bait and prey proteins (as well as the tags) are ligated. The altered mobility allows detection by Western blotting, while tags allow visualization by biochemical techniques (e.g. ELISA) (adapted from Yao et al. 2020).

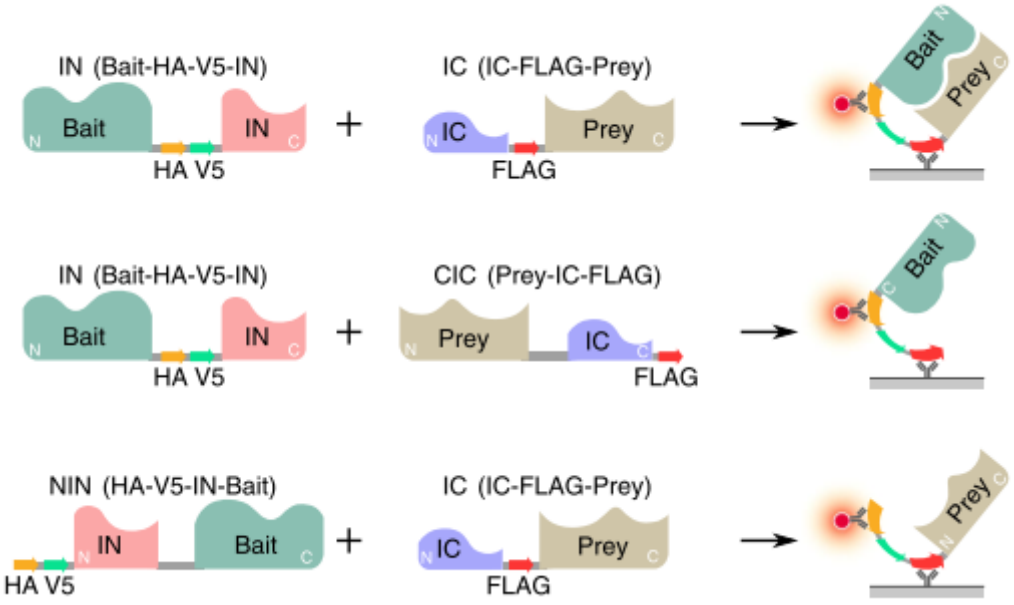


Figure 4. SIMPL formats. The IC moiety is fused C-terminally to prey with a downstream FLAG in the CIC design, while the IN moiety is fused N-terminally to bait with an upstream V5 in the NIN design. Interaction between bait and prey results in splicing where tags remain fused to bait and prey in CIC and NIN format, respectively (adapted from Yao et al. 2020).

1.6. Research aim

The aim of this research was to map interactions of three members of the Ras family of small GTPase proteins, HRAS, NRAS and MRAS, and their potential interactors using the newly developed Split Intein-Mediated Protein Ligation (SIMPL) assay. The newly discovered interactions may help in understanding the complex cellular processes, and interactions that play a part in disease development may serve as a target of future therapeutics.

2. Materials and methods

2.1. Cell culture

T-REx HEK 293 cells which stably express tetracycline repressor protein were used for the transient protein expression experiments. Cells were grown at 37 °C, 5% CO_2 in a cell culture incubator (Hera cell 150i, Thermo Fisher Scientific) in tissue culture plates containing Dulbecco's Modified Eagle Medium (DMEM) (Multicell, Wiesent Bio Products) with 4.5 g/L D-glucose, L-glutamine and sodium pyruvate, supplemented with 10% (v/v) fetal bovine serum (FBS) (Multicell, Wiesent Bio Products) and 1% (v/v) penicillin/streptomycin (10 000 µg/ml, Gibco). The cells were passaged after reaching high confluency (90-100%).

2.2. Cell passaging

The medium was aspirated and the cells were washed with 5-10 ml of room temperature Dulbecco's Phosphate Buffered Saline (1x PBS) (Multicell, Wisent Bio Products) to remove dead cells and leftover medium. PBS was then aspirated and the cells were covered with 1 ml of 0.25% trypsin-EDTA (1x) (Gibco). The plate was placed in the incubator for 5 minutes, or until the cells were detached from the plate bottom. To inactivate trypsin, 1 ml of DMEM was added, the cells were resuspended and 200 µl were added to a new culture plate containing 10 ml of DMEM. When seeding cells for experiment purposes, 5 µl of the cell suspension were diluted in 495 µl of the medium and counted with Scepter™ 2.0 Handheld Automated Cell Counter (Millipore).

2.3. cDNA library preparation

2.3.1. Heat-shock transformation of bacterial cells

Chemically competent *E.coli* DH5α cells (prepared by Inoue method) stored at -80 °C were thawed on ice and 20-30 µl were placed in sterile 8-strip PCR tubes. The Gateway cloning reaction mixure (2 µl) was added to the cells. The cells and DNA were mixed by flicking the tubes several times and then placed on ice for 20-30 minutes. The tubes were

heated at 42 °C for 45-50 seconds in a water bath (VWR), then placed on ice for 2 minutes to allow the bacteria to recover. To grow the bacteria, the cells were plated on room temperature 10-cm plates containing the LB agar medium with the appropriate antibiotic and placed in the bacteria incubator (VWR) at 37 °C overnight.

2.3.2. Colony picking and plasmid purification

Individual colonies were picked and inoculated in 3 ml of LB medium + 3 µl antibiotic in 15 ml cell culture Falcon tubes. The tubes were placed in a bacteria incubator shaker (Minitron, Infors HT) at 37 °C and 200 rpm for 16-20 hours. The plasmid DNA was then isolated using the commercial Presto™ Mini Plasmid kit (Geneaid) as per manufacturer's instructions.

2.3.3. Gateway cloning

2.3.3.1. Isolation of Entry clones

The FpClass tool (Jurisica lab, http://dcv.uhnres.utoronto.ca/FPCLASS/) was used to predict potential protein interactors of the selected bait proteins. Prey cDNA in pDONR223 plasmid was obtained from the Human ORFeome collection v8.1 (Stagljar Lab). NRAS and MRAS bait cDNA was obtained from Dr. Zhong Yao (Stagljar Lab) and HRAS from Dr. Anna Lyakisheva (Stagljar Lab). Bacterial cells containing the DNA of interest were streaked from frozen glycerol stocks onto agar plates with the appropriate antibiotic, then grown in a bacteria incubator at 37 °C overnight. Colonies were picked for liquid bacterial culture and the DNA was isolated as described above. As the only DNA not originally in pDONR223, HRAS cDNA was cloned into pDONR223 plasmid by Gateway BP cloning: 3 µl of 1xTE (Tris HCl, EDTA) Buffer (Tekanova), 0.5 µl of Gateway BP Clonase Enzyme II mix (Invitrogen), 0.5 µl of pDONR223 donor vector (150 ng/µl) and 0.5 µl of cDNA (100 ng/µl) were mixed and incubated at room temperature for 2-4 hours. The DNA was amplified and purified as previously described.

2.3.3.2. Generation of Expression clones

Given the fact that Ras proteins are C-terminally processed and any tagging on that termini is likely to result in the tag being cleaved off, bait proteins were N-terminally fused to IC (with FLAG and c-myc tags) and prey proteins were N- and C-terminally fused to IN (with V5 and HA tags; NIN and IN formats). Gateway LR cloning was used to clone the desired cDNA into SIMPL destination vectors of interest, designated IC2 (Figure 5), IN4b (Figure 6) and NIN4 (Figure 7). The reaction mixture containing 3 µl of 1xTE Buffer (Tekanova), 0.5 µl of LR Clonase Enzyme II mix (Invitrogen), 0.5 µl of destination vector (150 ng/µl) and 0.5 µl of entry clone cDNA (100 ng/µl) was prepared. The expression clones were amplified and purified as described above.

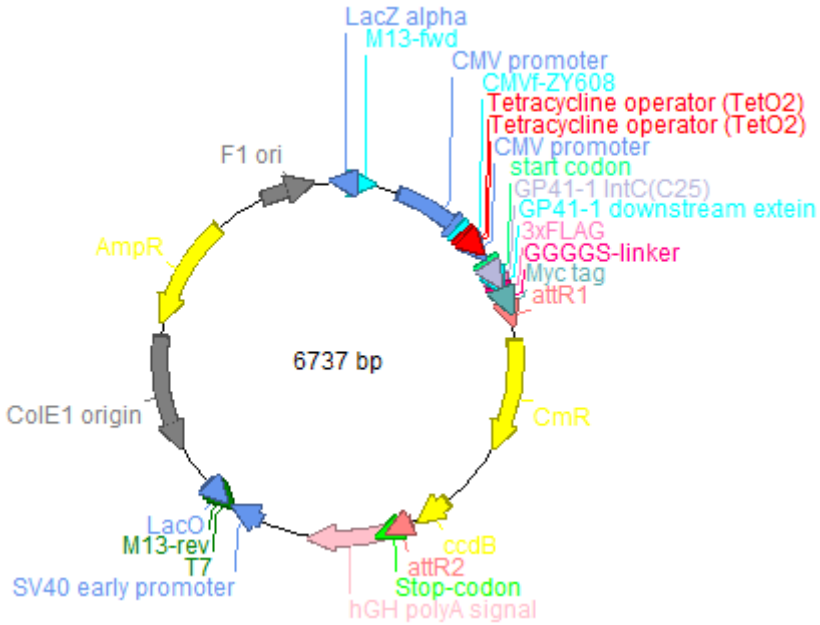


Figure 5. N-tagged SIMPL bait vector (IC2)

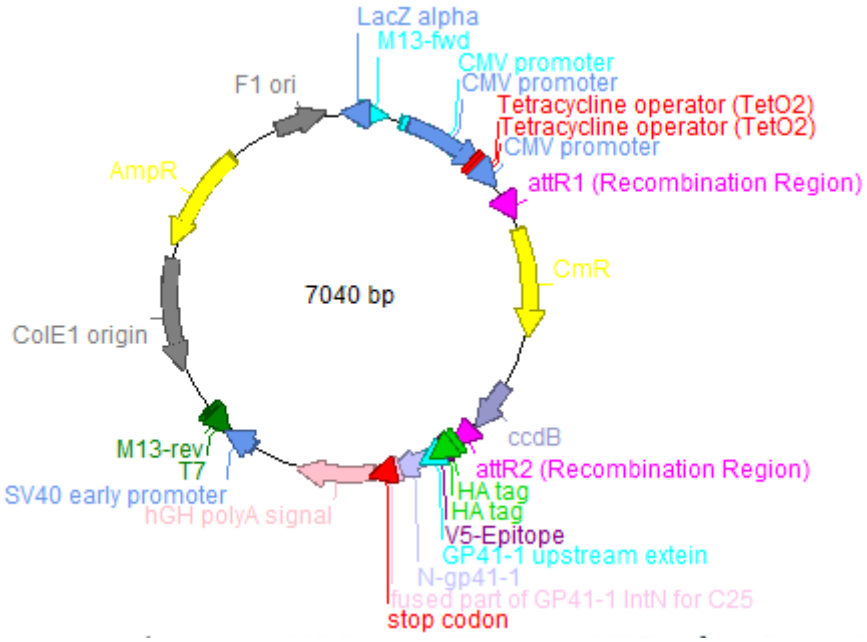


Figure 6. C-tagged SIMPL prey vector (IN4b)

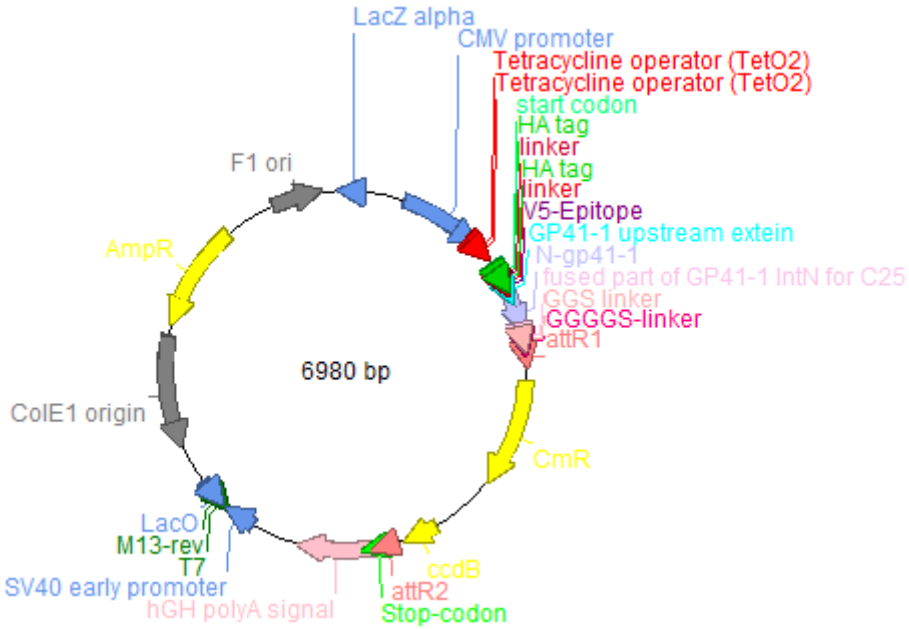


Figure 7. N-tagged SIMPL prey vector (NIN4)

2.3.4. DNA concentration measurement

DNA concentration and purity (A260/A280 ratio) were measured by spectrophotometry on CLARIOstar Microplate Reader (BGM Labtech).

2.3.5. Sequencing

All the bait proteins and around 12% of the total number of prey proteins at random were sent for sequencing at The Centre for Applied Genomics, Sick Kids Hospital, Toronto, using the CMV-forward primer and the sequencing results were analysed using the ApE programme (https://jorgensen.biology.utah.edu/wayned/ape/) and NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.4. ELISA

Day 1: Cell seeding and tranfection

HEK293 cells were seeded from a 10-cm plate to 96-well plates; 100 µl with 15 000 cells/well in the morning. Bait-prey DNA pair was transfected in the late afternoon. For each sample, 0.2 µg DNA was added to 10 µl of PBS and 0.6 µl of PEI (Sigma-Aldrich) was added to another 10 µl of PBS. The mixtures were combined and left at room temperature for 15-30 minutes, then 5 µl were added to each well.

Day 2: Induction of protein synthesis, ELISA plate preparation

To each well, 50 µl of media with 1.5 µg/ml tetracycline (for a final concentration of 0.5 µg/ml) were added 24 hours after transfection.

a-FLAG (Sigma) and a-V5 (BioRad) antibodies were diluted 1:100 in PBS. 20 µl were added to each well of the 384-well plates. The plates were incubated at 37 °C for 1 hour and shaken occasionally.

80 µl of blocking buffer (2% BSA in PBST; PBST: 0.05% Tween20 in PBS) were added to each well and the plates were incubated on a shaker at 4 °C overnight.

Day 3: Cell lysis and ELISA reading

The media were aspirated and the cells were lysed with 100 µl of lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 2 mM EDTA, 0.5 % Triton X-100, protease inhibitors from stock solutions: benzamidine (1000x) 1 M in ddH_2O, aprotinin (1000x) 10 mg/ml in ddH_2O, leupeptin (1000x) 10 mg/ml in ddH_2O, pepstain A (500x) 1 mg/ml in EtOH), sealed and sonicated for 1 minute at 4 °C in a bath sonicator (Branson 8510). 20 µl of cell lysate were added to each antibody coated well. The plates were rocked at room temperature for 2-3 hours and the wells washed 3 times with 60 µl of PBST. Detection antibodies were diluted in blocking buffer 1:2 000 for a-HA (GeneTex) and 1:5 000 for a-c-myc (Santa Cruz Biotechnology). Antibodies (30 µl) were added to the wells as follows: a-c-myc for detection of bait protein captured by a-FLAG, a-HA for detection of prey proteins interacting with a-FLAG captured bait proteins and a-HA for detection of prey proteins captured by a-V5 antibody. The plates were further incubated on a rocker at room temperature for 1 hour. After washing 3 times with 60 µl of PBST, 30 µl/well of substrate buffer prepared by mixing the two SuperSignal™ ELISA Pico reagents (Thermo Scientific) and ddH_2O 1:1:8 were added. Luminescence was read on CLARIOstar Microplate Reader (BGM Labtech) for integration time of 0.2 seconds/well.

2.5. Data analysis

The luminescence signal for all the readings (bait expression, prey expression, SIMPL signal) was obtained after substracting the corresponding values of 'mock' sample with no transfected DNA. The average and standard deviation was calculated for each sample. To normalize the data to prey expression, SIMPL signal for each sample was divided by the corresponding prey luminescence value.

2.6. Western blot

2.6.1. Transfection and sample preparation

The cells were seeded from a highly confluent 10-cm plate into 24-well plates in a 1:100 ratio with 0.5 ml media per well. For transfection, 12.5 µl of PBS with 0.25 µg DNA and 12.5 µl PBS with 0.75 µl PEI were mixed and left to sit at room temperature for 15-30 minutes prior to adding to the cells. The following day, tetracycline was added in final concentration of 1 µg/ml overnight. Keeping the plate on ice, the media were aspirated and 70 µl of H lysis buffer (50 mM β-glycerophosphate, 1.5 mM EGTA, 1 mM EDTA, 0.1 mM ortho-vanadate, 1 mM DDT and 1% Triton-100 supplemented with protease inhibitors) were added to each well. The samples were incubated for 10 minutes at 4 °C on the shaker and the lysates were collected in PCR tubes on ice. They were further centrifuged for 5 minutes at 4 °C at 4 000 rpm (5810R, Eppendorf). 54 µl of each supernatant were added to 18 µl of 4x Sample buffer (40% glycerol, 200 mM Tris/HCl pH 6.8, 8% SDS, 0.04% bromphenol blue, 0.1 M dithiothreitol) and heated at 95 °C for 3 minutes.

2.6.2. Gel electrophoresis and Western blotting

The sample proteins were separated by gel electrophoresis on 12% resolving gels. Per 8 ml, 2.7 ml of ddH_2O, 3.2 ml of 30% acrylamide/bisacrylamide solution, 2 ml of 4x Lower buffer (1.5 M Tris, pH 8.8, 0.4% SDS), 80 µl of ammonium peroxydisulfate (APS) and 5 µl of Tetramethylethylenediamine (TEMED) were added. Per 5 ml of 5% stacking gel, 2.8 ml of ddH_2O, 0.83 ml of 30% acrylamide/bisacrylamide solution, 1.25 ml of 4x Upper buffer (0.5 M Tris, pH 6.8, 0.4% SDS), 50 µl of APS and 5 µl of TEMED were added. 15 µl of each sample and 1 µl of PageRuler^{TM} Plus Prestained Protein Ladder (Thermo Scientific) were loaded onto the gel and run in 1x Running buffer (diluted from 10x Running buffer: 0.25 M Tris, 1.92 M Glycine, 1% SDS and ddH_2O to the final volume of 1 L) for 65-75 minutes in Mini-PROTEAN Tetra Cell (BioRad) apparatus under the constant voltage of 150 V.

The proteins were transfered onto a nitrocellulose membrane by wet transfer in 1x Transfer buffer (diluted from 10x Transfer buffer: 0.25 M Tris, 1.2 M Glycine and ddH_2O for the final volume of 1 L). The transfer was done using the Criterion^{TM} Blotter (BioRad)

apparatus for 90 minutes under the constant amperage of 300 mA. Following the transfer, the membraines were stained with Ponceau S (Sigma Aldrich) and washed with 1% acetic acid (BioShop) to verify transfer quality. They were further washed with water and 1x Tris buffered saline/Tween-20 (TBST) buffer (diluted from 10x TBST buffer: 0.2 M Tris pH 7.5, 1.5 M NaCl, 0.5% Tween-20 and ddH_2O for the final volume of 1 L) until all the dye had been washed out. They were then blocked for 2 hours with ~10 ml of 2% Bovine Serum Albumin (BSA, BioShop) in TBST. The membranes were incubated with 10 ml of the primary antibody diluted 1: 10 000 in TBST overnight at 4 °C on the rocking platform. They were washed three times for 15 minutes in TBST, then incubated in 10 ml of secondary antibody diluted 1:10 000 in TBST for 2 hours at room temperature on the rocking platform. After three more rounds of washing, the membranes were incubated with home-made enchanced chemiluminescence (ECL) substrate (10 ml of 0.1 M Tris pH 8.5, 50 µl of Luminol, 22.2 µl of p-Coumaric acid, 5.4 µl of 30% H_2O_2) for 1 minute. The film was developed in the dark room with different exposure times.

3. Results

3.1. Transient bait protein expression

To test whether bait proteins express transiently, bait cDNA was transfected into T-REx HEK 293 cells. Protein expression was induced by tetracycline addition (1 µg/ml) and their total levels were measured using anti-FLAG antibodies. Anti-α-tubulin antibodies were used to measure the loading control.

Bait protein expression was detected both in the presence and absence of Tet (Figure 8), with Tet induced samples having a stronger signal.

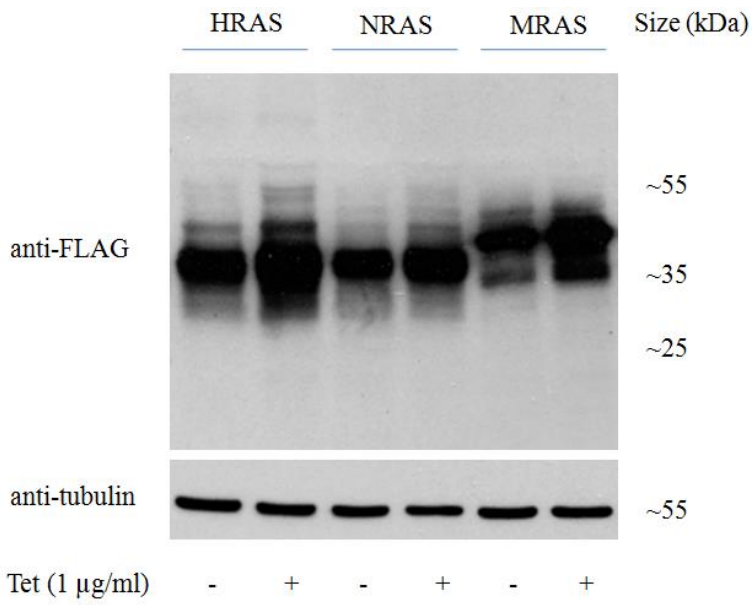


Figure 8. Transient expression of HRAS, NRAS and MRAS. Bait proteins were detected with and without Tet, α-tubulin was used as loading control. Detected protein sizes: HRAS ~ 38 kDA, NRAS ~ 38 kDA, MRAS ~ 40 kDa.

3.2. SIMPL assay for system validation using positive and negative protein interactors

To validate the SIMPL system, bait proteins were tested with positive and negative protein interactors. TRE-x HEK 293 cells were co-transfected in biological duplicates with

bait and prey protein cDNA. RAF-RBD-IN4b (Ras-binding domain) found in RAF proteins (Terrell and Morrison 2019) was used as a positive control, LSM3-IN4b as negative control and 'mock' as a control sample without any transfected DNA. LSM2/LSM3 bait/prey pair was additionally used as a positive control. Protein expression was induced using Tet (0.5 µg/ml).

Bait and prey protein expression, as well as SIMPL signal were measured using ELISA. Bait proteins were captured by α-FLAG antibodies, and prey proteins were captured using α-V5 antibodies. Bait proteins were detected by α-c-myc detection antibodies, prey proteins and SIMPL signal were detected by α-HA detection antibodies. As detection antibodies are conjugated with horse-radish peroxidase (HRP), luminescence was measured upon substrate addition. The average and standard deviation were calculated for each sample.

Bait, prey and SIMPL signal luminescence were normalized by substracting the corresponding luminescence values from the 'mock' sample, as it had no DNA and the measured values represent background luminescence signal. SIMPL signal was further normalized to prey protein expression by dividing SIMPL signal luminescence values with those of prey luminescence.

Bait and prey protein expression was detected in all samples (Figure 9). The strongest SIMPL signal (Figure 10) corresponds to the LSM2/LSM3 positive control. When baits were co-transfected with RAF-RBD, samples showed a higher SIMPL signal than with LSM3. After normalization (Figure 11), the LSM2/3 pair showed the strongest, and bait/LSM3 showed the weakest interaction. Bait/RAF-RBD pairs showed higher interaction compared to bait/LSM3 pairs.

a) b)

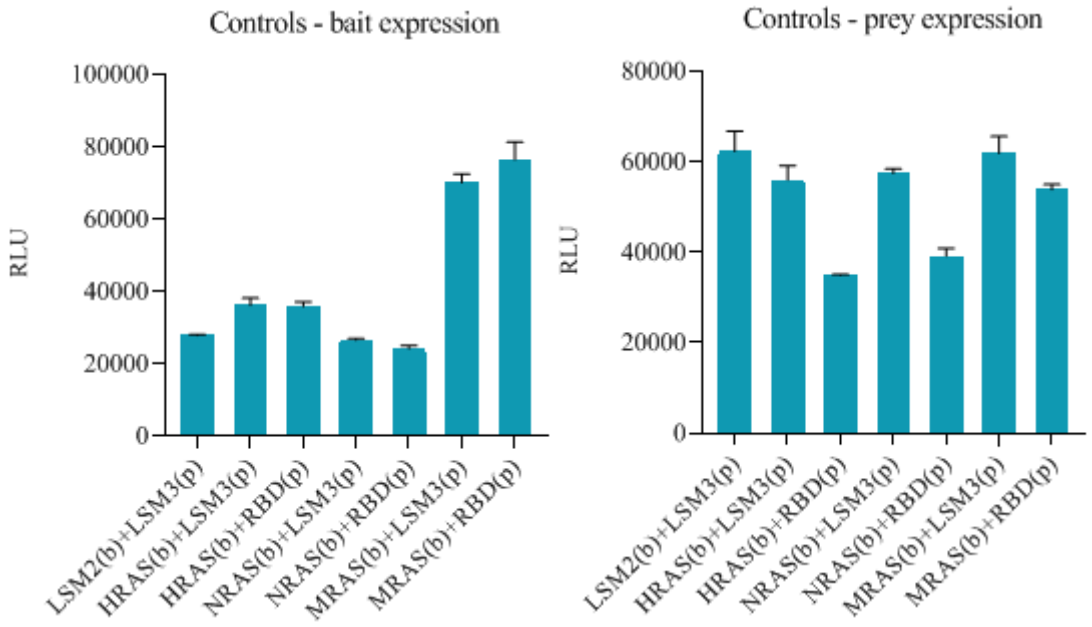


Figure 9. Bait and prey expression of positive and negative controls. a) Expression of bait proteins HRAS, NRAS, MRAS and LSM2, b) expression of prey proteins RBD and LSM3. Baits were captured with α-FLAG capture antibodies and detected with HRP-conjugated α-c-myc detection antibodies. Preys were captured with α-V5 antibodies and detected with HRP-conjugated α-HA detection antibodies. Error bars represent the standard deviation across biological duplicates; b-bait, p-prey, RLU-relative luminescence units.

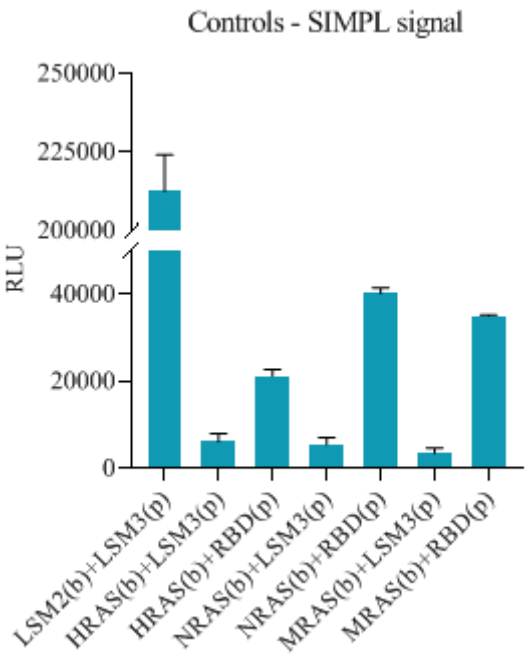


Figure 10. SIMPL signal of positive and negative controls. Bait proteins were captured with α-FLAG detection antibodies, and interacting prey proteins with HRP conjugated α-HA detection antibodies. Error bars represent the standard deviation across biological duplicates; b-bait, p-prey, RLU-relative luminescence units.

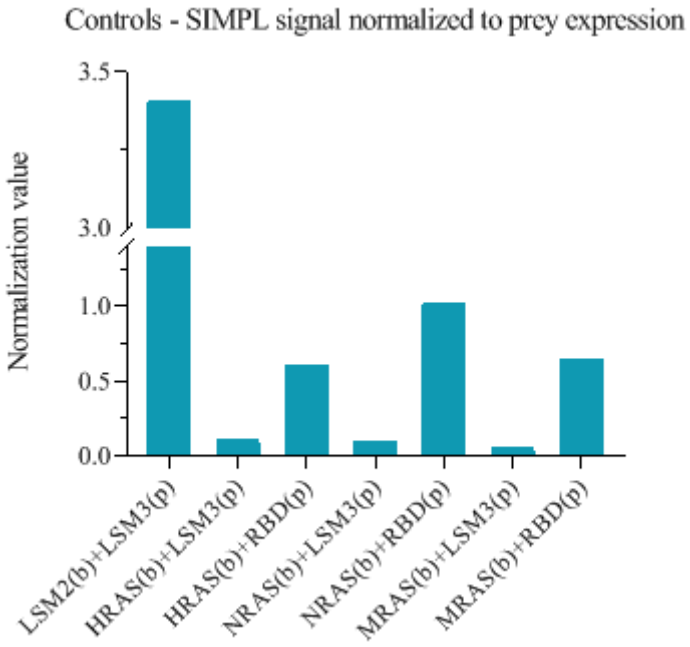


Figure 11. Normalized luminescence data of SIMPL assay for positive and negative controls. SIMPL signal luminescence was normalized to prey expression for all samples; b-bait, p-prey.

3.3. Screening of prey libraries

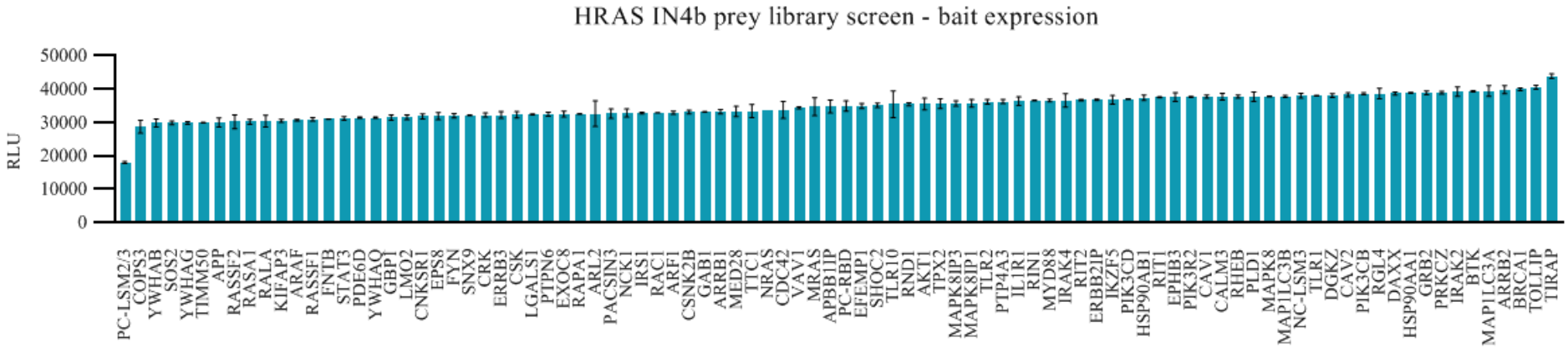
To detect protein-protein interactions of the selected Ras family members, IN4b and NIN4 prey libraries with potential protein interactors corresponding to each bait protein were screened. The two prey libraries, each containing 89 proteins for HRAS, 84 for NRAS, and 28 for MRAS, were generated by Gateway cloning into the C-tagged and N-tagged SIMPL prey vectors (designated IN4b and NIN4, respectively). The control prey proteins were in the IN4b vector, and the bait proteins in the N-tagged SIMPL bait vector (designated IC2). TRE-x HEK 293 cells were co-transfected in biological duplicates with bait and prey DNA, with RBD as a positive, LSM3 as a negative and LSM2/LSM3 bait/prey pair as an independent positive control. One day after transfection, protein expression was induced with 0.5 μg/μl Tet. The luminescence corresponding to bait and prey expression, as well as the SIMPL signal, was measured the following day. The background luminescence (from 'mock' sample; bait, prey and SIMPL signal) was substracted from the corresponding values for each sample. The luminescence data was further normalized as described above and the samples were sorted from lowest to highest value. The value 0.55 after SIMPL signal normalization to prey expression was chosen as the cut-off point identifying positive interaction.

3.3.1. HRAS IN4b and NIN4 prey library screens

Bait expression was detected in all samples for both screens (Figure 12 a) and Figure 15 a)). Of the 89 C-terminally tagged prey proteins, 13 showed a negative expression, while 7 additional proteins showed a low expression. The controls and the rest of the prey proteins showed a higher level of expression (Figure 12 b)). For 46 of the library proteins, SIMPL signal luminescence value was negative (Figure 13). Preys that had a negative and very low expression were removed from normalization to prey expression (Figure 14) to avoid deceptively positive values.

All of the N-terminally tagged prey proteins showed a positive luminescence value (Figure 15 b)), with one having a very low expression. Only one prey showed negative SIMPL signal value, while additional 3 proteins showed a low signal (Figure 16). SIMPL signal for all samples, with the exception of the one with the lowest prey expression expression was normalized as above (Figure 17).

a)



b)

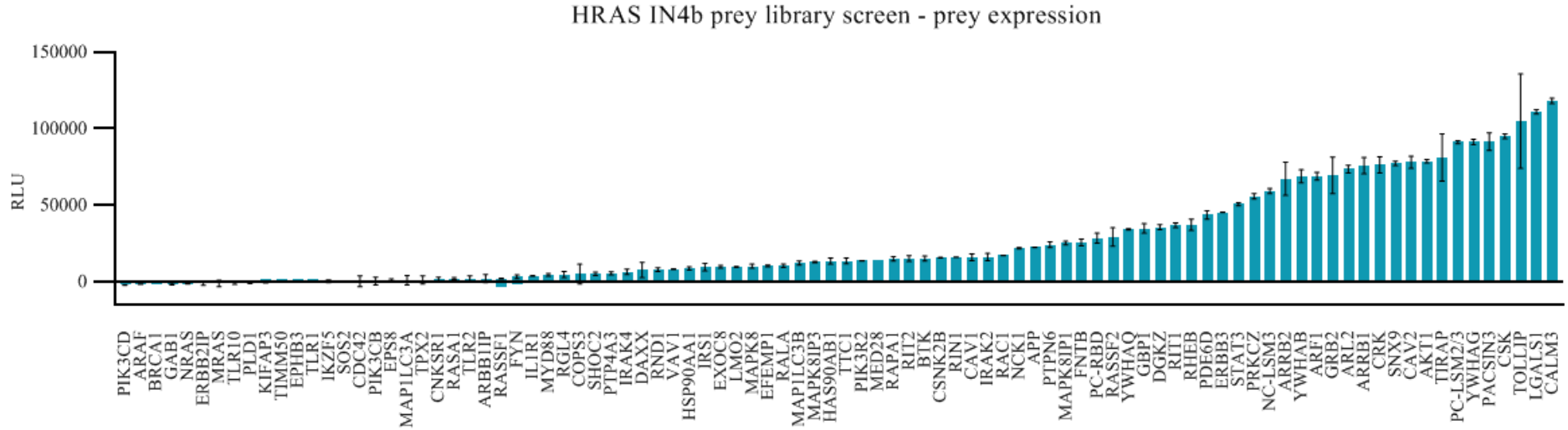


Figure 12. Bait a) and prey b) expression from the screen of IN4b prey library against HRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

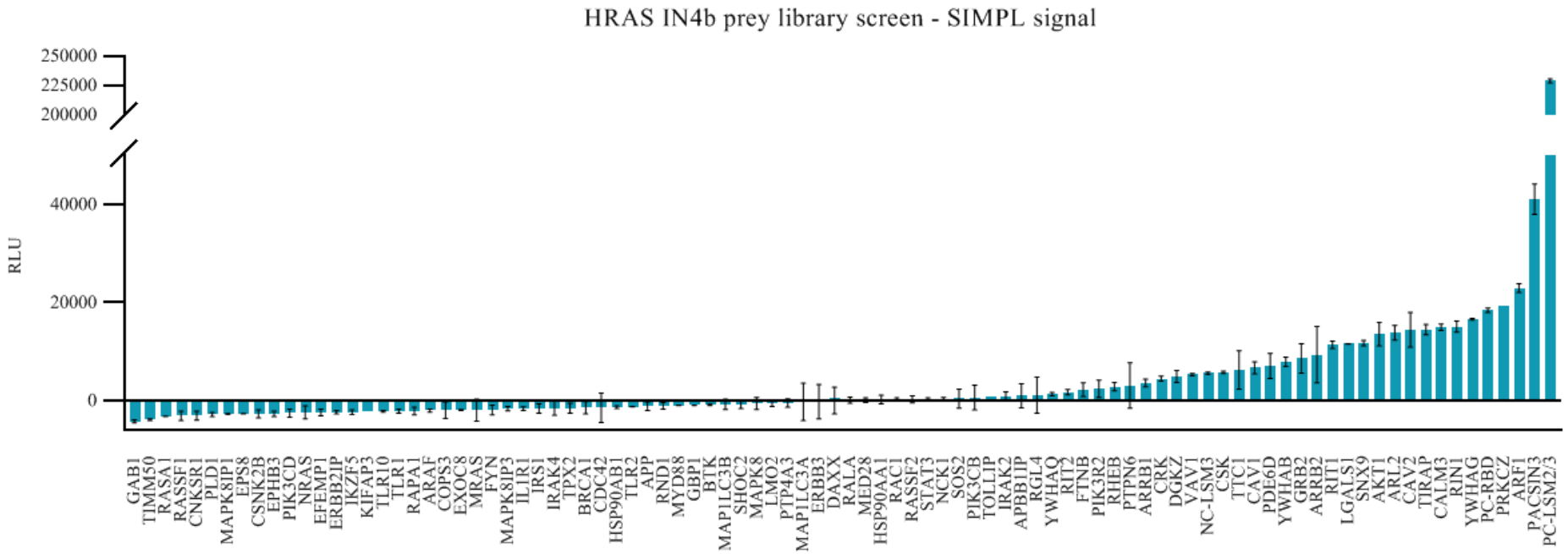


Figure 13. SIMPL signal from the screen of IN4b prey library against HRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

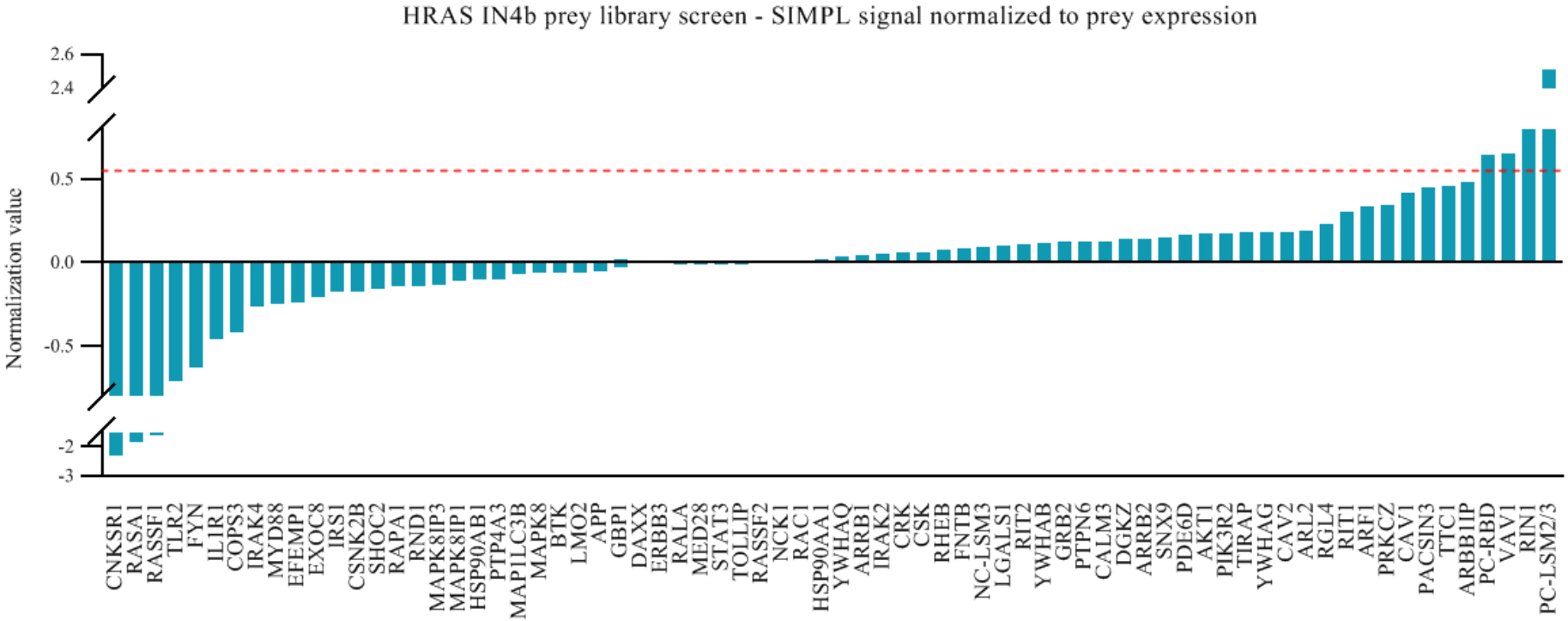
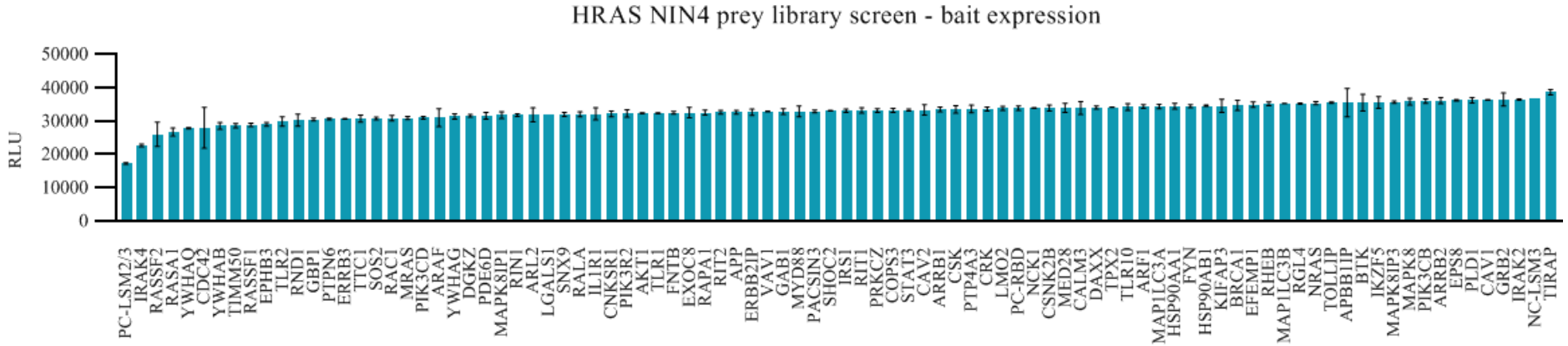


Figure 14. Normalized luminescence data of SIMPL signal from IN4b prey library screen against HRAS. The red line represents the 0.55 cut-off value.

a)



b)

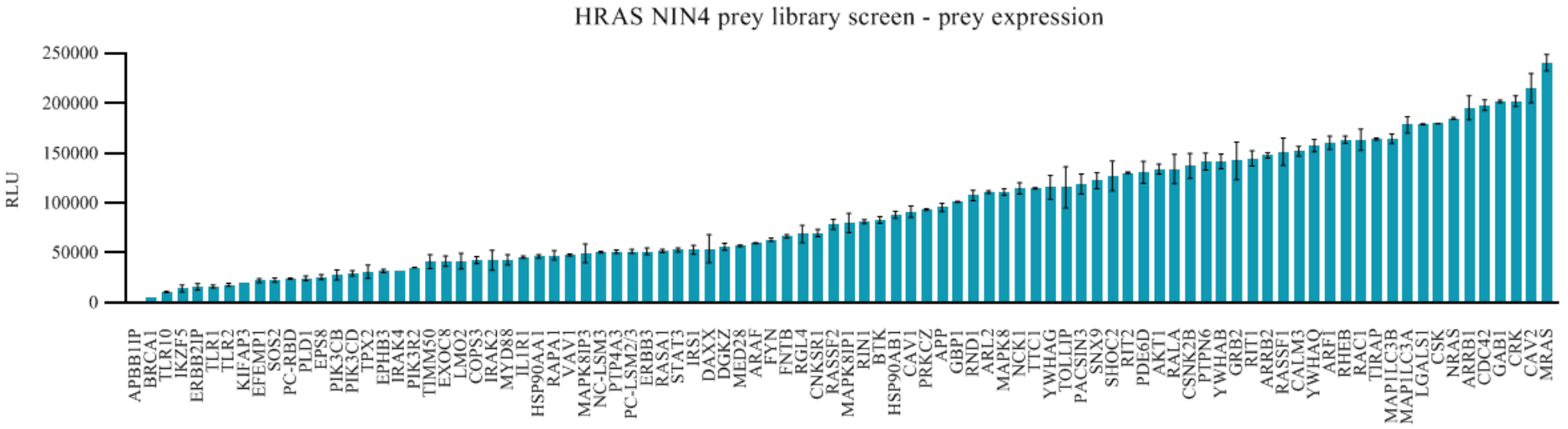


Figure 15. Bait a) and prey b) expression from the screen of NIN4 prey library against HRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

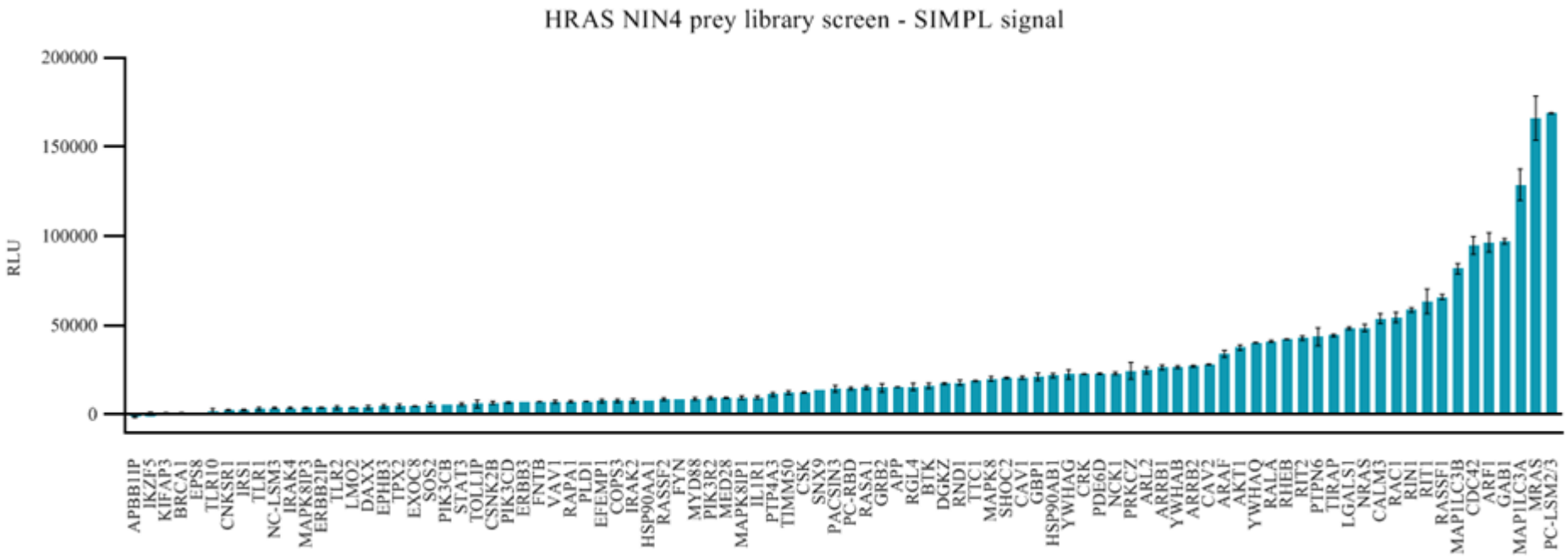


Figure 16. SIMPL signal from the screen of NIN4 prey library against HRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

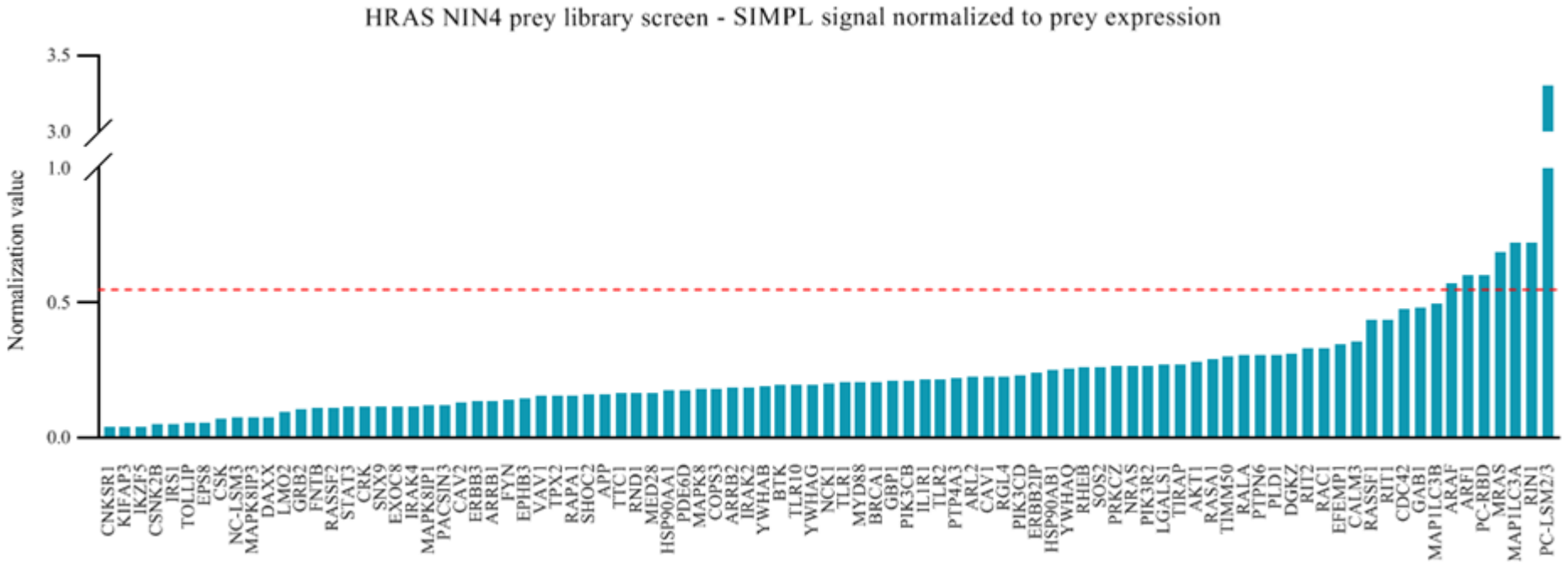


Figure 17. Normalized luminescence data of SIMPL signal from NIN4 prey library screen against HRAS. The red line represents the 0.55 cut-off value.

3.3.2. NRAS IN4b and NIN4 prey library screens

The screened libraries contained 84 C- and N-terminally tagged prey proteins. All samples showed a positive bait expression in both screens (Figure 18 a) and Figure 21 a)). In the IN4b prey library screen, 7 preys showed negative, and 4 showed a low expression (Figure 18 b)). 7 proteins had a negative, and 6 had s low SIMPL signal luminescence value (Figure 19). The 7 proteins with negative and 4 with low expression were excluded from normalization to prey expression (Figure 20).

In the NIN4 prey library screen against NRAS, all preys showed a positive expression (Figure 21 b)). SIMPL signal was negative for 3 samples, and one sample showed a low signal (Figure 22). The signal was normalized to prey expression for all samples (Figure 23).

3.3.3. MRAS IN4b and NIN4 prey library screens

The libraries screened against MRAS contained 28 prey proteins. All samples showed a positive bait expression in both screens (Figure 24 a) and Figure 27 a)). In the IN4b prey library screen, 2 proteins showed a negative and one a low expression (Figure 24 b)). 13 proteins had a negative, and 1 had a low SIMPL signal luminescence value (Figure 25). The three preys with negative and low expression were excluded from normalization (Figure 26).

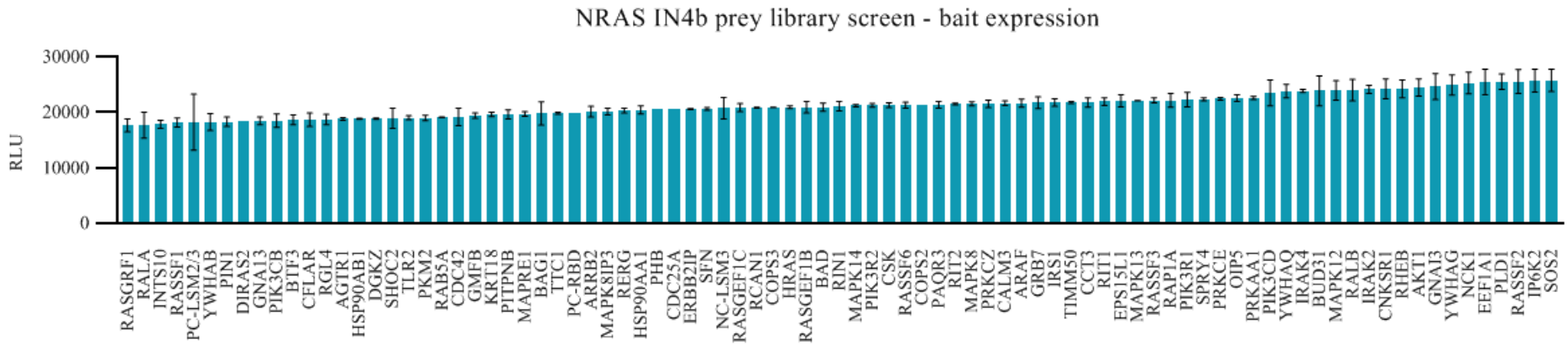
All preys in the NIN4 prey library screen showed a positive expression (Figure 27 b)). Two samples had a negative and one had a low SIMPL signal luminescence value (Figure 28). SIMPL signal was normalized to prey expression for all samples (Figure 29).

3.4. RAS dimerization

Although previously considered a monomeric GTPase, RAS dimer formation has been reported to occur on artificial membranes. Recent studies indicate RAS dimers may be critical in RAS organization and signaling *in vivo* and *in vitro* (Chen et al. 2015). As results from the screens indicated interaction between HRAS and NRAS, HRAS and MRAS and MRAS and NRAS, to verify dimer formation via SIMPL, bait proteins were aditionally tested against each other. HRAS, NRAS and MRAS used in the screen were N-terminally tagged (in NIN4

vector) and LSM3 was used as a negative control. All samples showed a positive bait and prey expression (Figure 30). SIMPL signal luminescence was low for the negative control. All tested combinations of HRAS, NRAS and MRAS showed a high SIMPL signal (Figure 31). When normalized to prey expression, all three RAS proteins showed higher values compared to the LSM3 negative control (Figure 32).

a)



b)

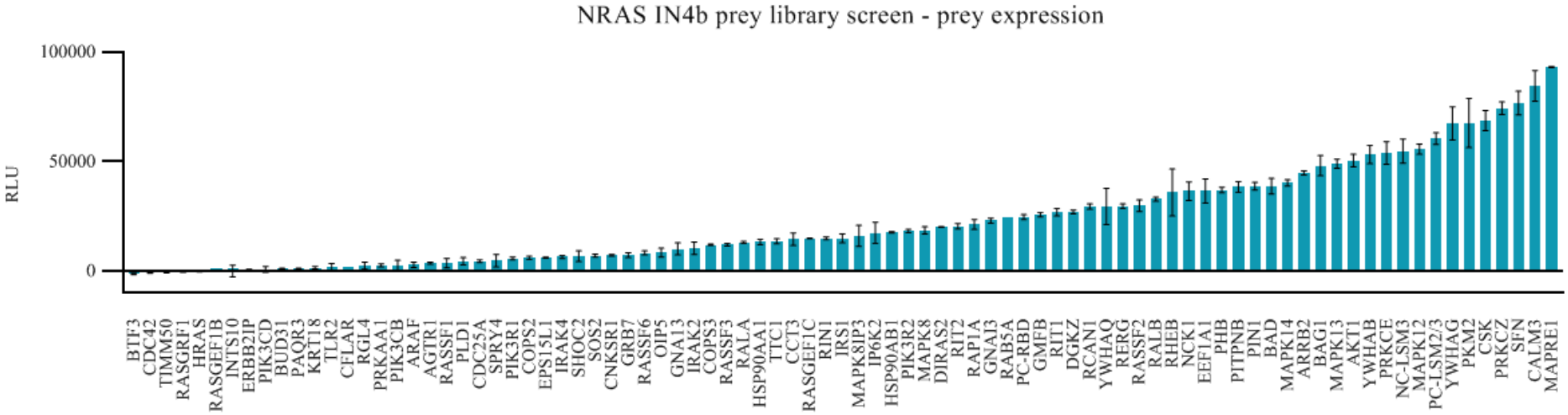


Figure 18. Bait a) and prey b) expression from the screen of IN4b prey library against NRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

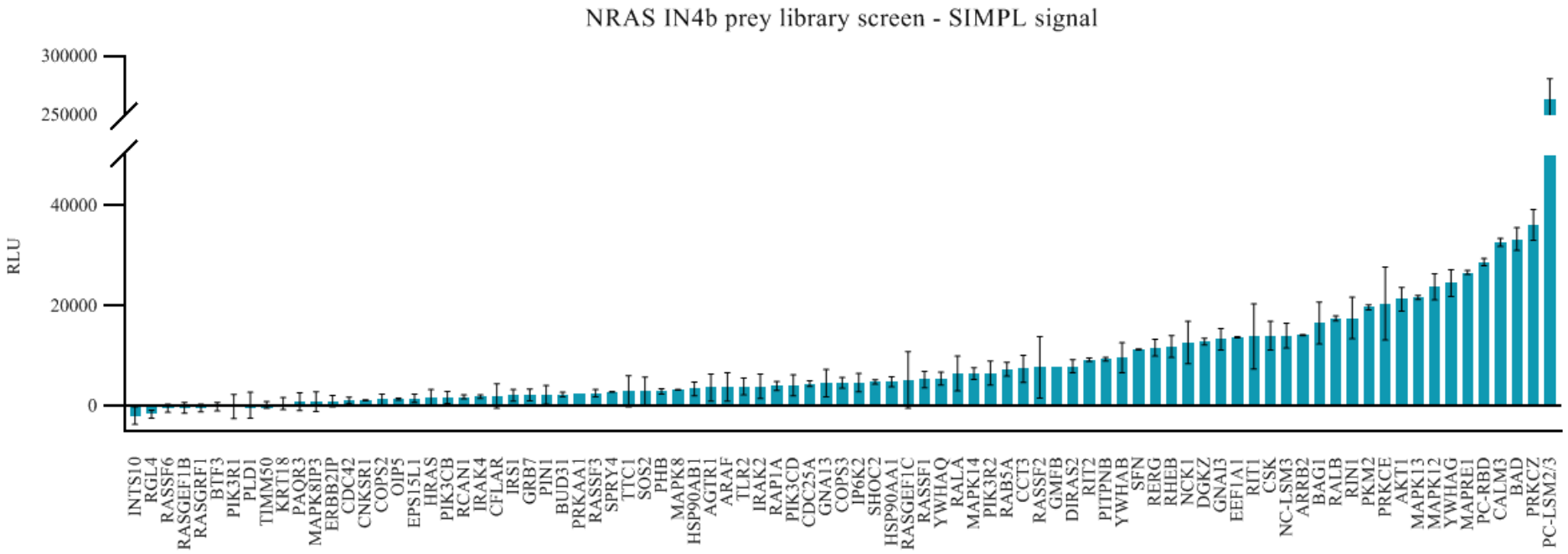


Figure 19. SIMPL signal from the screen of IN4b prey library against NRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

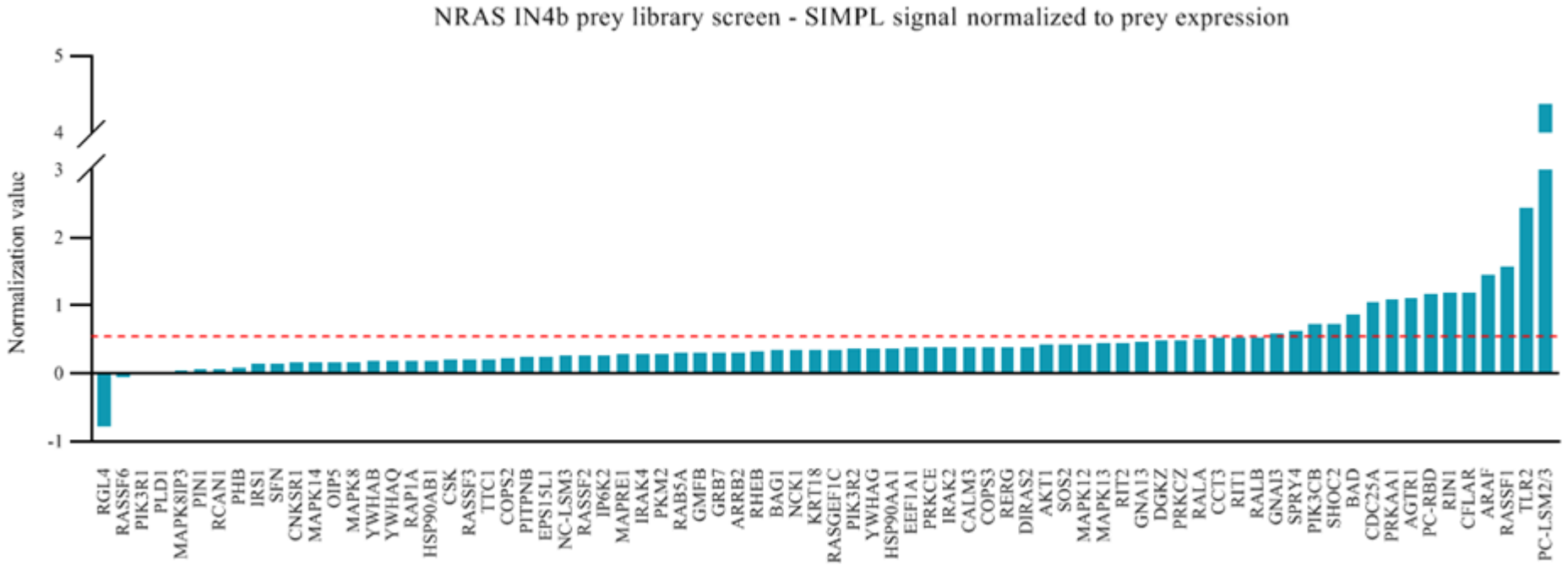
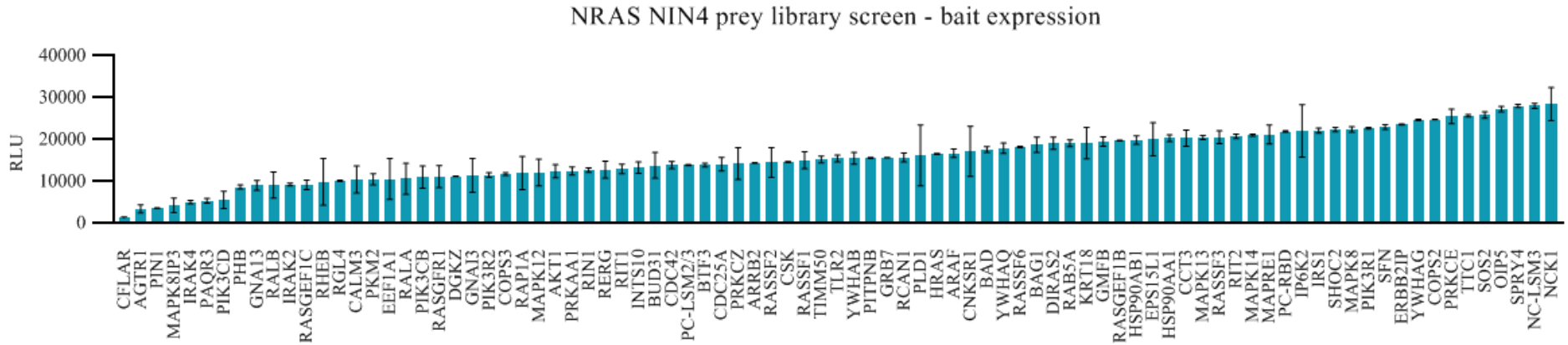


Figure 20. Normalized luminescence data of SIMPL signal from IN4b prey library screen against NRAS**.** The red line represents the 0.55 cut-off value.

a)



b)

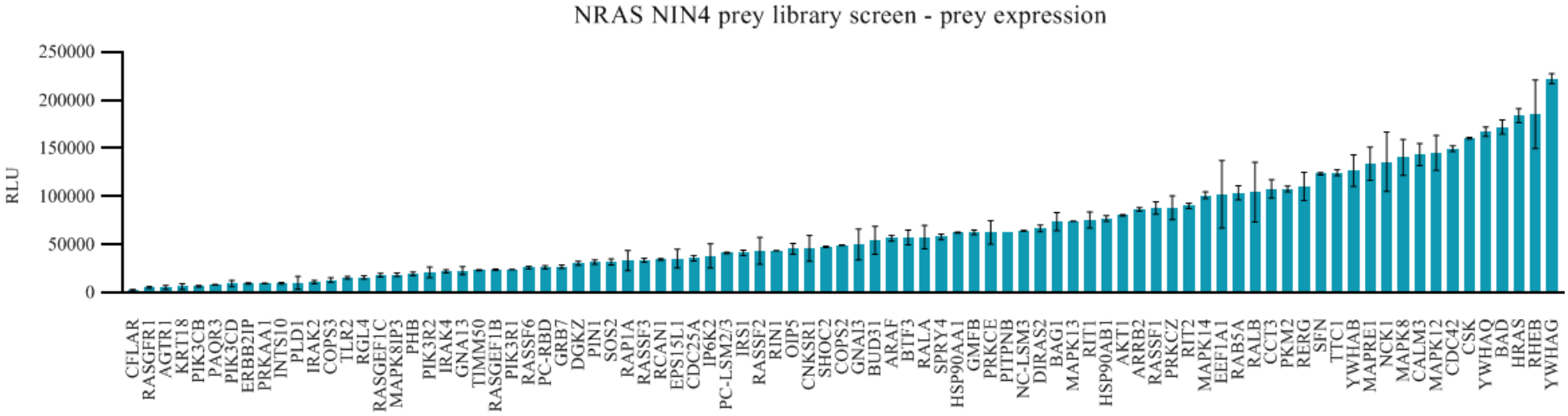


Figure 21. Bait a) and prey b) expression from the screen of NIN4 prey library against NRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

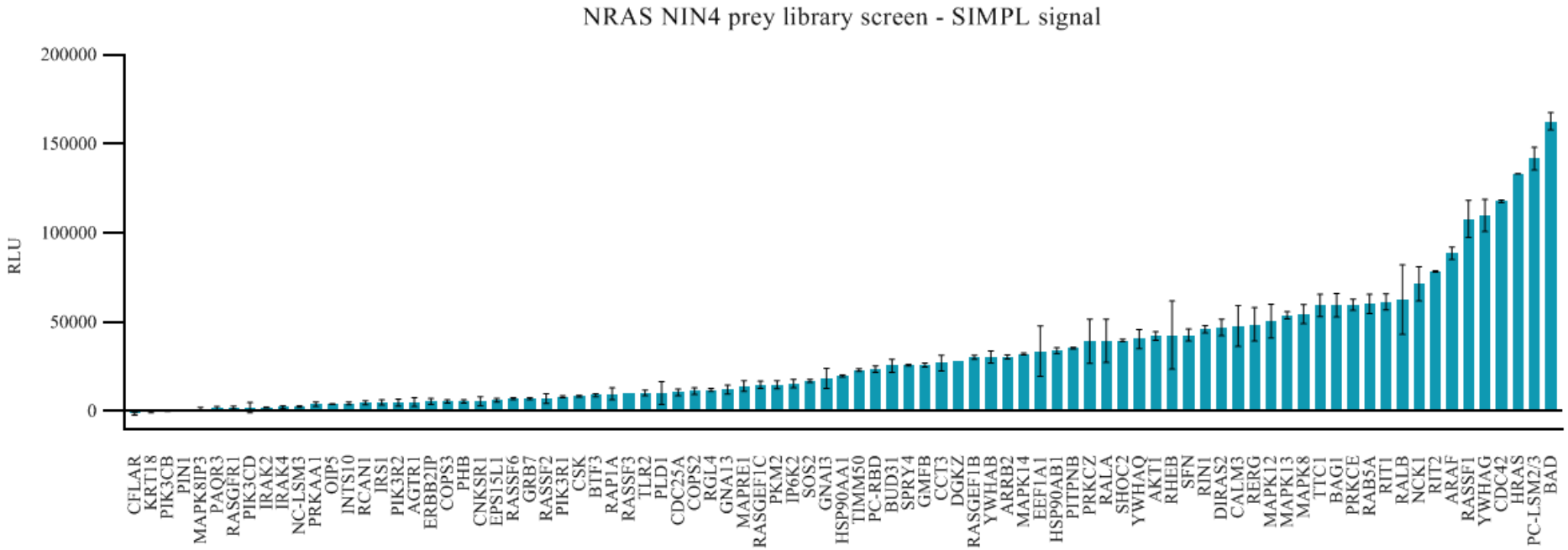


Figure 22. SIMPL signal from the screen of NIN4 prey library against NRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

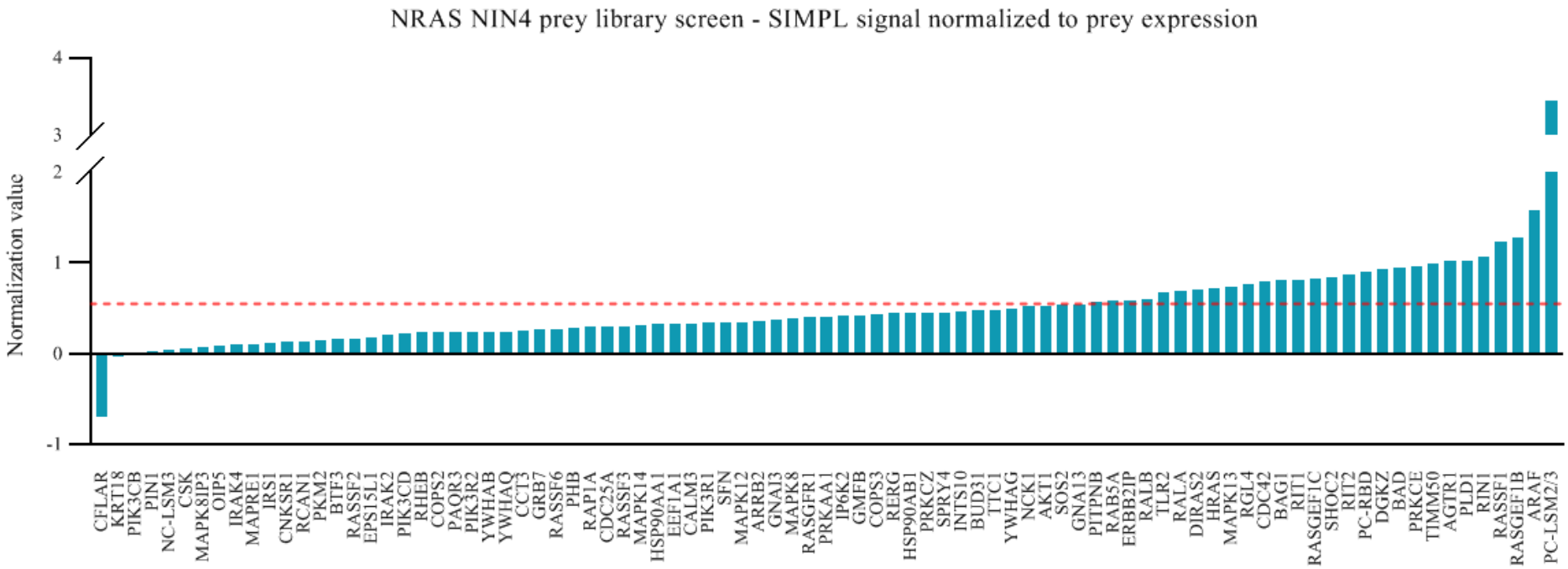
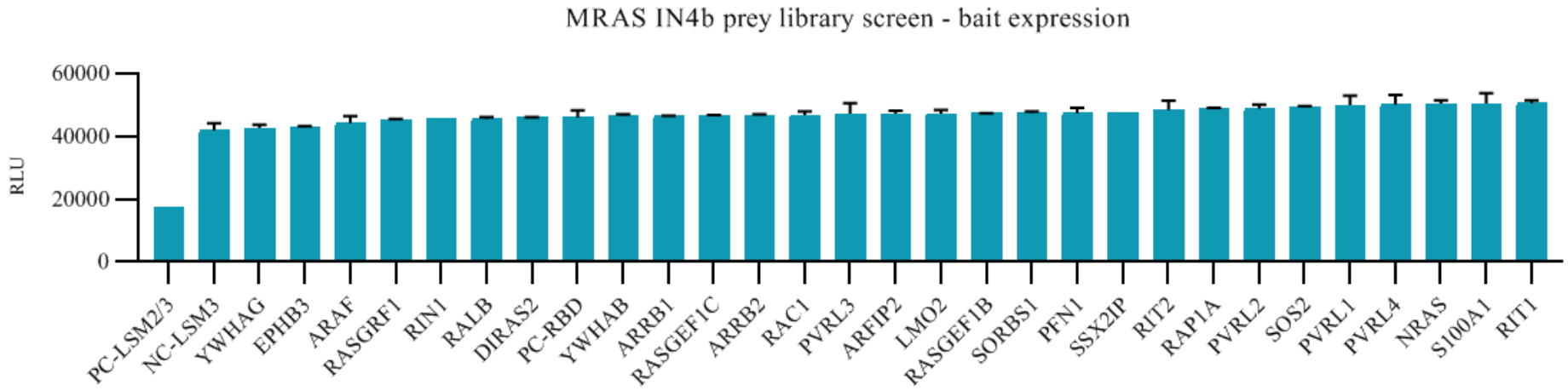


Figure 23. Normalized luminescence data of SIMPL signal from NIN4 prey library screen against NRAS. The red line represents the 0.55 cut-off value.

a)



b)

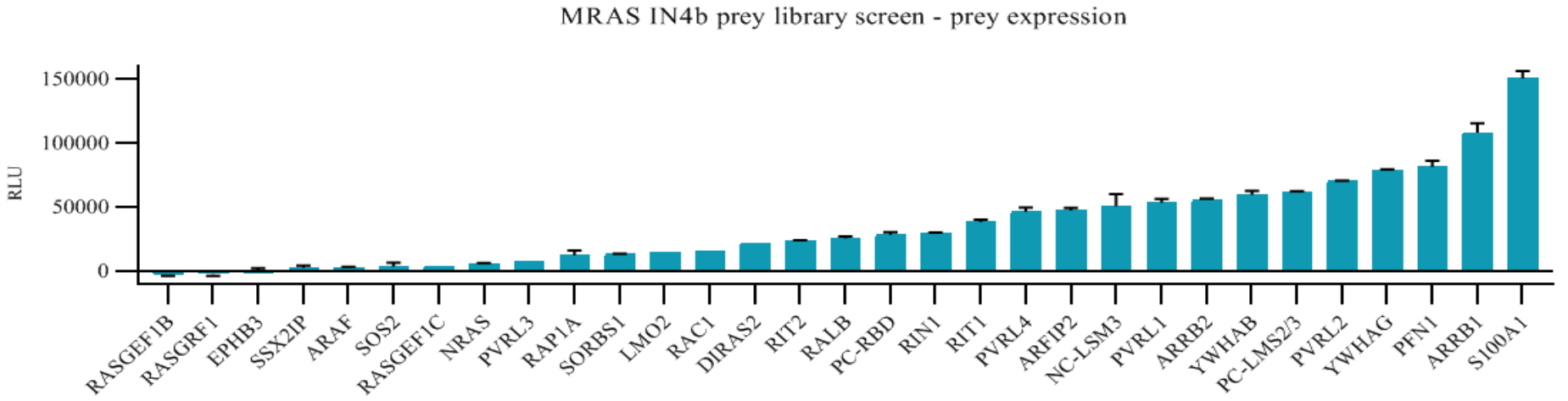


Figure 24. Bait a) and prey b) expression from the screen of IN4b prey library against MRAS**.** Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

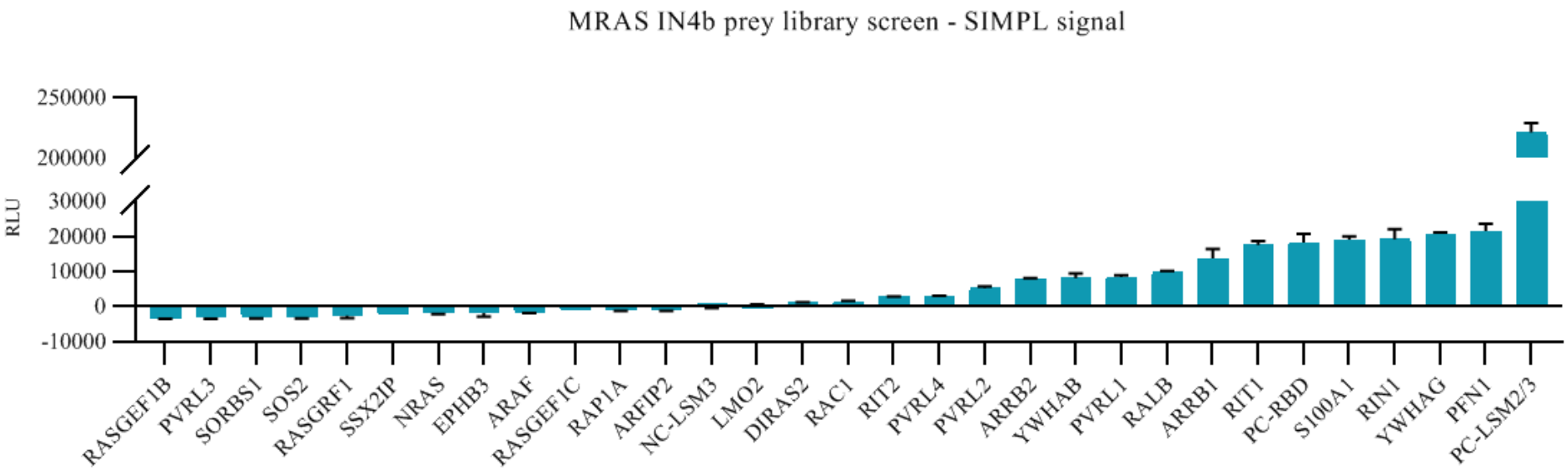


Figure 25. SIMPL signal from the screen of IN4b prey library against MRAS**.** Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

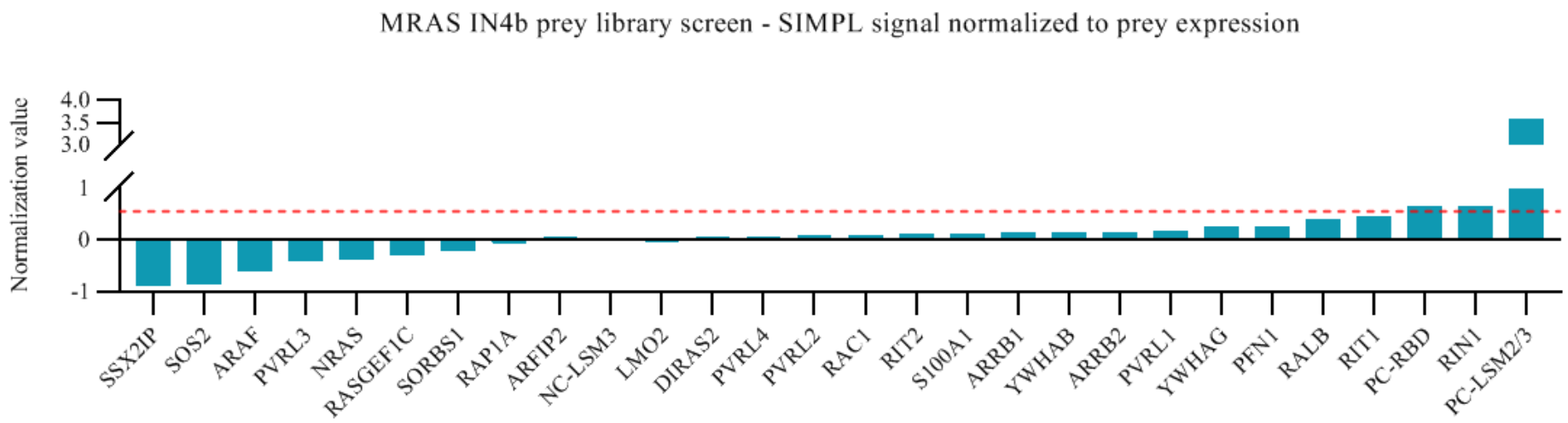
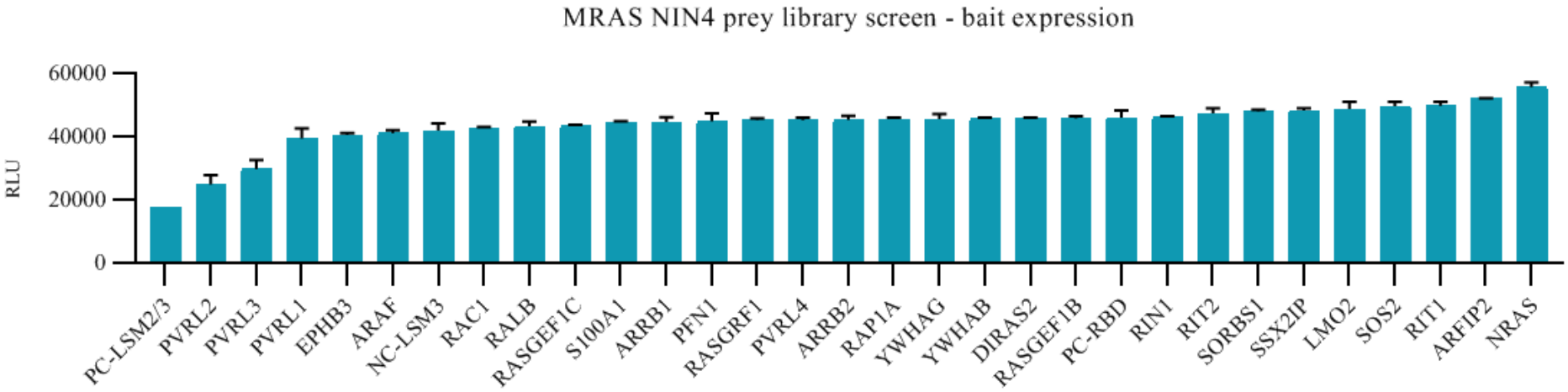


Figure 26. Normalized luminescence data of SIMPL signal from IN4b prey library screen against MRAS. The red line represents the 0.55 cut-off value.

a)



b)

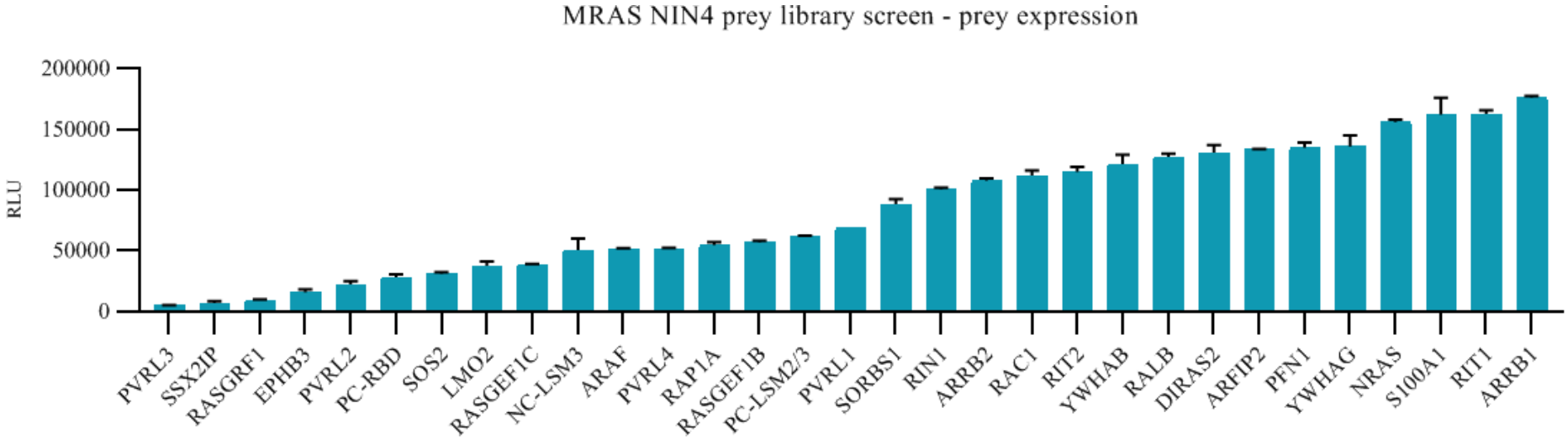


Figure 27. Bait a) and prey b) expression from the screen of NIN4 prey library against MRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

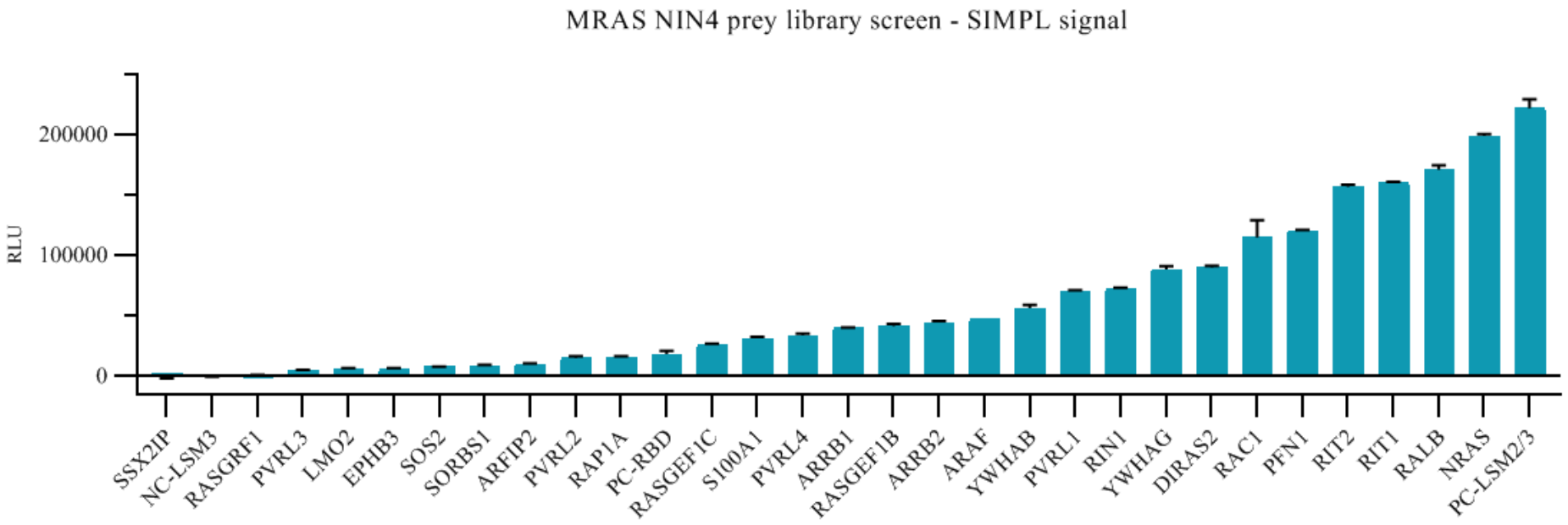


Figure 28. SIMPL signal from the screen of NIN4 prey library against MRAS. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

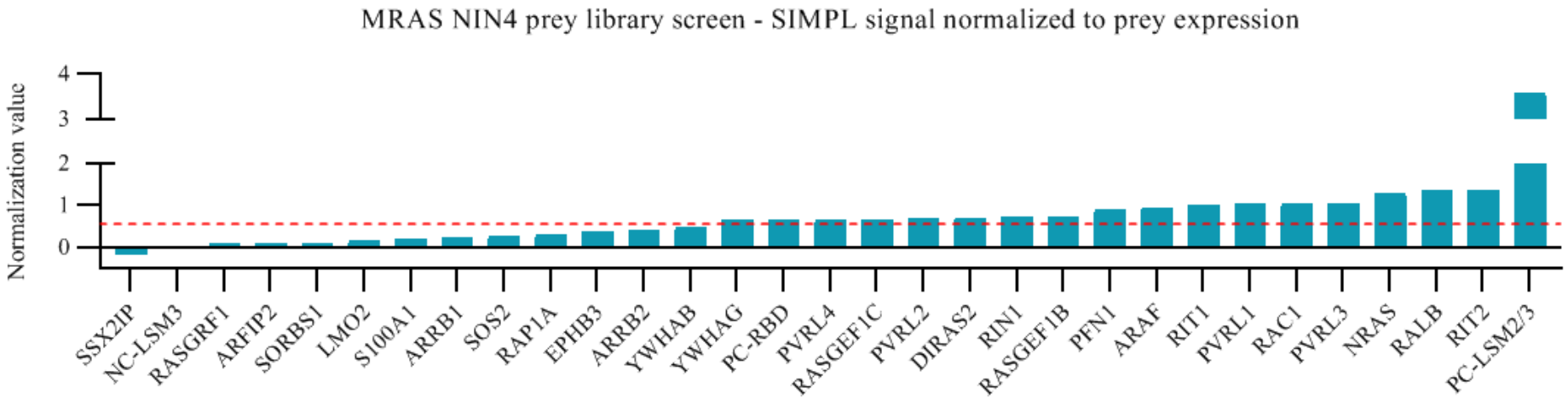


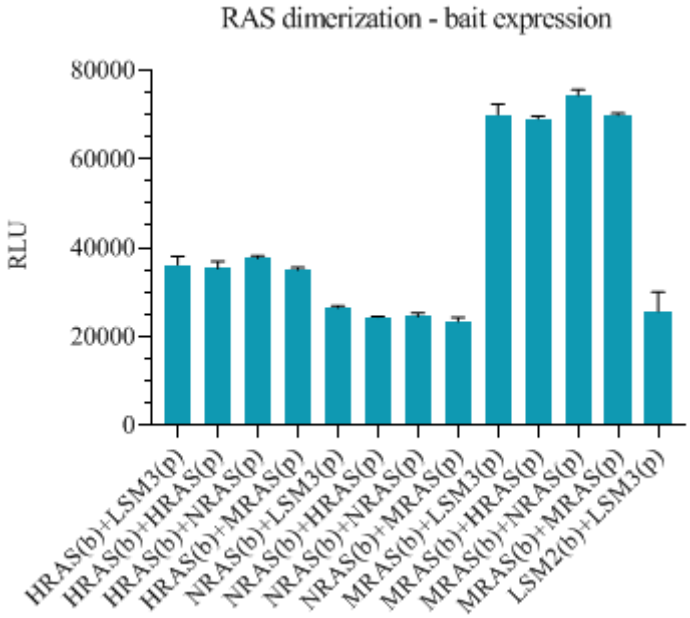
Figure 29. Normalized luminescence data of SIMPL signal from NIN4 prey library screen against MRAS. The red line represents the 0.55 cut-off value.

Table 1. Identified positive interactors for all three bait proteins from IN4b and NIN4 prey library screens. Known interactors are highlighted in green.

HRAS IN4b prey library						
	Normalization Value	Protein	FpClass score	Normalization value	Protein	FpClass score
	0,9485	RIN1	0,8013	0,6543	VAV1	0,5389
HRAS NIN4 prey library						
	0,7244	RIN1	0,8013	0,6030	ARF1	0,7085
	0,7232	MAP1LC3A	0,4550	0,5706	ARAF	0,8826
	0,6906	MRAS	0,7738			
NRAS IN4b library screen						
	2,4303	TLR2	0,3339	1,0435	CDC25A	0,4410
	1,5704	RASSF1	0,5644	0,8619	BAD	0,5194
	1,4545	ARAF	0,8826	0,7332	SHOC2	0,8826
	1,2009	CFLAR	0,3917	0,7313	PIK3CB	0,6824
	1,1926	RIN1	0,2928	0,6226	SPRY4	0,4966
	1,1057	AGTR1	0,4577	0,5835	GNAI3	0,3479
	1,0957	PRKAA1	0,5166			
NRAS NIN4 library screen						
	1,5729	ARAF	0,8826	0,8166	RIT1	0,6068
	1,2751	RASGEF1B	0,3220	0,8085	BAG1	0,4597
	1,2282	RASSF1	0,5644	0,7885	CDC42	0,2946
	1,0677	RIN1	0,2928	0,7626	RGL4	0,4796
	1,0207	PLD1	0,2766	0,7267	MAPK13	0,2727
	1,0189	AGTR1	0,4577	0,7252	HRAS	0,8826
	0,9923	TIMM50	0,4054	0,7048	DIRAS2	0,2557
	0,9575	PRKCE	0,2604	0,6856	RALA	0,6850
	0,9462	BAD	0,5194	0,6702	TLR2	0,3339
	0,9288	DGKZ	0,8049	0,6005	RALB	0,3097
	0,8725	RIT2	0,3295	0,5901	ERBB2IP	0,8826
	0,8369	SHOC2	0,8826	0,5805	RAB5A	0,2882
	0,8314	RASGEF1C	0,3133	0,5634	PITPNB	0,2811
MRAS IN4b library screen						

	0,6492	RIN1	0,2727			
MRAS NIN4 library screen						
	1,3631	RIT2	0,4166	0,8886	PFN1	0,5184
	1,3423	RALB	0,2715	0,7235	RASGEF1B	0,3017
	1,2735	NRAS	0,2647	0,7135	RIN1	0,2727
	1,0333	PVRL3	0,6264	0,6872	DIRAS2	0,2657
	1,2064	RAC1	0,7358	0,6769	PVRL2	0,5194
	1,0166	PVRL1	0,6414	0,6594	RASGEF1C	0,3346
	0,9862	RIT1	0,3731	0,6502	PVRL4	0,6025
	0,9331	ARAF	0,3184	0,6474	YWHAG	0,3327

a)



b)

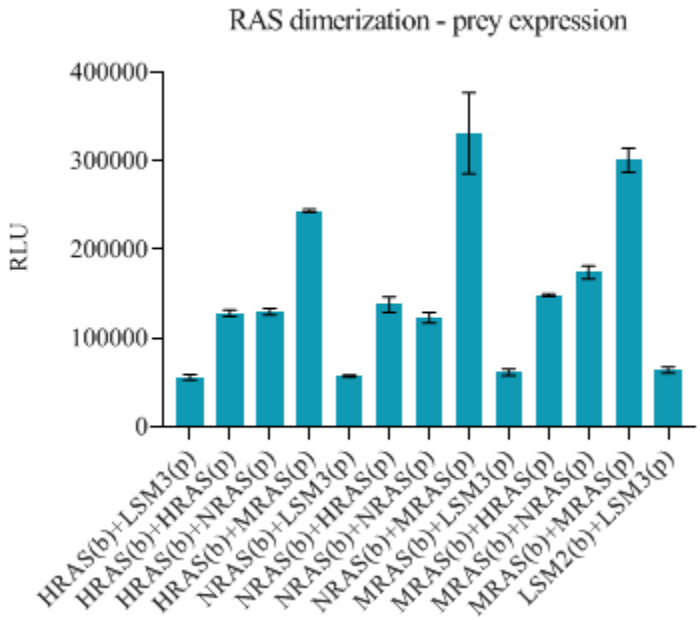


Figure 30. Bait a) and prey b) expression from the mini-screen of RAS proteins against each other. Error bars represent the standard deviation across biological duplicates; b-bait, p-prey, RLU-relative luminescence units.

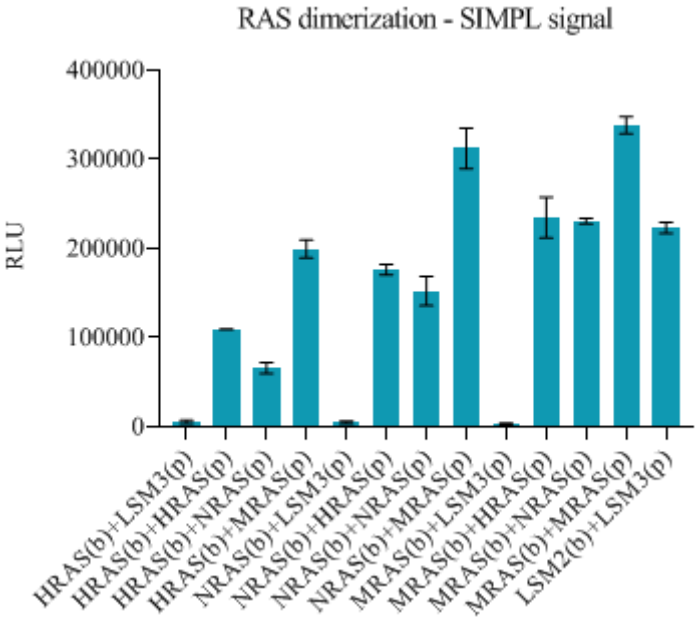


Figure 31. SIMPL signal from the mini-screen of RAS proteins against each other. Error bars represent the standard deviation across biological duplicates, RLU-relative luminescence units.

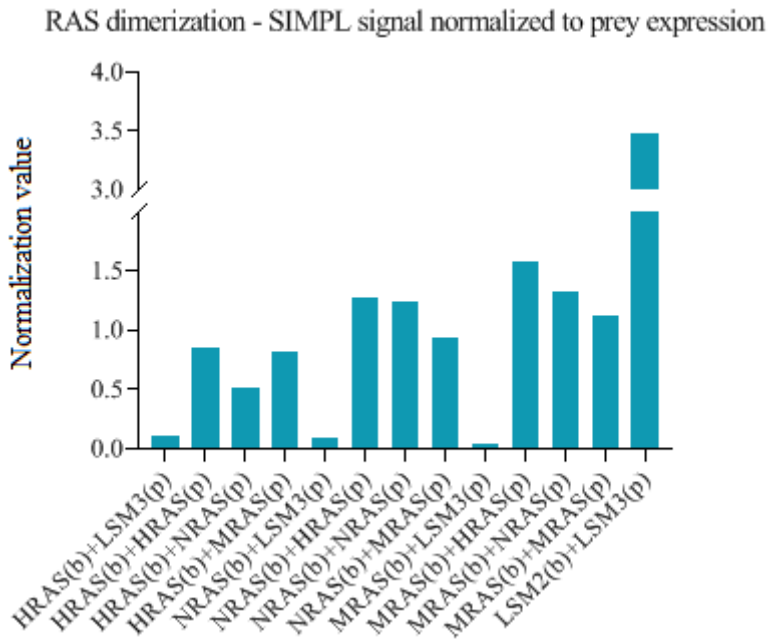


Figure 32. Normalized luminescence data of SIMPL signal from the mini-screen of RAS proteins against each other.

4. Discussion

4.1. Transient bait expression and testing of positive and negative interactors

Prior to the screenings, transient bait protein expression was tested using Western blot. The expression of RAS proteins was detected in both Tet induced and uninduced samples. As TRE-x HEK293 cells stably express Tet repressor, and bait protein expression is under control of a Tet operator, such a result may have arisen due to a high transfection efficiency as the amount of Tet repressor in the cells may have been deficient for proper repression. Design and construction of stable cell lines, where bait expression can be better regulated could be used to circumvent this issue.

To validate the SIMPL system, baits were tested against positive and negative interactors. LSM3 was chosen as a negative control as it had not been previousy reported to interact with any of the bait proteins. LSM2 and LSM3 pair were chosen as an independent positive control as they are part of a complex involved in mRNA degradation and interact with each other (Wu et al. 2014). RAF proteins are serine/threonine kinases and direct effectors of activated RAS proteins, crucial in RAF/MEK/ERK signaling. RAF-RBD makes initial contact with active RAS and binds with high affinity (Terrell and Morrison 2019) and was used as a positive control. A difference in interaction between positive and negative controls was observed. As prey expression was higher for LSM3 compared to RAF-RBD for all tested baits, and SIMPL signal was higher for RAF-RBD compared to LSM3, they were deemed suitable controls for prey library screens.

4.2. Screening results and RAS dimerization

Prey libraries for screening against the three bait proteins were generated by Gateway cloning of the ORFeome genes into two destination vectors, in order to obtain both N- and C-terminally tagged preys. The N-terminally tagged preys performed better in all screenings. This could be either due to the tags interfering with post-translational modifications, folding and protein function in C-terminally tagged preys, or due to a better spatial arrangement which enables interaction in N-terminally tagged preys. SIMPL signal is a direct

quantification of interaction between bait and prey proteins (Yao et al. 2020) and it can be normalized to bait and/or prey expression. Determining where to draw the line between positive and negative interactors presents a challenge due to the novelty of the technique and the lack of a statistics framework for analyzing the data. For this reason, value of 0.55 after normalization to prey expression was chosen as the cut-off value as it includes the RAF-RBD positive controls and is higher than the LSM3 negative controls for all screenings. Since prey expression largely differs between samples, normalization to prey expression is useful when comparing interaction strength between different proteins.

Six interactors were identified for HRAS from the screenings, and two additionally from the RAS dimerization experiment. Thirty-two were identified for NRAS, and sixteen for MRAS, with two additions for each from the RAS dimerization experiment. The complete lists of positive interactors identified from screenings for HRAS, NRAS and MRAS are in Table 1. According to the BioGRID database (https://thebiogrid.org/), two of the interactors are already known for HRAS and seven for NRAS, while none of the positive interactors for MRAS are previously reported interactors.

The role of several interactors in the context of RAS signaling has already been established. RIN1 (RAS and RAB interactor 1) binds to RAS with high affinity and competes with that of RAF1 (Wang et al. 2002). It activates the ABL tyrosine kinases and mediates actin remodeling involved in adhesion and migration of epithelial cells (Hu, Bliss, and Wang 2005). The effect of RIN1 activation is different in different tissues and tumour types. While RIN1 expression is increased in gastric adenocarcinoma and linked with poor prognosis (Yu et al. 2012), low levels in hepatocellular carcinoma are associated with tumour invasion (He et al. 2013). RAS also binds RAF (including ARAF) proteins with high affinity. Kinase activity of RAF leads to initiation of the MEK-ERK protein kinase cascade which affects transcription (Colicelli 2004). RIN1 and ARAF were identified as positive interactors for all three bait proteins.

RALA and RALB (Ras-like A and B) are activated by RalGEF, a major RAS effector (Mishra et al. 2010). RALA is needed for anchorage-independent proliferation of tumour cells while RALB is vital for tumour cell survival (Chien and White 2003). Both proteins have a key role in various activities through interaction with their downstream effectors, such as tumor formation, metastasis, vesicular trafficking and gene expression (Gentry et al. 2014). RIT1 and RIT2 (RAS-like without CAAX 1 and 2) were identified as interactors for both

NRAS and MRAS. They are crucial for neuronal cell survival (Colicelli 2004). RIT1 gain-of-function mutations have also recently been found to cause Noonan syndrome (Aoki et al. 2013) and have been reported in a variety of human cancers, including hepatocellular carcinoma, myeloid malignancies and breast cancer (Fang et al. 2016). DIRAS2 (Distinct Subgroup of RAS 2) is downregulated in ovarian cancer. Its re-expression has been shown to induce autophagy-mediated cell death by inhibition of the AKT1-mTOR and RAS-MAPK signaling pathways (Sutton et al. 2018). CDC42 (Cell Division Cycle 42) is a Rho family small GTPase that can regulate cell proliferation, polarity, vesicle trafficking and actin remodeling and has been found to become activated upon oncogenic RAS expression (Stengel and Zheng 2012). Another Rho GTPase, RAC1, has been linked to regulation of lamellipodium formation and membrane ruffling (Zhang, Chernoff, and Zheng 1998). RAC1 gain-of-function mutations in cancer that confer resistance to targeted therapies and contribute to tumour phenotype have been identified (De, Aske, and Dey 2019). BAD (BCL-2 Associated Agonist of Cell Death) is a BH3-only protein that inhibits antiapoptotic and promotes proapoptotic signals. It is a substrate of various kinases, including AKT (Matsuura et al. 2016) that functions downstream of RAS activated PI3K, a component in cell transformation (Colicelli 2004).

For the majority of discovered interactors, their role in RAS signaling as well as the functional effect of interactions with RAS proteins remains to be discovered. The proteins tested in these screenings were predicted using FpClass. FpClass identifies most likely interactors based on sets of compatible protein features that may act cooperatively (Kotlyar et al. 2014). More than 400 proteins were predicted for HRAS and NRAS and more than 100 for MRAS. Due to the entry clone availability in the ORFeome collection or due to a faulty DNA sequence, only a portion of these proteins were screened against a certain bait. Additionally, predicted protein interactors differed largely for all three bait proteins. An expanded, unbiased screen with same preys for all baits should be considered to get a better picture of PPIs involved with RAS proteins.

In the protein screens against HRAS, there were 24 known interactors, of which only 2 were identified as positive due to a high normalization cut-off value. Providing a proper statistical method to analyze this data is established, other known interactors may prove to also be identifiable using SIMPL. This also highlights the importance of using different protein-protein interaction identification methods for validating positive hits, as not all methods are equally successful.

The RAS dimerization experiment showed that HRAS, NRAS and MRAS interact with each other forming homo- and heterodimers. HRAS dimerization has been previously reported (Lin et al. 2014). NRAS dimerization has also been established on an artificial POPC membrane. The proteins may dimerize because it helps nanoclustering in the membrane (Güldenhaupt et al. 2012). Although the possible function for RAS dimer formation and potential mechanisms for the regulation of RAS dimer formation and signaling have been suggested, further research is needed (Chen et al. 2015).

5. Conclusion

Eight interactors were identified for HRAS, thirty-four for NRAS and eighteen for MRAS. The mapped interactions include homo- and heterodimerization of HRAS, NRAS and MRAS with each other.

6. References

Alcantara, Krizelle Mae M et al. 2019. “Non- Redundant and Overlapping Oncogenic Readouts of Non- Canonical and Novel Colorectal Cancer KRAS and NRAS Mutants.” *Cells* 8(12): 1557.

Aoki, Yoko et al. 2013. “Gain-of-Function Mutations in RIT1 Cause Noonan Syndrome, a RAS/MAPK Pathway Syndrome.” *American Journal of Human Genetics* 93(1): 173–80.

Aranko, A. Sesilja, Alexander Wlodawer, and Hideo Iwaï. 2014. “Nature’s Recipe for Splitting Inteins.” *Protein Engineering, Design and Selection* 27(8): 263–71.

Brunsveld, Luc, Herbert Waldmann, and Daniel Huster. 2009. “Membrane Binding of Lipidated Ras Peptides and Proteins - The Structural Point of View.” *Biochimica et Biophysica Acta - Biomembranes* 1788(1): 273–88.

Chen, Mo, Alec Peters, Tao Huang, and Xiaolin Nan. 2016. “Ras Dimer Formation as a New Signaling Mechanism and Potential Cancer Therapeutic Target.” *Mini-Reviews in Medicinal Chemistry* 16(5): 391–403.

Chien, Yuchen, and Michael A. White. 2003. “RAL GTPases Are Linchpin Modulators of Human Tumour-Cell Proliferation and Survival.” *EMBO Reports* 4(8): 800–806.

Chinton, Josefina et al. 2019. “Clinical and Molecular Characterization of Children with Noonan Syndrome and Other RASopathies in Argentina.” *Archivos Argentinos de Pediatria* 117(5): 330–37.

Colicelli, John. 2004. “Human RAS Superfamily Proteins and Related GTPases.” *Science’s STKE : signal transduction knowledge environment* 2004(250): re13–re13.

De, Pradip, Jennifer Carlson Aske, and Nandini Dey. 2019. “RAC1 Takes the Lead in Solid Tumors.” *Cells* 8(5): 382.

Engin, H. Billur, Daniel Carlin, Dexter Pratt, and Hannah Carter. 2017. “Modeling of RAS Complexes Supports Roles in Cancer for Less Studied Partners.” *BMC Biophysics* 10(S1).

Fang, Zhenhao et al. 2016. “Biochemical Classification of Disease-Associated Mutants of RAS-like Protein Expressed in Many Tissues (RIT1).” *Journal of Biological Chemistry* 291(30): 15641–52.

Fernández-Medarde, Alberto, and Eugenio Santos. 2011. “Ras in Cancer and Developmental Diseases.” *Genes and Cancer* 2(3): 344–58.

Gentry, Leanna R., Timothy D. Martin, David J. Reiner, and Channing J. Der. 2014. “Ral Small GTPase Signaling and Oncogenesis: More than Just 15 Minutes of Fame.” *Biochimica et Biophysica Acta - Molecular Cell Research* 1843(12): 2976–88.

Gogarten, J. Peter et al. 2002. “Inteins: Structure, Function, and Evolution.” *Annual Review of Microbiology* 56(1): 263–87.

Güldenhaupt, Jörn et al. 2012. “N-Ras Forms Dimers at POPC Membranes.” *Biophysical Journal* 103(7): 1585–93.

Gysin, Stephan, Megan Salt, Amy Young, and Frank McCormick. 2011. “Therapeutic Strategies for Targeting Ras Proteins.” *Genes and Cancer* 2(3): 359–72.

He, Hui et al. 2013. “Low RIN1 Expression in HCC Is Associated With Tumor Invasion and Unfavorable Prognosis.” *American Journal of Clinical Pathology* 140(1): 73–81.

Hu, Hailiang, Joanne M Bliss, and Ying Wang. 2005. “RIN1 Is an ABL Tyrosine Kinase Activator and a Regulator of Epithelial-Cell Adhesion and Migration.” *Current Biology* 15: 815–23.

Johnson, Derek S., and Youhai H. Chen. 2012. “Ras Family of Small GTPases in Immunity and Inflammation.” *Current Opinion in Pharmacology* 12(4): 458–63.

Karnoub, Antoine E, and Robert A Weinberg. 2014. “Ras Oncogenes : Split Personalities.” *Nature Reviews Mollecular Cell Biology* 9(7): 517–31.

Khan, Abdul Q. et al. 2019. “RAS-Mediated Oncogenic Signaling Pathways in Human Malignancies.” *Seminars in Cancer Biology* 54(March 2018): 1–13.

Kotlyar, Max et al. 2014. “In Silico Prediction of Physical Protein Interactions and Characterization of Interactome Orphans.” *Nature Methods* 12(1): 79–84.

Lin, Wan Chen et al. 2014. “H-Ras Forms Dimers on Membrane Surfaces via a Protein - Protein Interface.” *Proceedings of the National Academy of Sciences of the United States of America* 111(8): 2996–3001.

Machleidt, Thomas et al. 2015. “NanoBRET-A Novel BRET Platform for the Analysis of Protein-Protein Interactions.” *ACS Chemical Biology* 10(8): 1797–1804.

Maffeis, Valeria, Lorenzo Nicolè, and Rocco Cappellesso. 2019. “RAS, Cellular Plasticity, and Tumor Budding in Colorectal Cancer.” *Frontiers in Oncology* 9:1255.

Matsuura, K., K. Canfield, W. Feng, and M. Kurokawa. 2016. “Metabolic Regulation of Apoptosis in Cancer.” *International Review of Cell and Molecular Biology* 327: 43–87.

Mishra, P. J. et al. 2010. “Dissection of RAS Downstream Pathways in Melanomagenesis: A Role for Ral in Transformation.” *Oncogene* 29(16): 2449–56.

Nunez Rodriguez, N. et al. 2006. “Characterization of R-Ras3/M-Ras Null Mice Reveals a Potential Role in Trophic Factor Signaling.” *Molecular and Cellular Biology* 26(19): 7145–54.

Nussinov, Ruth, Chung Jung Tsai, and Hyunbum Jang. 2018. “Oncogenic Ras Isoforms Signaling Specificity at the Membrane.” *Cancer Research* 78(3): 593–602.

O’Bryan, John P. 2019. “Pharmacological Targeting of RAS: Recent Success with Direct Inhibitors.” *Pharmacological Research* 139: 503–11.

Petschnigg, Julia et al. 2014. “The Mammalian-Membrane Two-Hybrid Assay (MaMTH) for Probing Membrane-Protein Interactions in Human Cells.” *Nature Methods* 11(5): 585–92.

Petschnigg, Julia, Jamie Snider, and Igor Stagljar. 2011. “Interactive Proteomics Research Technologies: Recent Applications and Advances.” *Current Opinion in Biotechnology* 22(1): 50–58.

Phizicky, E. M., and S. Fields. 1995. “Protein-Protein Interactions: Methods for Detection and Analysis.” *Microbiological Reviews* 59(1): 94–123.

Prior, Ian a, Paul D Lewis, and Carla Mattos. 2012. “UKPMC Funders Group UKPMC Funders Group Author Manuscript A Comprehensive Survey of Ras Mutations in Cancer.” *Cancer Research* 72(10): 2457–67.

Rojas, Ana Maria, Gloria Fuentes, Antonio Rausell, and Alfonso Valencia. 2012. “The Ras Protein Superfamily: Evolutionary Tree and Role of Conserved Amino Acids.” *Journal of Cell Biology* 196(2): 189–201.

Shah, Neel H., and Tom W. Muir. 2014. “Inteins: Nature’s Gift to Protein Chemists.” *Chemical Science* 5(2): 446–61.

Shimanshu, Dhirendra K., Dwight V. Nissley, and Frank McCormick. 2017. “RAS Proteins and Their Regulators in Human Disease.” *Cell* 170(1): 17–33.

Snider, Jamie et al. 2015. “Fundamentals of Protein Interaction Network Mapping.” *Molecular Systems Biology* 11(12): 848.

Stengel, Kristy R., and Yi Zheng. 2012. “Essential Role of Cdc42 in Ras-Induced Transformation Revealed by Gene Targeting” *PLoS ONE* 7(6): e37317.

Suter, Bernhard, Saranya Kittanakom, and Igor Stagljar. 2008. “Two-Hybrid Technologies in Proteomics Research.” *Current Opinion in Biotechnology* 19(4): 316–23.

Sutton, Margie N. et al. 2018. “RAS-Related GTPases DIRAS1 and DIRAS2 Induce Autophagic Cancer Cell Death and Are Required for Autophagy in Murine Ovarian Cancer Cells.” *Autophagy* 14(4): 637–53.

Terrell, Elizabeth M., and Deborah K. Morrison. 2019. “Ras-Mediated Activation of the Raf Family Kinases.” *Cold Spring Harbor Perspectives in Medicine* 9(1): a033746.

Wang, Ying et al. 2002. “The RAS Effector RIN1 Directly Competes with RAF and Is Regulated by 14-3-3 Proteins.” *Molecular and Cellular Biology* 22(3): 916–26.

Wennerberg, Krister, Kent L. Rossman, and Channing J. Der. 2005. “The Ras Superfamily at a Glance.” *Journal of Cell Science* 118(5): 843–46.

Wood, David W., and Julio A. Camarero. 2014. “Intein Applications: From Protein Purification and Labeling to Metabolic Control Methods.” *Journal of Biological Chemistry* 289(21): 14512–19.

Wu, Donghui et al. 2014. “Lsm2 and Lsm3 Bridge the Interaction of the Lsm1-7 Complex with Pat1 for Decapping Activation.” *Cell Research* 24(2): 233–46.

Yao, Zhong et al. 2020. “Split Intein-Mediated Protein Ligation for Detecting Protein-Protein Interactions and Their Inhibition.” *Nature Communications* 11(1): 1–14.

Yu, Hai Feng et al. 2012. “High RIN1 Expression Is Associated with Poor Prognosis in Patients with Gastric Adenocarcinoma.” *Tumor Biology* 33(5): 1557–63.

Zhang, Baolin, Jonathan Chernoff, and Yi Zheng. 1998. “Interaction of Rac1 with GTPase-Activating Proteins and Putative Effectors. A Comparison with Cdc42 and RhoA.” *Journal of Biological Chemistry* 273(15): 8776–82.

Curriculum vitae

Personal information

Name: Josipa Mateševac

Date of birth: 13.09.1995.

Education

2017. – 2020. Graduate Programme of Molecular Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia

2014. – 2017. Undergraduate programme of Molecular Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia

Experience

May 21st – December 12th 2019. International visiting graduate student at Department of Biochemistry, Faculty of Medicine, University of Toronto, Toronto, Canada

- Master thesis research and student internship in Stagljar Lab under supervision of Dr. Igor Štagljar, Prof.

March – May 2018. Laboratory skill training at Laboratory for Plant Molecular Biology, Department of Molecular Biology, Faculty of Science, University of Zagreb

- student internship under supervision of Dr. Nataša Bauer, Assoc. Prof.

December 7th – December 8th 2017. Participant at Practical Workshop "Introduction to Bioprocessing", BIOCentar, Zagreb, Croatia

February – May 2017. Laboratory skill training at Department of Animal Phisiology

- student internship under supervision of Dr. Nada Oršolić, Prof.

2018. Organizer at Night of Biology, Faculty of Science, University of Zagreb