HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2021 October 19.

Published in final edited form as:

Methods Mol Biol. 2021 ; 2293: 117–141. doi:10.1007/978-1-0716-1346-7_9.

Deconvolution of multiple Rab binding domains using the batch yeast-2 hybrid method DEEPN

Tabitha A. Peterson, **Robert C. Piper***

Molecular Physiology and Biophysics, Carver College of Medicine, University of Iowa, Iowa City, Iowa, 52242

Summary

A hallmark of functionally significant interactions between Rab proteins and their targets is whether that binding depends on the type of nucleotide bound to the Rab GTPase. A system that can directly compare those sets of interactions mediated by a Rab in its GTP-bound conformation *vs* its GDP bound conformation can provide immediate with regard to the nucleotide-binding status of the Rab would provide a direct route to finding biologically relevant partners. Comprehensive large scale yeast 2-hybrid assays allow a potential method to compare one interactome against another provided that the same set of interacting partners is interrogated between samples. Here we describe the use of such a yeast-2 hybrid system that lends itself towards comparing pairs of Rab mutants, locked in either their GTP or GDP conformation. Importantly, using a complex library of protein fragments as potential binding ('prey') partners, identification of interacting proteins as well as the domain(s) mediating those interactions can be determined using a series of sequence analyses and binary validation experiments.

Keywords

Next generation sequencing; GTP; GDP; Rab GTPase

1. Introduction

Rab GTPases are protein interaction switches that are acutely regulated by toggling between an inactive GDP-bound form and an active GTP-bound form. Exchanging GDP for GTP causes a conformational change allowing Rab GTPases to bind specific partner proteins or effectors, which then assemble into functional complexes at discrete locations in the cell according to the regulatory mechanisms that govern the distribution of Rab proteins [1]. Knowing the various protein interactions in which Rab proteins participate in vivo is a major requirement for understanding how they function. Importantly, many Rab proteins have been found to have multiple partners that likely engage them under different cellular states and locations, emphasizing the critical goal of identifying specific Rab-interacting proteins. Perhaps the most important group of Rab-interacting proteins are those that bind Rabs in a nucleotide-specific manner, since those partners would be best implicated in mediating a Rab-regulated function.

*Correspondence: Robert-Piper@uiowa.edu.

Rab GTPases fit within a larger family of small molecular weight GTPases (~25 kDa) consisting of 6 central β-sheets coordinated 5 α-helices that provide a well conserved guanine-nucleotide binding fold [2]. Binding different guanine nucleotides in vivo causes substantial conformational changes in the 'switch' regions of Rab GTPases. Rab proteins have variable C-terminal tails that are covalently lipidated. These C-terminal tails also undergo a conformational change upon GTP-binding by disengaging their guanine dissociation inhibitor (GDI) and inserting into a lipid bilayer [3]. Thus, a nucleotide-specific Rab interacting protein would be expected to use these features to achieve specific binding.

An important tool that has driven functional studies of Rab GTPases as well as biochemical approaches aimed at finding Rab effectors is based on conserved mutations near the nucleotide binding region. Mutation of Q61 to L in Ras, which blocks GTPase hydrolysis, was originally found as an oncogenic form of Ras [4]. This position is highly conserved amongst Ras-family members, allowing a similar mutation to be introduced into Rab GTPases to lock them into a GTP-bound conformation [5,6]. Similarly, mutation of S34 to Nlocks Ras into a GDP-bound conformation, which can bias interactions towards proteins such as guanine-nucleotide exchange factors [7].

Mutant Rabs (GTP-bound 'Q>L' point mutants and GDP-bound 'S or T>N' point mutants) have been used in numerous biochemical and genetic approaches to find interacting partners. One particular approach is the yeast 2-hybrid (Y2H) system, in which fusion proteins containing a Rab (aka 'bait') and its interacting partner (aka 'prey'), bind to form a functional transcription factor that drives yeast growth [8-10]. The availability of sets of Rab fusion 'bait' proteins configured in a GTP-bound (Q>L) and GDP-bound (S or T>N) conformation have made it possible to use a Y2H based matrix to score a particular Rab interacting 'prey' protein for its nucleotide and Rab specificity across the whole family of Rab GTPases [11]. However, the limited scale and capacity a traditional Y2H screen has in sampling potential interacting partners within a given 'prey' cDNA library has prevented the use of these Rab GTPase 'bait' sets to perform comprehensive denovo Y2H screens for Rabs and their nucleotide specific interactions. Rather, a fiding or not finding an interacting partner in a typical Y2H screen is entirely stochastic and comes without the ability to make quantitative or statistical statements about whether potential interacting candidates do or do not interact with a given Rab GTPase in a nucleotide-specific manner [12].

The advent of deep sequencing has allowed this limitation to be addressed since the composition of a 'prey' library can be delineated and quantified from yeast populations grown under conditions that do and do not select for a positive Y2H interaction. We have developed one of these approaches called DEEPN (Dynamic Enrichment for the Evaluation of Protein Networks), which uses selection for Y2H interactions at a modest stringency in batch using liquid cultures [12]. The abundance of every prey-encoding plasmid is determined by deep sequencing and bioinformatic analysis using a dedicated stand-alone software package designed specifically for DEEPN datasets. For the identification of proteins that differentially interact with Rab GTPases, DEEPN offers two main advantages: One is that deep sequencing can confirm that the entire composition of the prey library population within yeast carrying one particular Rab 'bait' is the same as that of another particular Rab bait [13]. This allows for a direct comparison between what components

one Rab bait interacts with vs a different Rab bait. The DEEPN software also identifies the junctions that connect the prey insert with the expression plasmid, allowing one to computationally determine whether a particular prey gene fragment is in the proper translational reading frame and to determine what portion of a given reading frame encodes the interacting protein fragment [14]. This latter feature is especially useful when using a highly complex library of prey plasmids which contains several different open-reading frame fragments since interacting domains can be quickly delineated.

These features offer distinct advantages when searching for Rab interactions. In systematic matrix-driven Y2H screens, Rab proteins have limited representation because their biologically relevant interactions are driven by nucleotide binding and thus, they need to be presented in particular nucleotide-bound conformations that are not represented in genome-wide libraries [15-18]. In both matrix-driven Y2H screens and affinity-isolation/mass-spectrometry experiments, full-length proteins are analyzed rather than protein fragments. Thus, Rab interacting domains may be hidden within the context of a larger proteins. Moreover, there is not an immediate indication where in a protein a Rab-interacting domain could lie without interrogating multiple protein fragments later, whereas DEEPN interrogates several gene fragments to yield comparative interacting data for each fragment as an integrated part of the workflow.

Here we demonstrate how to analyze a DEEPN dataset for interactors that differentiate between distinct Rab proteins and their nucleotide conformation. Several proteins are known to bind multiple Rab proteins, often within distinct domains. Here we show how DEEPN Y2H data can identify subdomains with such proteins to yield a medium resolution interaction map and how computational reconstruction of plasmids that yield a positive Y2H interaction can inform downstream validation and hypothesis testing.

Major Outcomes and Possible uses.

The genome encodes a plethora of Rab proteins, yet only the function of a handful are largely known. Even for these few, how their functions are executed and the set of interacting effector proteins required for that execution remain underdetermined. Rab GTPases work by interacting with other proteins in a manner dependent on their bound nucleotide. To understand their function requires finding those nucleotide specific interactions and characterizing the structural basis of them enough to alter their Rab-interacting motifs and determine how that interaction is relevant to cellular process. The methods described here harness the inexpensive capabilities of high-throughput sequencing and the well-established yeast 2-hybrid protein interaction reporter system to not only query large sets of potential interacting proteins, but statistically determine whether each candidate has specificity for one nucleotide-bound state *vs* another. Moreover, with a dense library of open-reading frame fragments, one can use computational methods to extrapolate where in a given protein a Rab interacting domain is located. This method can be expanded to not only determine what interactors are dependent on a particular nucleotide-bound conformation but also which ones may be sensitive to disease-causing mutations within Rab proteins [19,20], thus offering a pathway to discover the biochemical basis for how Rab mutations cause disease.

2. Materials

2.1 Rab Expression constructs.

1. Rab fusion constructs in pTEF-GBD, encoding the Gal4 DNA binding domain. Rab GTPases with mutations that favor a GTP-bound conformation and a GDP-bound conformation. Rab open-reading frames are codon optimized to the *S. cerevisiae* and the isoprenylation consensus sequence (CAAX box, [21]) minimally mutated to avoid lipidation (Figure 1).

2.2 Data Processing

1. Illumina sequence datasets from a differential DEEPN Y2H screen. This includes sequence data from plasmid populations grown under non-selective conditions or conditions that select for a positive Y2H interaction (eg. media lacking Histidine) using the Gal4-DNA-binding bait vector alone or within a fusion construct with Rab mutants locked in their GDP and GTP bound state.
2. DEEPN software programs, including DEEPN, Stat_Maker, and Mapster. (https://github.com/emptyewer). Figure 2
3. Macintosh computer for data processing. (minimum requirements: OS 10.10 or above, quadcore Intel i3 processor, 8 Gb memory, 4 Tb hard disk drive)

3 Plasmid Reconstruction

1. Processed data files from DEEPN output (Note A).
2. Plasmid: pPL6343 – pGal4AD
3. Oligonucleotides to amplify gene fragments and clone into pGal4AD (Note B).

2.4 Validation

1. Plasmids: pPL6229 - pTEF-GBD, pPL6222 - pTEF*-GBD, pPL6343 – pGal4AD
2. Yeast Strains; PJ69-4A and PLY5725
3. Glucose Solution (50% w/v): for 500 mL add 220 mL milliQ water to a 600 mL beaker. Add in 250 g D-(+)-Glucose slowly till all dissolved. Add milliQ water up to 500 mL. Filter Sterilize with a 0.22 μm PES filter and store at room temperature.

NOTE A. DEEPN creates a number of datasets and files that allow easy analysis of sequence data to discern the nature of particular Y2H interactions. One of those datasets is viewed with the module Read_Depth, which displays the number of reads obtained along each transcript or ORF. This analysis will indicate which gene fragment(s) were present in each sample and indicate what portion of the gene sustained a Y2H interaction under selective conditions. The second dataset is a description and count of the junctions that join the Gal4-transcriptional activation domain to a particular gene of interest. Here, the 5' and 3' ends of gene fragments that encode interacting portions are collated and counted, and the data summary is viewed using Blast_query. Finally, a *.junctions file is created that lists all of the reads that cover the junctions between the Gal4-AD 'prey' vector and the gene inserts. From this file, one can search and retrieve the exact sequence of a given junction in order to computationally reconstruct the exact prey plasmid that supported a positive Y2H enrichment in the DEEPN experiment.

Note B. The process of plasmid reconstruction uses the processed DEEPN files to computationally reconstruct 'prey' plasmids that sustained a Y2H interaction. Once this is done, oligonucleotides are designed to amplify these fragments and clone these fragments into pGal4AD to yield a plasmid that can then be used to validate the Y2H interaction that was indicated by the DEEPN data.

4. Yeast Nitrogen Base (YNB): for 500 mL, add 400 mL milliQ water to a 600 mL beaker. Add 3.35 g Yeast Nitrogen Base without amino acids. Mix well and add milliQ water up to 480 mL. Autoclave and allow to cool till warm to touch. Add 20 mL 50% glucose. Swirl to mix. Store at room temperature.

5. Transformation Buffer: 1 M sorbitol, 1 M lithium acetate dihydrate, 10 mM tris pH 7.6, 0.5 mM EDTA, 0.2 mM calcium chloride. For 500 mL, add 300 mL milliQ water to a 600 mL beaker. Add 91.1 g Sorbitol, 10 mL 0.5 M Tris pH 7.6, 5.1 g lithium acetate dihydrate, 500 μl 0.5 M EDTA pH 8.0, 100 μl of 1 M Calcium Chloride. Mix well and add milliQ water up to 500 mL. Filter Sterilize with a 0.22 μm PES filter and store at room temperature.

6. Yeast extract peptone dextrose (YPD) plates: For 25 plates, weigh out and dissolve 10 g Peptone and 5 g Yeast Extract in 400 mL MilliQ water in a 600 mL beaker. Mix well and add milliQ water up to 480 mL. Add 7.5 g agar. Autoclave and allow to cool till warm to touch. Using a pipette, add 20 mL of 50% glucose. Mix well by swirling. By pipette pour 20 mL into a series of 100 mm plates.

7. Complete synthetic minimal media (CSM)-Trp plates: For 25 plates, weigh out and dissolve 3.35 g Yeast Nitrogen Base without amino acids in 400 mL MilliQ water in a 600 mL beaker. Mix well and add milliQ water up to 480 mL. Add 0.35 g -Trp-Met dropout mix, 10 mg of methionine, and 7.5 g agar. Autoclave and allow to cool till warm to touch. Using a pipette, add 20 mL of 50% glucose. Mix well by swirling. By pipette pour 20 mL into a series of 100 mm plates.

8. CSM-Leu plates: For 25 plates, weigh out and dissolve 5.025 g Yeast Nitrogen Base without amino acids in 400 mL MilliQ water in a 600 mL beaker. Mix well and add milliQ water up to 470 mL. Add 0.5025 g -Leu-Met dropout mix and 7.5 g agar. Autoclave and allow to cool till warm to touch. Using a pipette, add 30 mL of 50% glucose. Mix well by swirling. By pipette pour 20 mL into a series of 100 mm plates.

9. CSM-Leu-Trp- plates: For 25 plates, weigh out and dissolve 5.025 g Yeast Nitrogen Base without amino acids in 400 mL MilliQ water in a 600 mL beaker. Mix well and add milliQ water up to 470 mL. Add 0.5025 g -Trp-Leu+40Ade dropout mix, 120 mg adenine and 7.5 g agar. Autoclave and allow to cool till warm to touch. Using a pipette, add 30 mL of 50% glucose. Mix well by swirling. By pipette pour 20 mL into a series of 100 mm plates.

10. CSM-Leu-Trp-His plates: For 25 plates, weigh out and dissolve 5.025 g Yeast Nitrogen Base without amino acids in 400 mL MilliQ water in a 600 mL beaker. Mix well and add milliQ water up to 470 mL. Add 0.4875 g -Trp-Leu-His+40Ade dropout mix, 120 mg adenine and 7.5 g agar. Autoclave and allow to cool till warm to touch. Using a pipette, add 30 ml of 50% glucose. Mix well by swirling. By pipette pour 20 mL into a series of 100 mm plates.

11. Plate Reader or spectrophotometer to read OD_{600}

12. Sterile sticks and 1.7 mL microcentrifuge tubes

3. Methods

Goal:

The statistical software module in DEEPN assesses whether a particular gene, or gene fragment is enriched in the population of plasmids in the Y2H culture placed under conditions that select for a positive Y2H interaction. Typically, the number of sequence reads for a gene that has a true positive Y2H interaction are far higher in the selected population than the non-selective population. For the interaction to be specific for the bait protein of interest, enrichment or increase in the number of sequence reads should be far less for a control bait (eg. the Y2H vector alone that only expresses the Gal4-DNA binding domain alone). Discerning what is likely to be a true and specific enrichment above the noise in the experiment is provided by the DEEPN software package using a built-in statistical model. The DEEPN software also determines whether the enriched gene/ORF prey inserts are likely in the correct translational reading frame, or not. These two analyses then serve as the first step in identifying likely Y2H-interactors specific to one bait, but not another. A second aspect of informatic processing takes advantage of the fact that prey libraries, such as cDNA libraries, contain fragments of open-reading frames and rarely complete full-length clones. By determining the exact 5' and 3' ends of the gene fragment that is enriched, bioinformatic analysis can reveal what domains or parts of a particular interacting protein is sufficient for interaction. In addition, by monitoring what gene fragments are not enriched upon Y2H selection, DEEPN can also indicate what domains are not required for Y2H interaction. The more fragments that are in the prey library, the more granular DEEPN analysis can become to determine the relevant interaction domain. Previously, we have made high-density libraries from genomic DNA from S. cerevisiae, an organism with few introns allowing genomic DNA fragments to provide a useful array of open-reading frame fragments. We also have recently made a prey library of open-reading-frame fragments derived from the human ORFeome v8.1. These libraries allow for better sampling of a given ORF, which in turn, can provide a higher resolution analysis of interacting and non-interacting ORF fragments through informatic analysis of the sequencing data (Figure 3).

We obtained DEEPN sequence data from screening mutant Rab bait proteins locked in a GDP as well as GTP-locked conformation using a human ORFeome fragment prey library. Following are the steps to process sequence data, identify likely interacting proteins, discern the interacting portion of each candidate from sequence data, and reconstruct the particular library constituent plasmids to perform validation experiments to confirm deduced Y2H interactions.

3.1 Data Processing

1. Obtain Illumina 150 bp paired-end reads from PCR amplicons amplified from DNA isolated from different yeast populations grown under selective and non-selective conditions. PCR amplicons are randomly sheared to ~250-450 bp fragments, modified by bar-coded Illumina sequencing primers prior to sequencing in a flow cell. Between 8-20 million reads/sample is adequate for analysis.

2. Map reads to relevant genome. For screens using the human ORFeome library, map Illumina reads to the human genome hg38. If alternate prey libraries have been used, then use mm10 for the reference mouse genome, or Saccer3 for the reference yeast genome. The output must be in the form of a .sam file (sequence alignment map). Programs such as Tophat2 or HiSat2 that can accommodate mapping mRNA sequence data to a complete genome work comparably. For an easy-to-use interface, mapping can be accomplished using the Mapster program included in the suite of DEEPN bioinformatics software.

3. Process the .sam files with the 'Gene Count' module within the DEEPN software. Once complete, then run the 'Junction Make' module one time to find the 5' ends or the gene fragment insert, and a second time to find the 3' end. For the human ORFeome library, use the following junction sequences to find the 5' and 3' ends of the gene fragment inserts, respectively: CCTCTGCGAGTGGTGGCAACTCTGTGGCCGGCCCAGCCGGCCATGTCAGC, CATGGCCCGGGAGGCCTAGATGAATAATAGAAGACGGGAGACACTAGCAC.

4. Identify in-frame interacting partners with StatMaker. Statmaker results display statistical rankings for the differential interaction of a given prey with vector alone, a Rab protein in its GTP-bound conformation, and a Rab protein in its GDP-bound conformation (Figure 4). This 3-way comparison assigns a probability for finding gene products that specifically interact with only the GTP- or GDP-bound form. Filter the specific interactors to identify those that are in the correct translational reading frame and within the open-reading frame of interest.

5. Use the Blast Query module to analyze the set of 5' and 3' junctions to find what fragments of a given interacting gene is sufficient to yield a positive Y2H interaction. The data that BlastQuery uses is a collection of all the sequences that flank the 5' and 3' ends of each insert. These are subjected to a Blastp search to determine what gene they correspond to and what portion of the open-reading frame is encompassed in the fragment (Figure 5). Click save to CSV to export data to a spreadsheet file located in QueryResults.

6. Use the ReadDepth module to find the number of sequence reads that were obtained across the prey gene of interest (Figure 6). Use the gene identifier (NM_*) that corresponds to the gene of interest and also select the correct sample from the selected population that contains the enriched gene of interest. Adjust the interval of sequence match to 20-30 bp to increase resolution of the analysis. Click the export button to write a spreadsheet file with the ReadDepth results.

7. Collate the data from the BlastQuery tables and the exported data from ReadDepth (Figure 7).

3.2 Plasmid Reconstruction

Goal.—There are two methods to reconstruct a prey plasmid that encompasses the gene fragment within the original Y2H library that produced the positive Y2H interaction that led to the amplification of the plasmid during growth under selective conditions. One is approximate, in that the known stretch of residues within the gene of interest are determined and cloned into the prey plasmid in frame with the Gal4 Transcriptional Activation domain. The other is more precise and relies on finding the sequence reads that correspond to the 5' and 3' junctions of the gene fragment that yields a positive Y2H interaction. These are found manually by searching through the list of junctions compiled in the *.junctions.txt file.

1. Approximate method. The data in Figure 7 indicate a single gene fragment of interest if a single region is indicated in ReadDepth and flanked by one high-abundance 5' junction and one high-abundance 3' junction (Note C). The junction data from BlastQuery will define the beginning and end codons of the gene region of interest. Using the strategy in Figure 8, construct a DNA map in which the span of DNA encoding the residues of interest within the prey gene of interest are inserted in-frame downstream of the Gal4 transcriptional activation domain. This construct should also include a stop codon after the gene fragment.

2. Precise method. The data in Figure 7 show the 5' and 3' junction data that correspond to the gene of interest. However, the fragment of interest may be flanked by other nucleotides as a result of the way it was cloned into the vector. Finding an actual sequence that encompasses those junctions allows for the precise reconstitution of the likely library plasmid that gave rise to the Y2H positive interaction. The precise junction of the Gal4-AD vector with the 5' and 3' ends of the gene fragment may affect the linker region preceding the fragment and the C-terminal region before translation is terminated. The data required from BlastQuery on the gene fragment of interest is the NM_* code of the gene of interest, the Position, and the q-start (Figure 8). The files needed are the *sample*.blast.txt and the corresponding *sample*.junctions.txt file. Open the *.blast.txt file. Search for the NM_* code of the gene of interest and find a Blast match that also contains the q-Start and Position numbers that match the q-start and s. start values, respectively. This line also contains the original sequence read identification number found within the original FASTQ file of the Illumina dataset. Copy the sequence ID and search for it in the corresponding *junctions.txt file. Each line within the *.junctions.txt file will contain the original sequence read, the sequence downstream of the flanking sequence used to find each junction, and the amino-acid sequence of the translated flanking region. Using the sequence from both the 5' and 3' *.junctions.txt file, construct a DNA map in which the span of DNA encoding the residues of interest

Note C. Finding the gene fragment of interest to test in further analysis is straightforward when a single fragment was amplified during selection as indicated by a single 5' and 3' junctions that flank a single region visualized by ReadDepth (Figure 6). However, multiple fragments may also be amplified during selection, making it less clear where fragment boarders are and whether multiple Rab-binding domains are present or whether overlapping fragments simply share a single Rab-interacting region. An example is shown in Figure 7, where Rab43-GTP interacts with a single fragment of SSX2IP whereas Rab5-GDP interacts with two regions that can be seen using the ReadDepth data.

within the prey gene of interest are inserted in-frame downstream of the Gal4 transcriptional activation domain.

3. Making new Gal4-AD 'prey' plasmid. From the electronic map generated in 3.2.1 or 3.2.2, design oligonucleotide primers that have 22-24 bp of homology with the gene insert of interest and that have 15-20 bp of homology with the regions of pGal4AD that flank the insertion site. Using those primers, amplify the fragment of interest from the genomic DNA previously isolated from the genomic DNA sample of the relevant yeast population that was used to generate the Illumina sequence data. These samples should be greatly enriched for the gene fragment of interest due to their amplification during growth under selection conditions. Once the PCR product is checked for the correct size, use the Gibson method to recombine the fragment into pGal4AD double cut with SfiI. Once sequenced and verified, these prey gene fragment AD plasmids can be used in binary Y2H validation experiments (Figure 9).

3.3 Validation

Goal: The goal of this step is to perform a series of binary assays in a traditional format to directly compare Y2H interactions across multiple baits including empty vector alone. This serves to validate the possibly Y2H interactions that computational analysis predicts happened during the batch selection process within the original DEEPN experiment. When used with a variety of bait proteins (such as a large set of Rab protein mutants locked in ether GTP- or GDP-bound conformations), this analysis can reveal one of 3 different outcomes. One is that a given prey fragment has the same interactions with the bait GTPases as found by the batch DEEPN analysis in that a subset of Rab proteins interact while others, including vector alone, do not. Another is that more Rab proteins and possibly vector alone are found to produce a Y2H interaction that was not detected computationally be DEEPN. The reason for this is that each bait in a DEEPN run can interact with a different set of prey proteins. And while some of those preys may also interact with other bait proteins, the preys may not all be enriched to the same extent across samples. The likely reason is that differential interactions across two baits will create different enriched populations that may crowd out detection of one authentic interacting prey in one sample but not the other. This emphasizes the need to perform the binary validation studies described here. A third possible outcome is that binary assays fail to show a Y2H interaction with a prey fragment computationally identified by DEEPN (Figure 10).

1. Lithium Acetate transform vector only pTEF-GBD or pTEF*-GBD and bait pTEF-GBD or pTEF*-GBD constructs into PJ69-4A and newly constructed prey AD validation constructs along with vector only pPL6343 into PLY5725. Spread transformations onto CSM-Trp and CSM-Leu-Met plates respectively (*see* Note D). Incubate plates in a 30 °C incubator for 2-3 days or the appearance of colonies.

Note D. We found that by making the competent cells within 3 months of transformation as well as not freeze thawing these cells allows for more efficient transformations. Plates made within the last week also allow for better transformations efficiencies.

2. Using a sterile stick, make quarter size patches of about 5-6 single colonies of each transformation onto the appropriate selection plate. Incubate plates 30 °C 1 day. (*see* Note E)

3. Using sterile technique, set-up a mating reaction of each pTEF-GBD or pTEF*-GBD vector only and bait construct with Library AD validation constructs on a YPD plate. Briefly, take a matchhead of Library AD validation construct, dab it at the top center of the plate, and streak it down the YPD plate. Using a sterile stick, take a matchhead of pTEF-GBD or pTEF*-GBD vector only and bait constructs, dab to the left of the already streaked Library AD validation construct, and streak it to the right through the Library AD validation construct. Do this for all library AD validation constructs/pTEF-GBD or pTEF*-GBD constructs needing validated. Incubate plates in a 30 °C incubator for 1 day.

4. Using sterile technique, streak the mated yeast onto CSM-Leu-Trp plates to select for a diploid population. Incubate plates in a 30 °C incubator for 1 day.

5. For each mating reaction, label a sterile 1.7 ml microcentrifuge tube with construct names. Pipette 500 μl of Yeast Nitrogen Base into each tube.

6. Using sterile techniques and a sterile stick, obtain a small matchhead amount of diploid yeast cells and twirl it into the correct 1.7 ml microcentrifuge tube. Do this for all diploid possibilities in a given set. (*see* Note F)

7. Make sure samples are mixed well and pipette 140 μl of samples into a 96 well plate. Save the remaining yeast diploid suspension in each tube for further dilution, which will depend on the values obtained from the plate reader.

8. Measure the OD_{600} of the cells. Record the OD. (*see* Note G)

9. Based on the values from the plate reader, make up a 1:10 dilution series using the YNB starting with an initial concentration of diploids at 0.5 OD and continuing until you have a total of 6 different concentrations. This can be done in either a 96 well plate or in sterile 1.7 ml microcentrifuge tubes. (*see* Note H)

Note E. The ability to analyze the validation result in confidence hinges on knowing that the bait BD constructs as well as the newly constructed Library AD constructs express protein, therefore it is imperative that the expression of the constructs be checked and confirmed by western blot analysis before continuing.

Note F. Before doing this step, it is imperative that you think through what combinations of diploids you plan on validating on a given set of CSM-Leu-Trp and CSM-Leu-Trp-His plates. For example, you may want to compare one newly constructed library AD construct across multiple Rab GTPases or you may want to compare a few different library AD construct regions across the same set of Rab GTPases in a GDP vs GTP specific manner. You need this planned out for a couple reasons. One, we have seen growth variability between plates so if you want to be able to make direct comparisons, they really should either be all on the same plate or you have a representative control (pTEF-GBD and Library AD construct) on each set of plates. Second, even though diploids are happier in YNB than sterile water, you do not want to leave them sit there for hours while you are setting up validation of other diploid combinations. You can always complete the protocol from here with one set and return to complete for iterative sets.

Note G. Make sure the cells are properly suspended prior to checking the OD_{600} reading as cells that have settled will give a false reading. Also, it is imperative that if a plate reader is used to determine the OD_{600}, the reading is multiplied by the correction efficient that allows for the same OD_{600} reading that would be achieved with a 1 mL solution 1 cm pathlength.

Note H. In order to make accurate dilutions, make sure to mix the suspension prior to using it to make the next 1:10 dilution set in the series. If this is not mixed properly for dilution here, you will not make an accurate dilution which will make your final result less interpretable. This will be seen by the growth or lack of growth when you compare constructs on the CSM-Trp-Leu plate at the end of the validation. You should see the same growth or ability to grow on CSM-Trp-Leu plates at the end of the validation.

10. Obtain a CSM–Trp-Leu and a CSM–Trp-Leu-His plate, label/mark the top as top, as well as label it accordingly to the combinations of diploids that will be spotted. (*see* Note I)

11. Using a pipette, spot 5ul, left to right and least to most concentrated respectively in one row. (*see* Note J)

12. When all the spots have dried, place all plates in the 30 °C incubator for 3 days.

13. After 3 days, pull all the plates out of the 30 °C incubator and scan for archive and interpretation. (*see* Note K, L).

References

1. Pfeffer SR (2013) Rab GTPase regulation of membrane identity. Curr Opin Cell Biol 25 (4):414–419. doi:10.1016/j.ceb.2013.04.002 [PubMed: 23639309]
2. Vetter IR, Wittinghofer A (2001) The guanine nucleotide-binding switch in three dimensions. Science 294 (5545):1299–1304. doi:10.1126/science.1062023 [PubMed: 11701921]

Note I. Make the CSM-Trp-Leu and CSM-Trp-Leu-His plates the day before spotting. Fresh plates will allow the diploids the best opportunity to grow and give an interpretable result. Plates that are even three days old here, will cause a slow growth or what seems to be a stunted growth ability for diploids even on the CSM-Trp-Leu plates alone.

Note J. It is imperative that you mix the concentrations efficiently right before spotting. This will further ensure that any growth difference seen in diploid comparison sets are accurate and thus interpretable. As mentioned above at the end you should see comparable growth on the CSM-Trp-Leu plates this will give you confidence that any difference seen on the CSM-Trp-Leu-His plates is indeed real. Improper mixing leads to an inability to compare sets and the need to repeat.

Note K. When interpreting your plates, you want to make sure that the growth is similar across all constructs on the CSM-Trp-Leu plate. If one pair of diploids did not grow as well as another there are a couple reasons for that. One, of which is highlighted throughout the notes, is the potential that at one step or another the diploids were not mixed well enough leading to the wrong concentration of diploids being spotted. The other is a little more complicated in the regard that the expression level of the construct is too high and is thus toxic to the cell and a growth defect in the form of cell viability is observed. If expression is in fact too high, you can switch from pTEF-GBD vector to pTEF*-GBD that will allow for a lower protein expression level and potentially help with cell viability. If the growth is similar of all diploids across the CSM-Trp-Leu plates than a direct interpretation of growth on the CSM-Trp-Leu-His plates can be made. For example, if there is no growth observed on the CSM-Trp-Leu-His plates, then one would conclude that there is no direct two-hybrid interaction between the Rab GTPase(s) of interest with a given library AD validation construct. You may also observe growth for both the GTP and GDP nucleotide bound states of the Rab GTPase with a given library validation construct of which you could conclude that indeed there is a two-hybrid interaction, however it is nucleotide non-specific. Lastly, you may observe that there is growth on only the GTP- or GDP-bound nucleotide state of which you would conclude that the two-hybrid interaction is nucleotide specific for the given pair. If you need to further tease apart the strength of the two-hybrid interaction between different pairs of diploids of the Rab(s)/Validation AD library constructs it is common to add 0.1mM to 10mM 3-aminotriazole to the CSM-Trp-Leu-His plates before pouring them. With the system described here, analysis of Rab proteins with the yeast strains and plasmids described, the use of 3-aminotriazole is not necessary.

Note L. By uncovering Rab interacting protein fragments/domains with a DEEPN screen and then validating those putative interactions with binary Y2H assays over a broad range of Rab GTPases, the Rab-specificity and nucleotide-bound conformational specificity of interactions can be determined. These can help provide immediate insight into how proteins that bind multiple Rab proteins may do this, either by having a single Rab domain that can interact with several different Rab GTPases, or multiple Rab binding sites each with a more exclusive specificity for Rabs. Many proteins that interact with multiple Rab proteins have been described such as RabEP1 [22], Optineurin [23,24], and MICALL1 [25,26], all of which have been identified by the approach above to contain multiple Rab-binding domains that each have different subsets of Rab specificity. In the example used here, the protein SSX2IP (Synovial Sarcoma X breakpoint 2 Interacting Protein) was found to have 2 Rab-interacting regions as validated by binary Y2H interactions. One binds Rab5A but only in its GDP-bound conformation, indicating this may be a region that could help SSX2IP work as a nucleotide-exchange factor of Rab5A. A different region binds Rab43, but only in its GTP-bound conformation, indicating that SSX2IP is a specific Rab43 effector. SSX2IP has known roles in centriole assembly and cilia biogenesis [27-29]. Little is known about Rab43, however, one role is to promote trafficking from the ER to the Golgi [30,31]. Rab5 is known to play a large role in endosomal trafficking, but it also plays a role in maturation of the spindle pole, providing a possible function for an interaction with SSX2IP [32]. Figure 11 summarizes the differential interactions of Rab43 and Rab5A with SSX2IP. Interestingly, the region that interacts with Rab43-GDP does not appear to do so in the context of full-length protein containing the region that interacts with Rab5A-GTP, suggesting that the binding of one Rab may regulate the binding of another. This suggests two models. One is a Rab cascade whereby Rab5-GTP could bind to SSX2IP and activate its nucleotide exchange activity for Rab43, which would coordinate activities associated with Rab5 with those of Rab43 [33]. Another is that SSX2IP provides a mechanism to localize Rab5 exchange activity to Rab43 membrane compartments or Rab43-containing complexes when Rab43 is activated by GTP binding.

3. Ullrich O, Stenmark H, Alexandrov K, Huber LA, Kaibuchi K, Sasaki T, Takai Y, Zerial M (1993) Rab GDP dissociation inhibitor as a general regulator for the membrane association of rab proteins. J Biol Chem 268 (24):18143–18150 [PubMed: 8349690]
4. Der CJ, Finkel T, Cooper GM (1986) Biological and biochemical properties of human rasH genes mutated at codon 61. Cell 44 (1):167–176. doi:10.1016/0092-8674(86)90495-2 [PubMed: 3510078]
5. Li G, Stahl PD (1993) Structure-function relationship of the small GTPase rab5. J Biol Chem 268 (32):24475–24480 [PubMed: 8226999]
6. Stenmark H, Parton RG, Steele-Mortimer O, Lutcke A, Gruenberg J, Zerial M (1994) Inhibition of rab5 GTPase activity stimulates membrane fusion in endocytosis. EMBO J 13 (6):1287–1296 [PubMed: 8137813]
7. Feig LA, Cooper GM (1988) Relationship among guanine nucleotide exchange, GTP hydrolysis, and transforming potential of mutated ras proteins. Mol Cell Biol 8 (6):2472–2478. doi:10.1128/mcb.8.6.2472 [PubMed: 3043178]
8. Fukuda M (2010) How can mammalian Rab small GTPases be comprehensively analyzed?: Development of new tools to comprehensively analyze mammalian Rabs in membrane traffic. Histol Histopathol 25 (11):1473–1480. doi:10.14670/HH-25.1473 [PubMed: 20865669]
9. Kail M, Barnekow A (2008) Identification and characterization of interacting partners of Rab GTPases by yeast two-hybrid analyses. Methods Mol Biol 440:111–125. doi:10.1007/978-1-59745-178-9_9 [PubMed: 18369941]
10. Langemeyer L, Barr FA (2012) Analysis of rab GTPases. Curr Protoc Cell Biol Chapter 15:Unit 15 18. doi:10.1002/0471143030.cb1518s57 [PubMed: 23208545]
11. Fuchs E, Haas AK, Spooner RA, Yoshimura S, Lord JM, Barr FA (2007) Specific Rab GTPase-activating proteins define the Shiga toxin and epidermal growth factor uptake pathways. J Cell Biol 177 (6):1133–1143. doi:10.1083/jcb.200612068 [PubMed: 17562788]
12. Pashkova N, Peterson TA, Krishnamani V, Breheny P, Stamnes M, Piper RC (2016) DEEPN as an Approach for Batch Processing of Yeast 2-Hybrid Interactions. Cell Rep 17 (1):303–315. doi:10.1016/j.celrep.2016.08.095 [PubMed: 27681439]
13. Peterson TA, Stamnes MA, Piper RC (2018) A Yeast 2-Hybrid Screen in Batch to Compare Protein Interactions. J Vis Exp (136). doi:10.3791/57801
14. Krishnamani V, Peterson tA, Piper RC, Stamnes MA (2018) Informatic Analysis of Sequence Data from Batch Yeast 2-Hybrid Screens. J Vis Exp (136). doi:10.3791/57802
15. Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, Brignall R, Cafarelli T, Campos-Laborie FJ, Charloteaux B, Choi D, Cote AG, Daley M, Deimling S, Desbuleux A, Dricot A, Gebbia M, Hardy MF, Kishore N, Knapp JJ, Kovacs IA, Lemmens I, Mee MW, Mellor JC, Pollis C, Pons C, Richardson AD, Schlabach S, Teeking B, Yadav A, Babor M, Balcha D, Basha O, Bowman-Colin C, Chin SF, Choi SG, Colabella C, Coppin G, D'Amata C, De Ridder D, De Rouck S, Duran-Frigola M, Ennajdaoui H, Goebels F, Goehring L, Gopal A, Haddad G, Hatchi E, Helmy M, Jacob Y, Kassa Y, Landini S, Li R, van Lieshout N, MacWilliams A, Markey D, Paulson JN, Rangarajan S, Rasla J, Rayhan A, Rolland T, San-Miguel A, Shen Y, Sheykhkarimli D, Sheynkman GM, Simonovsky E, Tasan M, Tejeda A, Tropepe V, Twizere JC, Wang Y, Weatheritt RJ, Weile J, Xia Y, Yang X, Yeger-Lotem E, Zhong Q, Aloy P, Bader GD, De Las Rivas J, Gaudet S, Hao T, Rak J, Tavernier J, Hill DE, Vidal M, Roth FP, Calderwood MA (2020) A reference map of the human binary protein interactome. Nature 580 (7803):402–408. doi:10.1038/s41586-020-2188-x [PubMed: 32296183]
16. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R, Kamburov A, Ghiassian SD, Yang X, Ghamsari L, Balcha D, Begg BE, Braun P, Brehme M, Broly MP, Carvunis AR, Convery-Zupan D, Corominas R, Coulombe-Huntington J, Dann E, Dreze M, Dricot A, Fan C, Franzosa E, Gebreab F, Gutierrez BJ, Hardy MF, Jin M, Kang S, Kiros R, Lin GN, Luck K, MacWilliams A, Menche J, Murray RR, Palagi A, Poulin MM, Rambout X, Rasla J, Reichert P, Romero V, Ruyssinck E, Sahalie JM, Scholz A, Shah AA, Sharma A, Shen Y, Spirohn K, Tam S, Tejeda AO, Trigg SA, Twizere JC, Vega K, Walsh J, Cusick ME, Xia Y, Barabasi AL, Iakoucheva LM, Aloy P, De Las Rivas J, Tavernier J, Calderwood MA, Hill DE, Hao T, Roth FP, Vidal M (2014) A proteome-scale map of the human interactome network. Cell 159 (5):1212–1226. doi:10.1016/j.cell.2014.10.050 [PubMed: 25416956]

17. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H, Szpyt J, Tam S, Zarraga G, Pontano-Vaites L, Swarup S, White AE, Schweppe DK, Rad R, Erickson BK, Obar RA, Guruharsha KG, Li K, Artavanis-Tsakonas S, Gygi SP, Harper JW (2017) Architecture of the human interactome defines protein communities and disease networks. Nature 545 (7655):505–509. doi:10.1038/nature22366 [PubMed: 28514442]

18. Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K, Dong R, Guarani V, Vaites LP, Ordureau A, Rad R, Erickson BK, Wuhr M, Chick J, Zhai B, Kolippakkam D, Mintseris J, Obar RA, Harris T, Artavanis-Tsakonas S, Sowa ME, De Camilli P, Paulo JA, Harper JW, Gygi SP (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. Cell 162 (2):425–440. doi:10.1016/j.cell.2015.06.043 [PubMed: 26186194]

19. Banworth MJ, Li G (2018) Consequences of Rab GTPase dysfunction in genetic or acquired human diseases. Small GTPases 9 (1-2):158–181. doi:10.1080/21541248.2017.1397833 [PubMed: 29239692]

20. Li G (2011) Rab GTPases, membrane trafficking and diseases. Curr Drug Targets 12 (8):1188–1193. doi:10.2174/138945011795906561 [PubMed: 21561417]

21. Seabra MC, Goldstein JL, Sudhof TC, Brown MS (1992) Rab geranylgeranyl transferase. A multisubunit enzyme that prenylates GTP-binding proteins terminating in Cys-X-Cys or Cys-Cys. J Biol Chem 267 (20):14497–14503 [PubMed: 1321151]

22. Kalin S, Hirschmann DT, Buser DP, Spiess M (2015) Rabaptin5 is recruited to endosomes by Rab4 and Rabex5 to regulate endosome maturation. J Cell Sci 128 (22):4126–4137. doi:10.1242/jcs.174664 [PubMed: 26430212]

23. Song GJ, Jeon H, Seo M, Jo M, Suk K (2018) Interaction between optineurin and Rab1a regulates autophagosome formation in neuroblastoma cells. J Neurosci Res 96 (3):407–415. doi:10.1002/jnr.24143 [PubMed: 28843006]

24. Hattula K, Peranen J (2000) FIP-2, a coiled-coil protein, links Huntingtin to Rab8 and modulates cellular morphogenesis. Curr Biol 10 (24):1603–1606. doi:10.1016/s0960-9822(00)00864-2 [PubMed: 11137014]

25. Kobayashi H, Etoh K, Ohbayashi N, Fukuda M (2014) Rab35 promotes the recruitment of Rab8, Rab13 and Rab36 to recycling endosomes through MICAL-L1 during neurite outgrowth. Biol Open 3 (9):803–814. doi:10.1242/bio.20148771 [PubMed: 25086062]

26. Rahajeng J, Giridharan SS, Cai B, Naslavsky N, Caplan S (2010) Important relationships between Rab and MICAL proteins in endocytic trafficking. World J Biol Chem 1 (8):254–264. doi:10.4331/wjbc.v1.i8.254 [PubMed: 21537482]

27. Klinger M, Wang W, Kuhns S, Barenz F, Drager-Meurer S, Pereira G, Gruss OJ (2014) The novel centriolar satellite protein SSX2IP targets Cep290 to the ciliary transition zone. Mol Biol Cell 25 (4):495–507. doi:10.1091/mbc.E13-09-0526 [PubMed: 24356449]

28. Kurtulmus B, Wang W, Ruppert T, Neuner A, Cerikan B, Viol L, Duenas-Sanchez R, Gruss OJ, Pereira G (2016) WDR8 is a centriolar satellite and centriole-associated protein that promotes ciliary vesicle docking during ciliogenesis. J Cell Sci 129 (3):621–636. doi:10.1242/jcs.179713 [PubMed: 26675238]

29. Yukawa M, Ikebe C, Toda T (2015) The Msd1-Wdr8-Pkl1 complex anchors microtubule minus ends to fission yeast spindle pole bodies. J Cell Biol 209 (4):549–562. doi:10.1083/jcb.201412111 [PubMed: 25987607]

30. Cox JV, Kansal R, Whitt MA (2016) Rab43 regulates the sorting of a subset of membrane protein cargo through the medial Golgi. Mol Biol Cell 27 (11):1834–1844. doi:10.1091/mbc.E15-03-0123 [PubMed: 27053659]

31. Li C, Wei Z, Fan Y, Huang W, Su Y, Li H, Dong Z, Fukuda M, Khater M, Wu G (2017) The GTPase Rab43 Controls the Anterograde ER-Golgi Trafficking and Sorting of GPCRs. Cell Rep 21 (4):1089–1101. doi:10.1016/j.celrep.2017.10.011 [PubMed: 29069590]

32. Das S, Hehnly H, Doxsey S (2014) A new role for Rab GTPases during early mitotic stages. Small GTPases 5. doi:10.4161/sgtp.29565

33. Ortiz D, Medkova M, Walch-Solimena C, Novick P (2002) Ypt32 recruits the Sec4p guanine nucleotide exchange factor, Sec2p, to secretory vesicles; evidence for a Rab cascade in yeast. J Cell Biol 157 (6):1005–1015. doi:10.1083/jcb.200201003 [PubMed: 12045183]

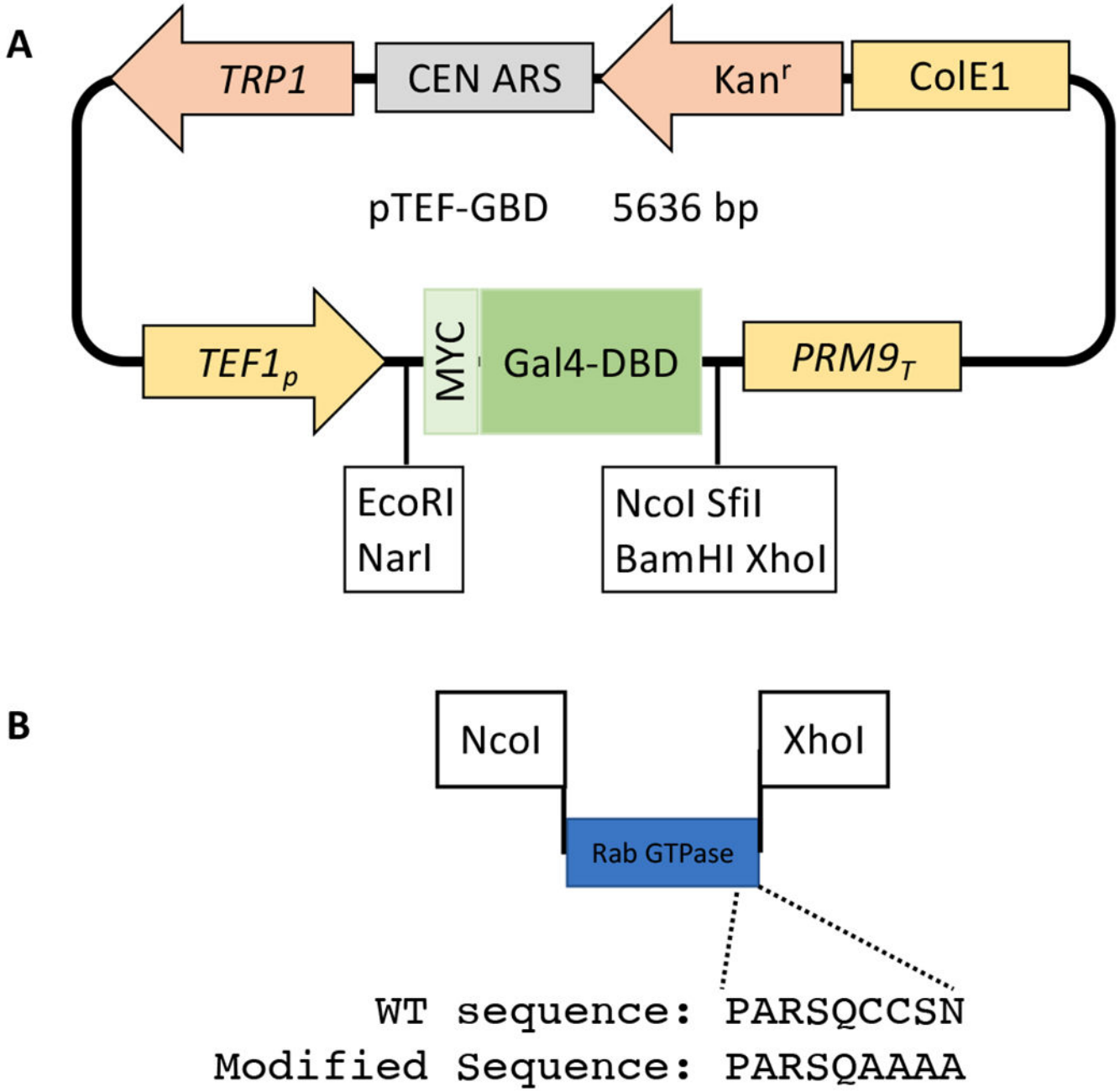

Figure 1.

Construction of Rab bait Y2H plasmids. (A) Schematic of the Gal4-DNA-binding domain expressing plasmid pTEF-GBD, a low-copy yeast plasmid with *TRP1* for selection in yeast and Kan^r for selection in bacteria. (B) Shown is the cloning of a Rab GTPase with alteration of the Rab C-terminal CAAX box. Clone the Rab GTPase of interest into the pTEF-GBD plasmid linearized with either EcoRI/NarI or NcoI/XhoI for N- or C-terminal cloning respectively to the Gal4-DNA-binding domain. The CAAX box on Rab proteins directs their lipidation (geranylgeranylation) which would direct their interaction with GDI and also membranes. This would hamper their ability to translocate to the nucleus and thus

their ability to interact with Gal4-activation-domain ‘prey’ in the nucleus. To avoid this, the Cysteine residues and remaining residues that define the CAAX box are replaced with alanine codons.

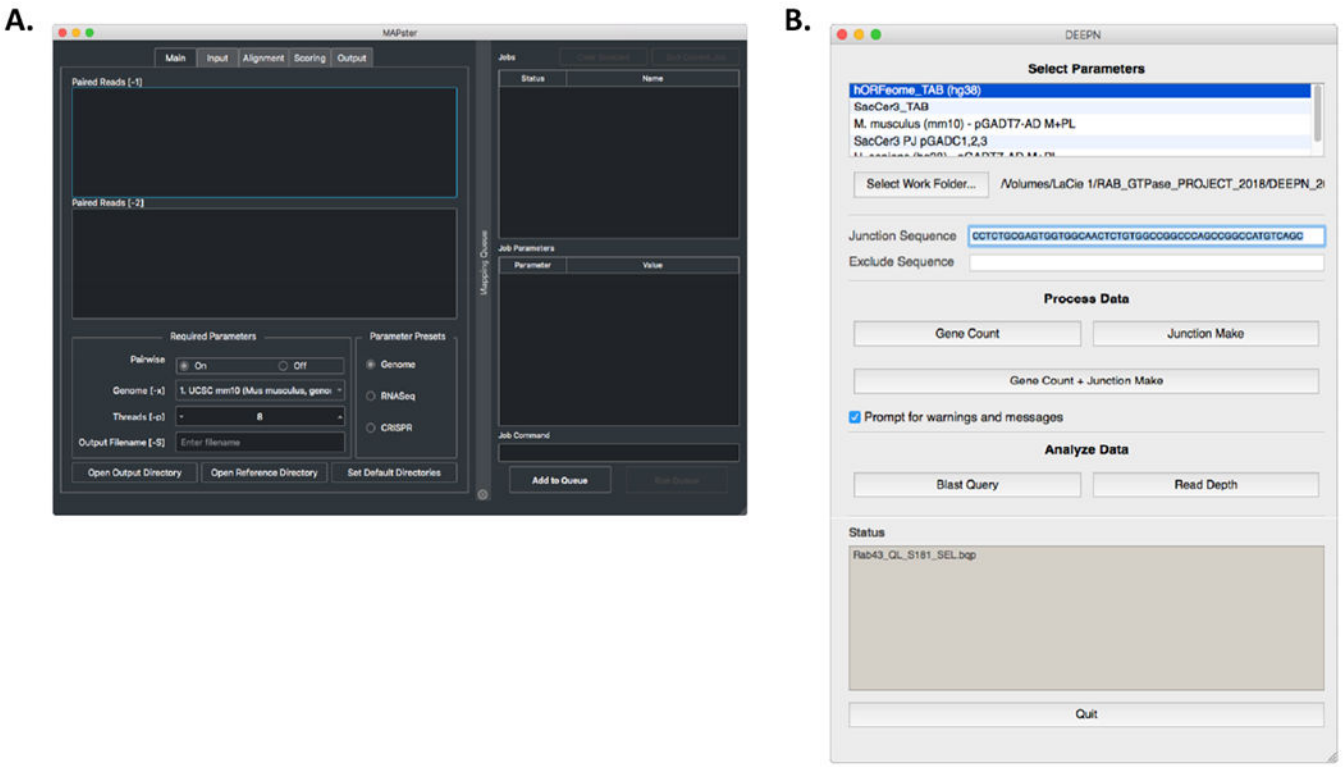

Figure 2.

User interface of two software programs used to computationally analyze Illumina sequence data from batch Y2H assays. (A) shows the Mapster program that can be used to Map sequence reads to the relevant genome to create a .sam file used for all subsequent analysis. (B) shows the DEEPN main user window that allows access to analysis modules such as GeneCount, BlastQuery, and ReadDepth.

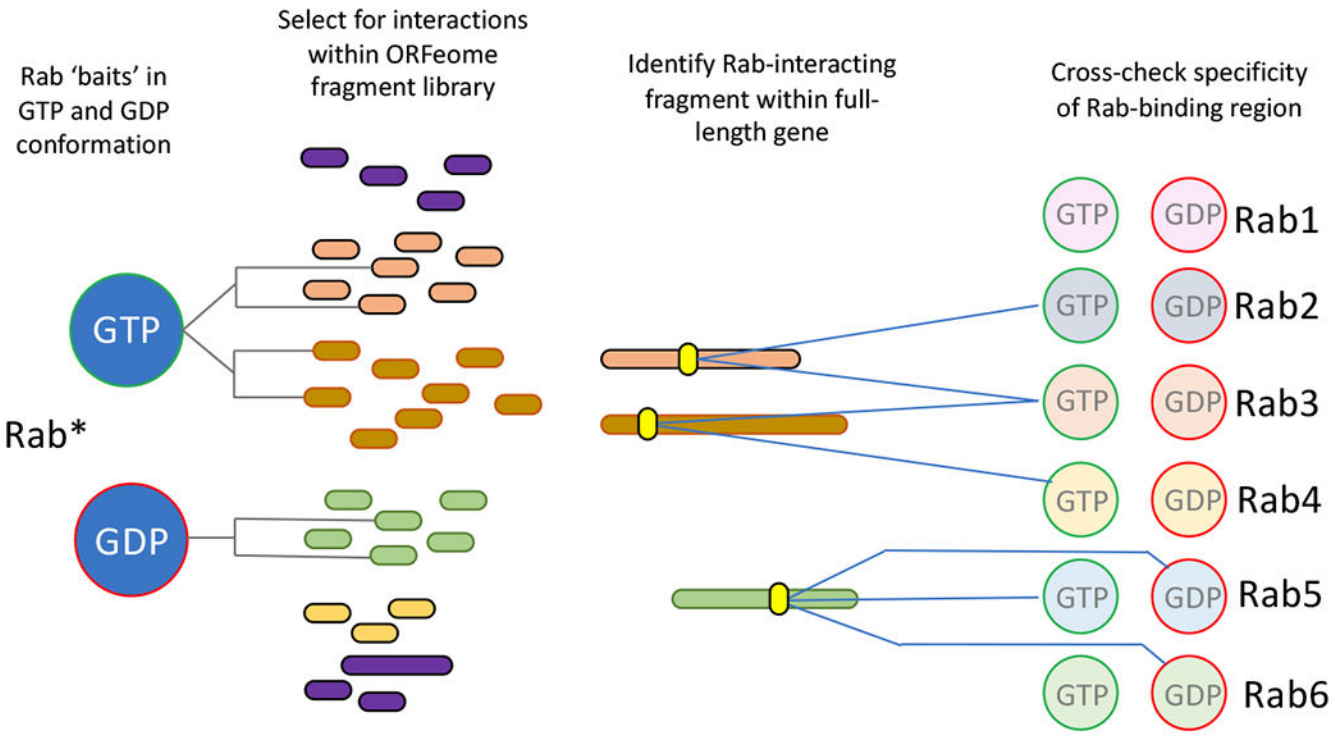

Figure 3.

Overview of the process to find differential Rab interacting proteins using DEEPN, a batch yeast 2-hybrid approach, and subsequent validation and characterization. Rab 'bait' plasmids are constructed as shown in Figure 1. These are introduced into yeast containing a diverse Y2H 'prey' library comprised of gene fragments. After selection for positive Y2H interactions, the entire population of remaining prey plasmids is sequenced to find genes that are enriched, indicating a positive Y2H interaction, and data are further analyzed to determine what gene fragments were selected for. These fragments also define the Rab-interacting region(s) of each prey protein (yellow regions). The computationally deduced plasmids that produce the Rab-interacting protein fragments are then reconstructed and tested in a series of binary Y2H interactions with a matrix of Rab proteins in their GDP and GTP-bound conformations to determine the specificity of interactions.

Bait 1: Rab43 QL														
Bait 2: Rab43 TN														
Gene	Base	Vec	Bait1	Bait2	Enr1	Enr2	AdjEnr1	AdjEnr2	pBait1_Vec	pBait2_Vec	pBait1_Bait2	pBait2_Bait1	pBait1	pBait2
SSX2IP	147.1346932	4.80345542	10997.52743	2.116454375	11.1458859	-1.16367639	8.623252169	-2.961213819	0.99987414	0.019400812	0.999990827	9.17E-06	0.99986497	1.78E-07
			Bait1_Selected_1								Bait2_Selected_1			
inframe_inorf	upstream	in_orf	downstream	in_frame	backwards	intron		inframe_inorf	upstream	in_orf	downstream	in_frame	backwards	intron
99.8	0	100	0	99.8	0.1	0		21.1	0	100	0	21.1	47.4	0
Bait 1: Rab5 QL														
Bait 2: Rab5 SN														
Gene	Base	Vec	Bait1	Bait2	Enr1	Enr2	AdjEnr1	AdjEnr2	pBait1_Vec	pBait2_Vec	pBait1_Bait2	pBait2_Bait1	pBait1	pBait2
SSX2IP	156.6158666	4.68457424	3.12031303	3400.568134	-0.5786094	9.4883466	-3.43836295	5.006179014	5.93E-05	0.998835209	8.40E-08	0.999999916	4.99E-12	0.99883512
			Bait1_Selected_1								Bait2_Selected_1			
inframe_inorf	upstream	in_orf	downstream	in_frame	backwards	intron		inframe_inorf	upstream	in_orf	downstream	in_frame	backwards	intron
15.5	2.3	97.7	0	15.5	55.8	0		96.5	0	100	0	96.5	1.2	0

Figure 4.

Data from StatMaker identifying SSX2IP as an interacting proteins with Rab43 (top) and Rab5A (bottom). The first row shows the number of sequence reads found for SSX2IP for 3 different datasets, Vector alone, Bait 1 (Rab43_QL or Rab5_QL), and Bait 2 (Rab43_TN or Rab5_SN) in ppm. A 3 way statistical model used to evaluate the likelihood that there is a specific enrichment in the number or reads for SSX2IP (Y2H interaction) for a given bait vs Vector alone and vs the other bait. These are observed as pBait1 and pBait2 having a maximum probability of 1. These data indicate that SSX2IP specifically interacts with Rab43 in its GTP-bound conformation, but not its GDP-bound conformation nor Vector alone. The data below indicate that that SSX2IP specifically interacts with Rab5 in its GDP-bound conformation, but not its GTP-bound conformation nor Vector alone. The data also summarize the percentage of junctions (those sequence reads that span the Gal4-activation domain and the prey protein of interest), that are in the correct translational frame and within the open-reading frame (inframe_inorf). Having a large enrichment of reads for a prey hit as well as a large proportion of those being in the proper reading frame are both criteria to determine the authenticity of the computationally derived Y2H interaction.

Figure 5.

Shown is the display from BlastQuery that compares various fragments of SS2XIP from 3 different datasets: Vector alone, Rab5A_SN (locked in the GDP-found form), and Rab5A_QL (locked into the GTP-bound form). BlastQuery displays data on the different junction fragments of the selected gene that were found in the sequence results in **(A)** tabular format and **(B)** graphical format. Junction fragments are those that span the C-terminal end of Gal4-activation domain and the fusion point into the gene of interest. The position within the gene of interest where it is fused to the Gal4 activation domain is listed as well as the number of times this particular junction was found in the dataset (in ppm). Also shown, is a calculation of whether the fragment of the gene of interest is in the coding region and whether it is in the same translational frame as the Gal4-activtion domain. For the example here, a fragment of SSX2IP that begins at position 890 is greatly enriched in in the Rab5A-SN dataset (1390 ppm), over what its abundance was in the vector only

dataset (0.217 ppm) and Rab5A_QL dataset (0.177). The exact genbank identifier of the annotated cDNA/transcript to which blast matches were found is given in the top-center and the sequence of that cDNA is provided in the bottom window for unambiguous identification of the prey gene of interest.

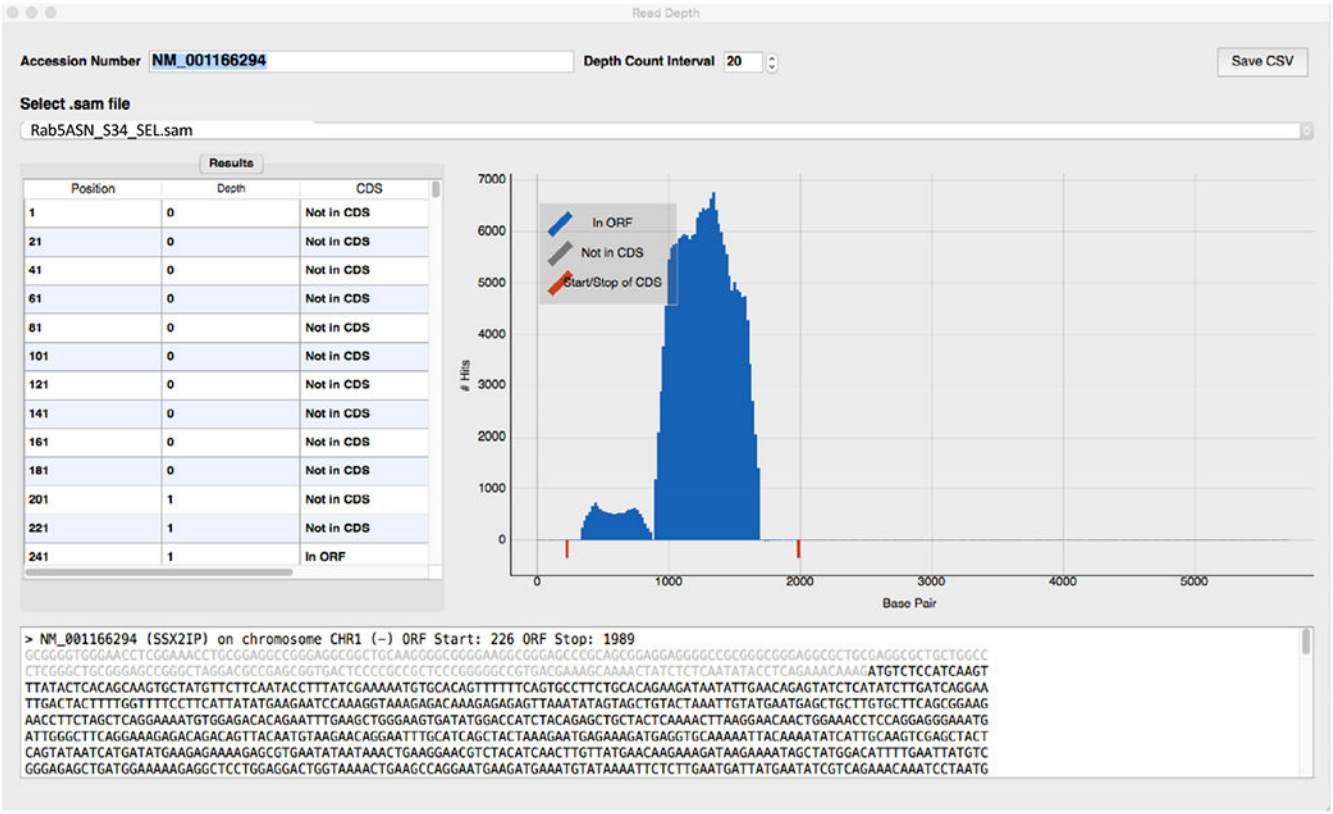

Figure 6.

Display of ReadDepth that shows the sequence coverage of a prey gene of interest designated by a Genbank identifier in a particular dataset. The level of sequence coverage indicates the level of enrichment of the corresponding gene region found in the data. This feature can identify which fragments were selected for or enriched by Y2H interactions and which fragments were not. In this case, the interacting gene of interest is SSX2IP, selected for an interaction with Rab5A_SN locked in the GDP-bound form. Two regions appear in the ReadDepth display, a very abundant fragment beginning at position ~900bp and a more minor fragment beginning at position ~340bp.

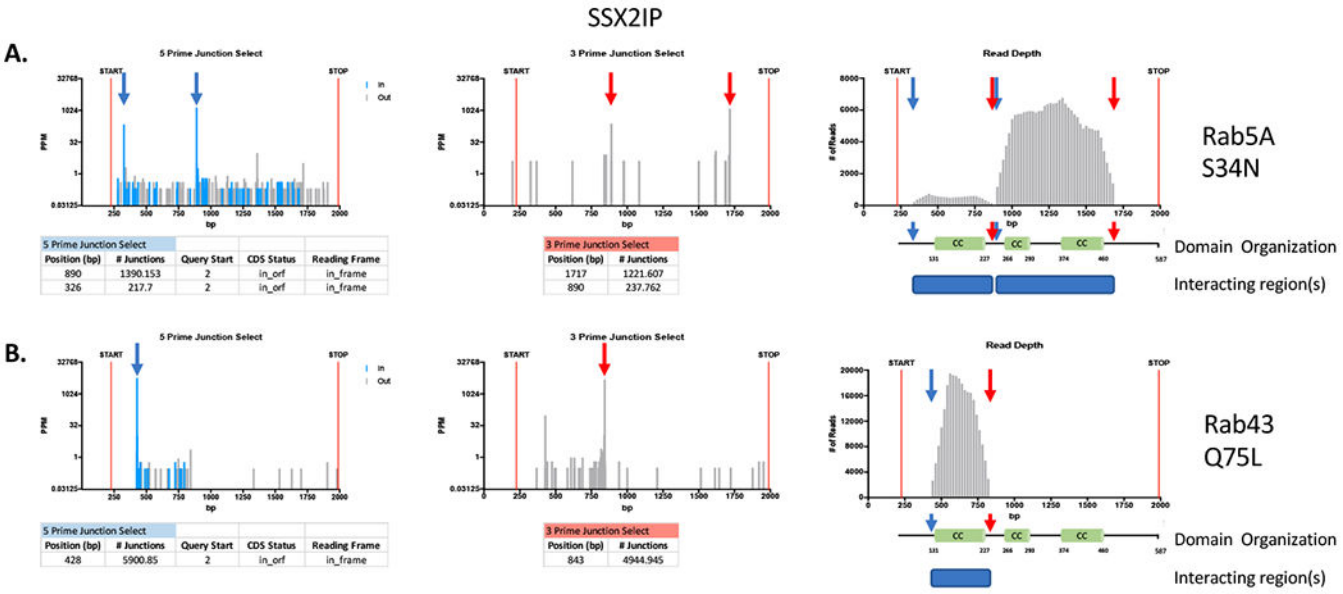

Figure 7.

Collated data for Rab Y2H interaction. Shown are the collated data for interaction of SSX2IP with Rab GTPases. DEEPN datasets for multiple Rab GTPases were generated and two different Rab proteins were found to interact with SSX2IP. However, the fragments of SSX2IP mediating the interaction were different between Rab5A_SN (GDP-bound conformation) and Rab43_QL (GTP-bound conformation). By summarizing the junction fragment data available from BlastQuery and the sequence coverage from ReadDepth, a helpful picture of these interactions can be generated for comparison. (A) shows the number of 5′ junctions found containing SSX2IP and their position along the length of the SSX2IP cDNA (left) for the dataset for Rab5A_SN interactions. Lines in blue are in the same reading frame as the up-stream Gal4-activation domain, whereas the grey are out of frame fusions. Blue arrows indicate the 5′ end of fragments that are must abundant and that are shown in tabular format below. Similar junction fusion points are shown for the 3′ end of SSX2IP fragments, with the most abundant 3′ junction indicated with red arrows and highlighted in tabular format below. Data from ReadDepth (right) is shown overlayed with blue and red arrows corresponding to the abundant 5′ and 3′ fragment junctions, respectively. Below is the domain organization of the SSX2IP protein, scaled to the nucleotide positions within the read depth data above. Delineated in blue are the putative Rab interacting regions extrapolated from the data. (B) Same analysis in A but for SSX2IP interactions with Rab43-QL (GTP-bound conformation) showing interaction with the 5’ fragment corresponding to residues 68-206.

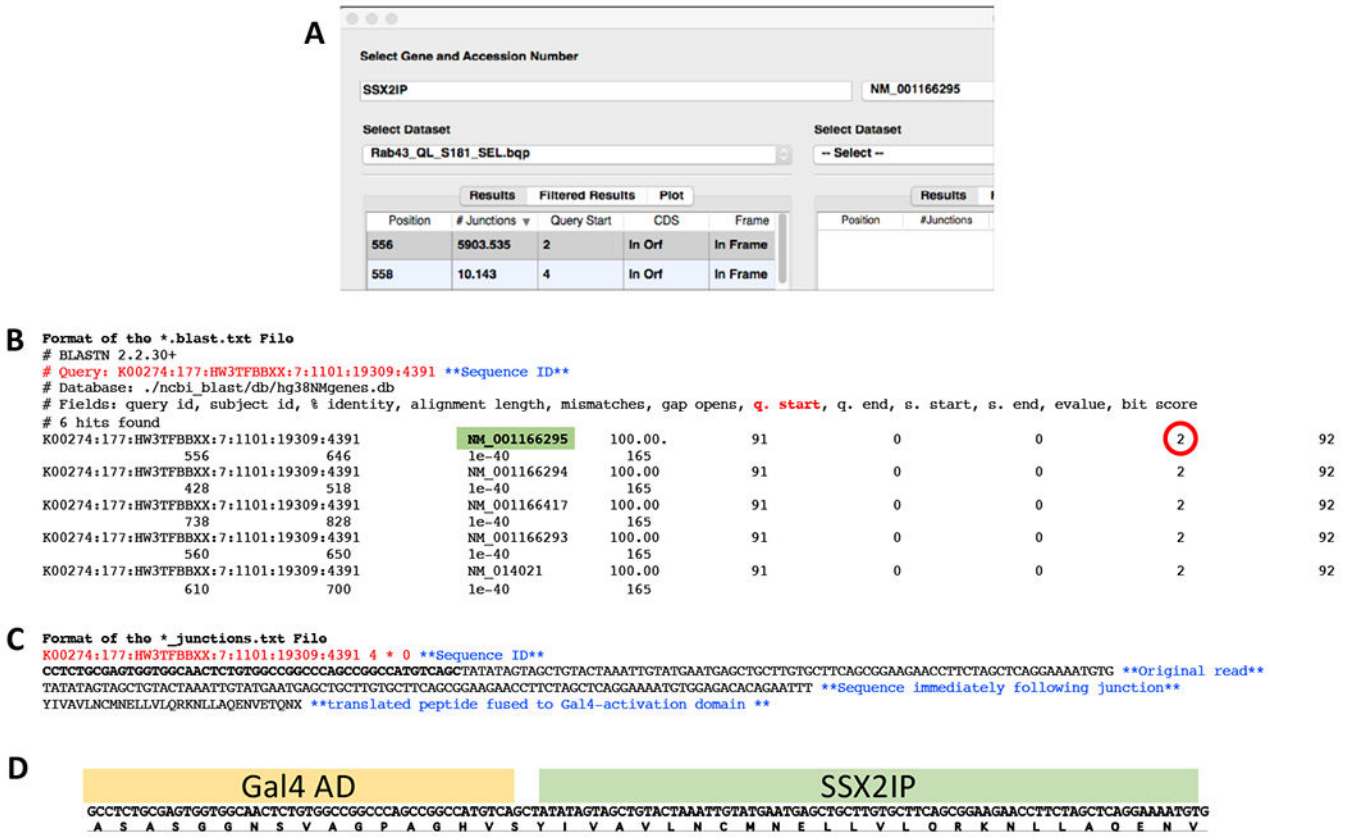

Figure 8.

BlastQuery can identify particular fragments that are greatly enriched under selection. These fragments are identified by their junction sequences, which are the sequences that are immediately adjacent to the sequence of the Gal4-activation domain with which they are fused. With this information in hand, one can trace back to the original sequence. The data that BlastQuery displays can be found in the associated *.blast.txt file, which contains a series of blast results for each junction sequence found in the original Illumina datafiles. In this example, what is sought is the original sequence that spans the Gal4-activation domain and the gene of interest.

A. BlastQuery shows that the interacting gene is SSX2IP, and the junction homology is found to start at a region beginning at base-pair position 558, with a Q-start of 2 meaning that there is an insertion of a bp in between the end of the Gal4-activation domain and the beginning of the SS2XIP coding region. The exact gene-ID for the SSX2IP that was matched is found in the BlastQuery window as NM_001166295.

B. To find the original sequence read identifier, one needs to open the corresponding *.blast.txt file that lists all the blast hits and search for a blast hit for NM_001166295 and find one that begins at position 558 with a Q-start of 2. Shaded in grey is the Illumina read identifier, in this case: K00274:177:HW3TFBBXX.

C. By searching for the read identifier in B in the *.junctions.txt file, the original sequence read can be found. The junctions.txt file is arranged with the following format: read ID, SAM flag (typically ‘4’ because the read was not mapped), contig name (typically ‘*’ because the read was not mapped), map position (typically ‘0’ because the read was not

mapped), the original read (which is what is needed to retrieve), the junction sequence downstream of the Gal4-activation domain (which is used to in the Blast search), and the amino acid translation of the junction sequence. The original sequence read can be extracted from the file and used to reconstruct what the precise sequence is that joins the Gal4-activation domain in the prey plasmid with the gene of interest fused to it.

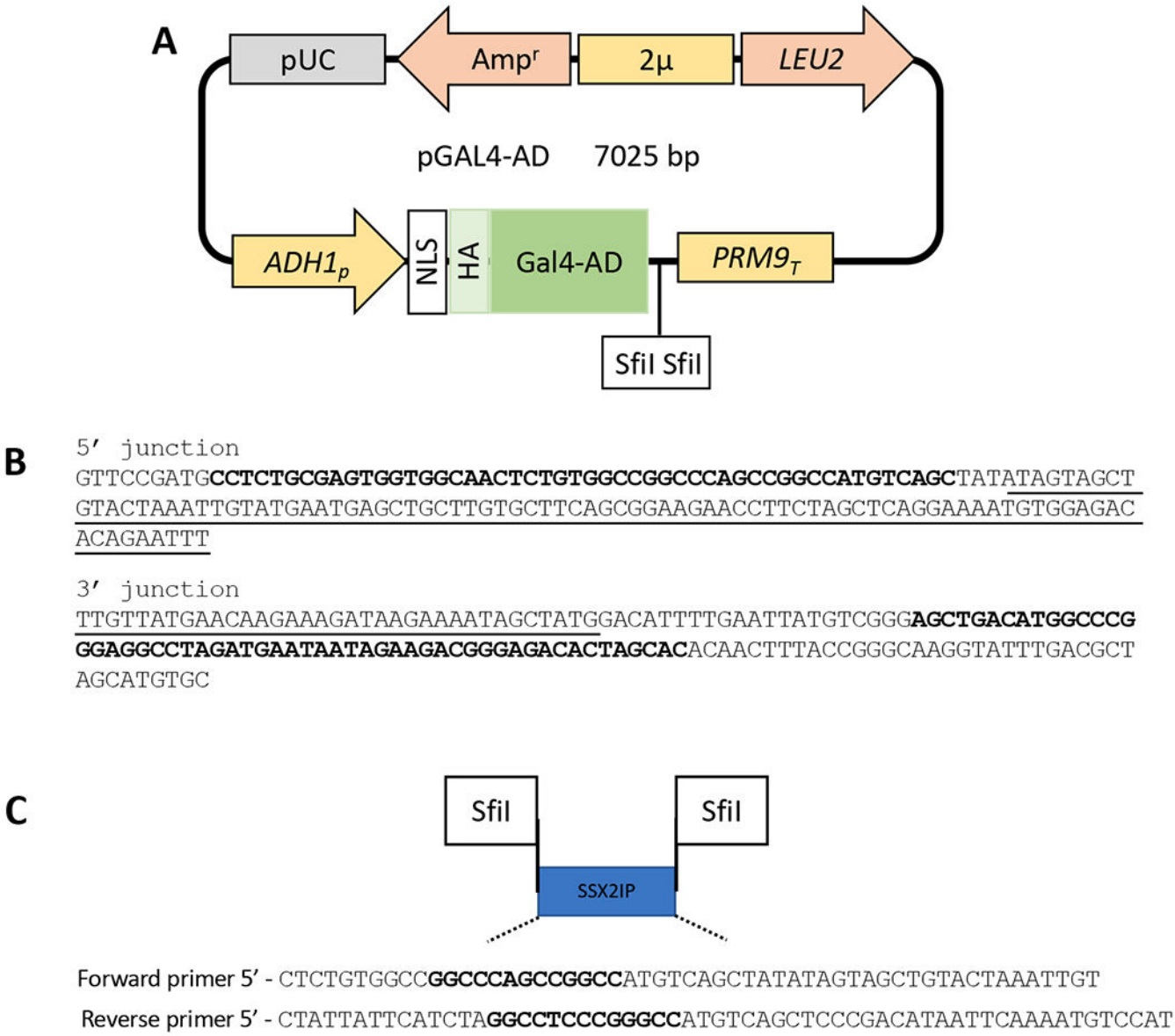

Figure 9.

Reconstructing prey plasmids. (A) Schematic of pGAL4-AD, a high-copy LEU2-containing plasmid that expresses an HA-tagged Gal4 activation domain and is used to house yeast genomic and human ORFeome fragment libraries. (B) Sequences corresponding to the flanking regions of a particular fragment of SSX2IP (which was identified to interact with Rab43) as determined by computational reconstruction using BlastQuery as noted in Figure 8. Underlined portion is the region encoding SSX2IP, bold is the region from the prey plasmid pGAL4-AD. (C) Example oligo primers that are used to amplify the designated fragment of SSX2IP and clone into SfiI double cut pGAL4-AD. Bold shows the SfiI sites.

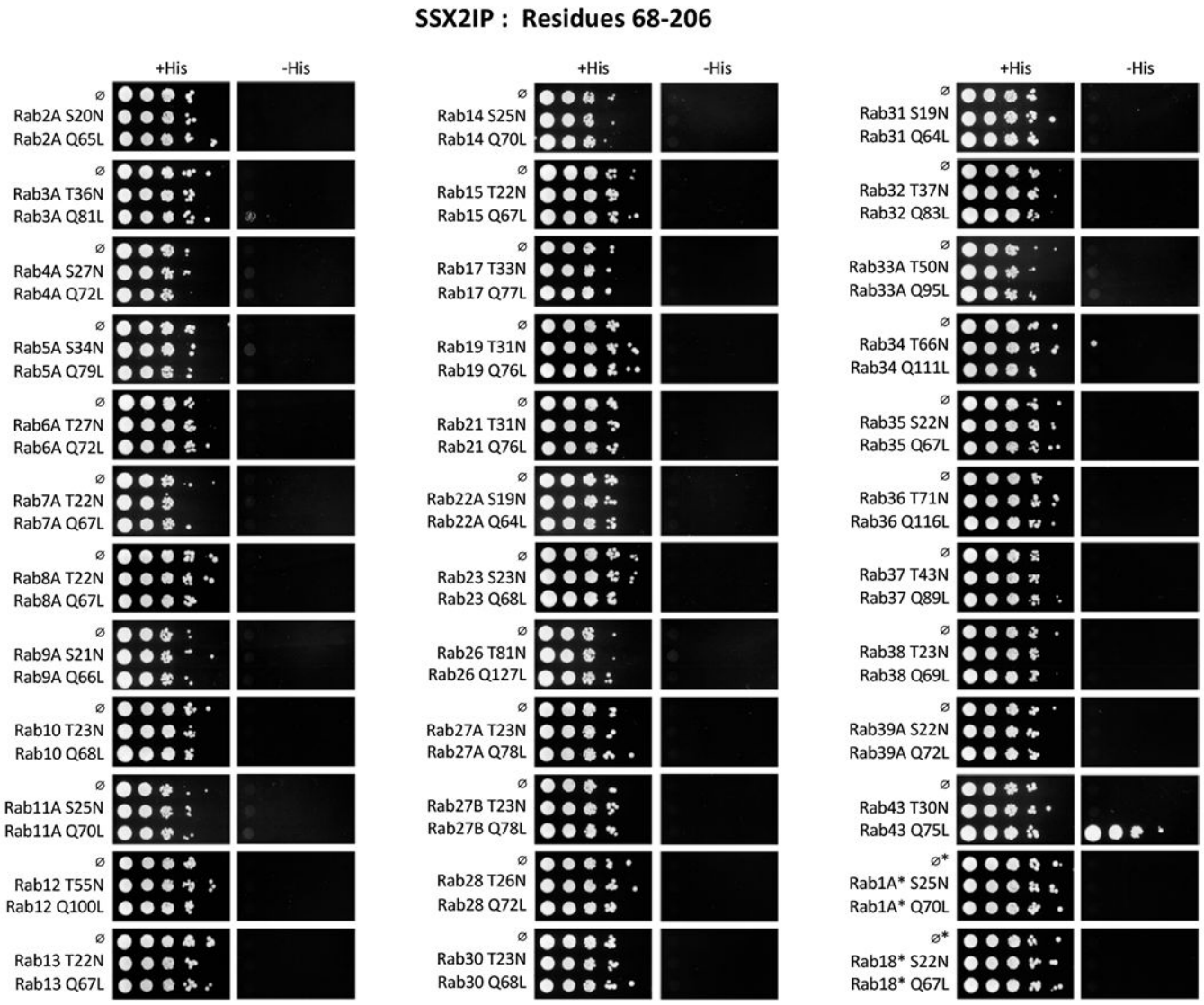

Figure 10.

Binary Y2H interactions. A ‘prey’ plasmid encoding an interacting portion of SSX2IP, as extrapolated from the sequence data, was reconstructed and introduced into diploid yeast also containing the indicated pTEF-GBD fusion ‘bait’ plasmids expressing Rab fusions (in either the GDP or GTP conformation) as well as vector only (ø). The diploid cells had as their sole source of HIS3 gene a version under the control of the promoter that requires a 2-hybrid interaction to complement a split Gal4 transcription factor. Yeast were serially diluted and plated on media containing histidine (+His) or lacking histidine (-His), the latter of which reveals positive Y2H interactions indicated by colony growth. To determine Rab specificity, a large matrix is constructed to find whether the particular fragment of SSX2IP interacts with multiple Rab proteins and what the conformational specificity is. In these data, SSX2IP (residues 68-206) is specific only for the GTP-bound conformation of Rab43.

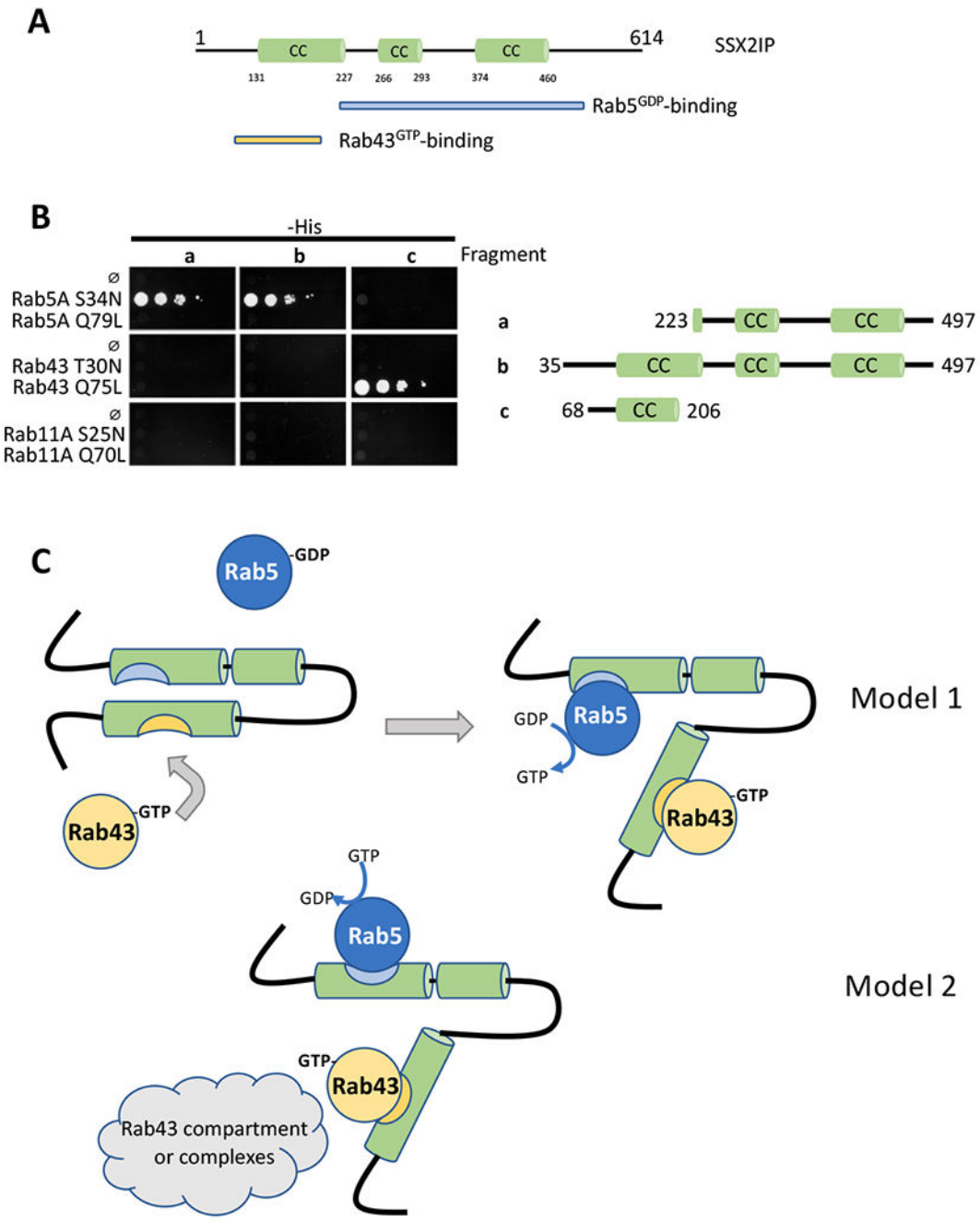

Figure 11.

Hypothesis for SSX2IP. (A) Schematic of the full length SSX2IP protein where CC denotes coiled coil regions of the protein. Also depicted are the regions identified by DEEPN and validation studies for interaction with Rab43-GTP and Rab5A-GDP, which are housed in separate parts of the protein. (B) Binary Y2H analysis with Rab43, Rab5 and an irrelevant control Rab, Rab11, each in their GDP- or GTP-bound conformations alongside of vector alone (ø). Growth is shown only on plates lacking histidine whereby growth indicates a positive protein interaction by Y2H. Shown (right) are the three fragments (a,b,c; residues

223-497, 35-497, and 68-206, respectively) tested for a binary Y2H interaction. (C) Two models for how Rab5 and Rab43 binding could be functionally integrated. Model 1 proposes a Rab cascade where Rab5-GDP binding is triggered by binding of Rab43-GTP. Model 2 proposes that activation of Rab43 allows it to recruit the Rab5 exchange activity of SSX2IP to other Rab43 effectors or Rab43 enriched locations.