Structure-based prediction of Ras-effector binding affinities and design of 'branchegetic' interface mutations

Philipp Junk [2,3*], Christina Kiel [1,2,3]

Affiliations

[1] Department of Molecular Medicine, University of Pavia, 27100 Pavia, Italy

[2] Systems Biology Ireland, School of Medicine, University College Dublin, Dublin 4, Ireland.

[3] UCD Charles Institute of Dermatology, School of Medicine, University College Dublin, Dublin 4, Ireland

*Correspondence: Philipp Junk (philipp.junk@ucdconnect.ie)

Summary

Ras is a central cellular hub protein controlling multiple cell fates. How Ras interacts with a variety of potential effector proteins is relatively unexplored, with only some key effectors characterized in great detail. Here, we have used homology modelling based on X-ray and AlphaFold2 templates to build structural models for 54 Ras-effectors complexes. These models were used to estimate binding affinities using a supervised learning regressor. Furthermore, we systematically introduced Ras 'branch-pruning' (or branchegetic) mutations to identify 200 interface mutations that affect the binding energy with at least one of the model structures. The impacts of these branchegetic mutants were integrated into a mathematical model to assess the potential for rewiring interactions at the Ras hub on a systems level. These findings have provided a quantitative understanding of Ras-effector interfaces and their impact on systems properties of a key cellular hub.

Keywords

AlphaFold, branch pruning, edgetics, enedgetics, effectors, FoldX, homology modelling, hub, networks, Ras

Introduction

Ras is a key cellular signaling hub and oncogene [1]. The first correct Ras structure was famously described in 1989 [2 3] and consists of the G domain super-fold (six β-strands and five α-helices; [1]). It's main structural and functional characteristic is the nucleotide binding site (the typical α,β-fold of nucleotide binding proteins), which can bind GTP and hydrolyze it to GDP. Then, GDP gets released and a new nucleotide (favorably GTP since it is higher abundant in cells) gets bound. Depending on which nucleotide is bound, the functional state of RAS is different. GDP bound Ras assumes a so-called inactive conformation, whereas upon the binding of GTP, the conformation of Ras is called active. The difference between the active and the inactive conformation is that, in the active conformation, two loop regions called switch 1 and switch 2 are interacting with the GTP and thereby tightly bound to it [1] (Figure 1A). This rearranges the interface of Ras in a specific way so that effector proteins with a specific structural motive can interact with Ras. The structural motive required for the interaction with Ras has a ubiquitin domain super-fold [4]. There exist three families of these domains, called Ras binding domain (RBD), Ras association domain (RA domain) and PI3K-Ras binding domain (PI3K RBD) with only minor sequence and structural differences between them. In the following, for the sake of simplicity, all of these will be called RBDs.

Ras·GTP can interact and activate many downstream effectors, some of which are better studied than others [5]. Well-studied effectors include PI3-kinases, RalGDS, and Raf kinases, which control important cellular processes such as survival, polarization, adhesion, migration, and proliferation. In our previous work, we characterized a set of 56 RBD-containing proteins as potential Ras effectors, which converge into 12 classes that are linked to different downstream cellular processes and phenotypes [6 7 8 9].

Ras-effector interactions are interesting from a systems biology and network point of view. Effectors use a mutually exclusive binding site on Ras, hence competition for binding can occur under certain conditions [10]. Also, the experimental and predicted affinities between RBDs and Ras·GTP vary, ranging from nano- to micromolar K_d values [8]. Previously we analyzed how the amount of effectors in complex with Ras varies in different cell types [6 7]. We find that only 9 effectors are predicted to bind in significant amount to Ras·GTP using the RBD alone [7]. However, 31 effectors are predicted to form significant complexes with Ras if they are additionally recruited to the plasma membrane via other domains present in effectors (piggyback mechanism, [11]). Hence, even weak binding affinities on the level of RBD-RAS binding can turn into significant complex formation if effectors are recruited (e.g. in a context

specific manner). Thus, for systems analyses and computational models, we need good estimations for all binding affinities between RBDs and Ras.

Structural analysis can lead to a deeper understanding of protein-protein interaction (PPI). Previous we used homology modelling, based on experimentally determined complex structures, to predict binding between effectors and Ras·GTP [12 13]. While this has provided insights into binding propensities, a limitation of this work was that no quantitative binding affinities were predicted, but qualitative binding by classifying effectors into categories of 'binding', 'non-binding', and 'twilight'. It is now timely to revisit homology modelling of Ras-RBD complexes, as new template structures are available by (i) X-ray crystallography (8 of 56 effectors are crystallized in complex with RAS) and (ii) AlphaFold [14 15].

Ras is frequently mutated in different cancers, where aberrant Ras signaling plays a role in cancer initiation, progression, and metastasis via alterations in metabolism, proliferation, and survival [16]. As Ras cannot be directly targeted (or only certain Ras mutations) [17], much hope rests on network-centric approaches [18] that involve Ras-effector interactions or downstream pathways [9]. Therefore, tinkering with Ras-effector binding is an attractive alternative strategy for finding suitable targets for therapeutic interventions. Previously we developed a 'branch-pruning' (or "branchegetic", in analogy to "edgetics" [19] and "enedgetics" [20]) strategy for Ras effector interactions, where mutations are introduced into Ras that differentially impact binding to effectors [21]. For example, introducing a mutation can result in a steric clash in the interface formed with one effector, but not with another; hence the interaction with one effector is broken while intact for another.

In this work, we first generate homology models of all Ras-effector (RBD) interactions and predict affinities of the RBDs in complex with Ras·GTP. We then use the generated model structures to employ a systematic branchegetic strategy that explores the impact of Ras interface mutants on binding to all effectors. Altogether, our results contribute to a quantitative and systems-level understanding of Ras-effector interactions and further our understanding of Ras in health and disease.

Results

Structural analysis of experimental complex structures

From the few experimentally determined complex structures of Ras with these RBDs, there are some structural similarities that can be determined (Figure 1). The main interface on the site of Ras consists of the two switch regions, switch 1 and switch 2. One of the main structural features

of the interface between Ras and RBDs is the formation of an intra-molecular β-sheet between β2 on RAS and β2 on the RBD. This interaction is a highly conserved structural feature across all available complex structures, with deviations in orientation of less than 1 Å (Figure 1B, see Methods for MAE).

Analyzing the energetic profile of the interface *in silico* using the FoldX force field shows that there are also some recurring hotspots in the interface (highlighted in Figure 1C). I36, D38 and Y40 are well characterized as important residues for the interaction between Ras and effector domains. Additionally, for the structures 3DDC and 1LFD, the mutation E31K was introduced to stabilize the complex for crystallization. Our analysis confirms that this interface mutation has indeed been favorable for the complex formation. Finally, the energetic contributions of the function regions switch 1 and 2 to the interface were analyzed. Both relative and absolute contributions are diverse, although switch 1 contributions to binding dominate (Figure 1D). Altogether, the analysis of existing Ras-RBD effector structures indicates that although there are many common features of the Ras RBD interface such as the intra-molecular β-sheet or the hydrophobic patch around I36, the actual energetic contributions can come from different parts of the interface. It also sets the basis for a successful homology modelling approach.

Homology modelling and characterization of modelled interfaces

In order to study these interface features in a more diverse set of structures, homology models of the complex between RAS and RBD were constructed for all proteins containing an RBD in the human proteome.

The homology modelling pipeline is based on the already existing complex models (Table S1). Also, with the recent release of AlphaFold2 [14] and the accompanying AlphaFold Protein Structure Database [15], the RBD domains of all potential effectors were extracted from that database and used. The structures of RBDs are predicted with good confidence by AlphaFold2 and our analysis indicates that AlphaFold2 is reliable at predicting the RBD fold (Figure S1). Additionally, AlphaFold2 complex modelling was attempted for all potential complex structures and models which AlphaFold2 were confident in (by Predicted Alignment Error (PAE)) and where the β-sheet alignment of the interface was within a tolerance to what has been observed in crystal structures, were used as templates as well (Figure S2 and S3, compare MAE Figure 1). There are two kinds of templates to use: 1) complex templates which comprised of the already experimentally determined complex structures of Ras and RBDs, as well as “good” AlphaFold2 predicted complex templates (Table S1), and 2) templates of the

RBDs alone which were extracted from the AlphaFold2 Protein Structure Database (Table S2). An overview over the pipeline is depicted in Figure 2. Homology modelling was performed using homelette [22] with modeller and altMod [23-25]. Evaluation of predicted structures was performed using QMEAN, MolProbity and SOAP potentials [26-29]. The top 300 models for each target for each source of complex templates (experimental or AlphaFold2) were selected by combining the different scores. Subsequently, analysis using FoldX (interaction energy and alanine scan) was performed.

In order to evaluate our approach, we generated a validation set, in which we created models for the structures already solved by X-ray crystallography without using information from the specific structure. Models generated in this validation set were compared to the underlying ground truth by assessing their correlation of the *in silico* alanine scan results to those of the crystal structures of interest. Using this ground truth, different methods to select representative models from the hundreds of structures were assessed. In particular, we evaluated the hyperparameters of an unsupervised learning pipeline comprised of different feature selection and dimensionality reduction strategies followed by clustering with the OPTICS algorithm (see Methods for more details about the hyperparameter space). After clustering, three representative structures were chosen. Based on the performance in the validation set, the optimal set of hyperparameters was chosen (Table S3, see Figures S4). The described strategy for identifying representative structures was then applied to all target complexes and we ended up with three representative complex structures for each effector.

Analogous to how we characterized the interfaces of the experimentally determined complex structures before, we performed the same analysis on the complex models. The overall FoldX interaction energies for the models are diverse, indicating that maybe some of the complexes are energetically unfavorable and would not form (Figure S5B). In general, the binding energies are lower than what would be observed in crystal structures, which is to be expected. There are one or two outliers with regards to FoldX binding energy, namely RASSF8 and PIK3C2B. In particular, RASSF8 is also showing an uncommon hotspot profile, with multiple unfavorable hotspots that are only appearing for this set of structures (Figure S5A). Based on this behavior, RASSF8 is excluded from further analysis.

The analysis of hotspots confirmed the already established hotspots. I36, Y40, D38 and E37 are the most commonly observed hotspots (Figure 3AB, Figure S5A). Interestingly, while I36, Y40, and E37 are exclusively favorable to the interaction with the effector protein, D38 seems to be also unfavorable in some of the structures (Figure 3B). The energetic diversity of the hotspot D38 was further investigated in the models. For this, two models were picked for

which D38 was a favorable hotspot in the alanine scan analysis (Figure 3C: RASSF1, Figure 3D: RGL3), and two models were picked for which it was unfavorable (Figure 3E: ARAP2, Figure 3F: RAPGEF3). Next, neighboring amino acids were analyzed. For favorable interactions, we were able to observe positively charged amino acids. On RASSF1, we find K216 and H217, whereas on RGL3, there is R630. These amino acids probably form strong interactions with the negatively charged D38 on Ras. In contrast, for the models where D38 comes up as an unfavorable hotspot in the interface, we observe an uncharged, mostly hydrophobic neighborhood.

Estimation of binding energies

One of the applications of the structural models that were generated was to use them for the estimation of binding energies. Since experiments measuring the binding energy between two proteins are experimentally very challenging and error-prone [30], we were implementing an *in silico* approach. Also, while FoldX is good at predicting energy changes to interaction energy or protein stability on mutation, the absolute interaction energies for protein complexes usually are not well correlated with experimental values [31]. Because of this, a supervised learning regression pipeline was built based on features extracted from the modelled structures and a collection of experimentally determined binding energies of different Ras-effector complexes (Table S5). From a combination of different regressors, feature selection procedures, and hyperparameters, the best approach was determined using a cross validation strategy (see methods). A support vector machine-based regressor (see hyperparameters in Table 4) performed best in cross-validation with an R2 score of 0.53. Then, the performance of the best approach was evaluated in an out-of-sample test set, where it achieved an R2 score of 0.77. The model was then used to predict the interaction energies for the complexes without prior experimental measurements (Figure 4, Table S5). The predicted binding affinities for our models range 4 orders of magnitude, between the highest predicted affinity for BRAF of 0.02 μM to PIK3C2B with the lowest predicted affinity of 588 μM. The highest binding effectors are quite well characterized (RAF family, PI3K, RASSF5, RIN1, RalGDS, AFDN [8]). As experimental measurements of the PI3K family members are difficult as the RBD is not easy to express and purify in isolation, it is noteworthy that we assign three of the good binding affinities to PI3K family members (PIK3CA, PIK3CD, PIK3CG). A big group of effectors has affinities in the range of 1 to 10 μM. For example, RASIP1 was previously in the “likely no binding” category [8], and is now predicted to have an affinity in complex with Ras of 2.7 μM. RAPGEF5 and RGL3 were previously in the “unknown” [8] category and have predicted

affinities of 9.2 and 5.6 μM, respectively. Another big group of effectors has affinities in the range of 10 to 100 μM. Especially the first one could be interesting for modulation of binding affinity by the piggy back mechanism [7].

Switch contributions to binding

Having Ras-effector structural models available allowed us to analyze the individual contributions of switch 1 and switch 2 to binding using the results from alanine scanning (similarly as done before for the X-ray Ras-effector structures). We find that generally most structures are dominated by favorable switch 1 contributions (Figure 5 and Figure S6). Switch 2 contributions are surprisingly small. We also predict more contributions from regions outside switch 1 and switch 2 as in the X-ray structures. For the two weak affinity binding groups (10 μM (K_d < 100 μM and (K_d > 100 μM) switch 1 contributions are in the range of 4 to 5 kcal/mol with small (~1 kcal/mol) contributions from switch 2 and remaining parts involved in interface formation. For the two strong affinity groups (K_d < 1 μM and 1 μM (K_d < 10 μM) switch 1 contributions increase to 6 to 9 kcal/mol with also increasing switch 2 contributions (1 to 2 kcal/mol). We also observed negative switch contributions (mainly for switch 1), indicating that these proteins are either not well predicted or non-binders. Indeed, all structures with negative switch energies are predicted to be weak binders.

Branch pruning analysis using Ras-effector model structures

Next, we were interested in exploring surface mutations on Ras that would selectively influence the binding to some, but not all effectors. Both enhancing and inhibiting mutations are of interest. This could enable the engineering of the Ras effector system to respond to stimuli in different ways and to study selective sets of effectors. We previously reported a framework for the identification and evaluation of so called 'branch pruning' mutations [21]. Since our protein is interacting with a many different effectors at the same time through the same interface it will be quite unlikely to identify mutations that selectively target only one protein. Instead, it is more likely that we will identify mutations that enhance in interaction with some proteins while inhibiting some others.

Figure 6 shows a heatmap of all identified mutations of interest and their effect on all effectors. Some interesting mutations to highlight are mutations around I36, that are almost exclusively unfavorable while affecting almost every structure. D37 mutation are more selective and also exclusively disruptive. D38 is mixed, as our analysis of the hotspot already indicated. This is probably the best point to disrupt the system. Y40, interestingly, while being

a ubiquitous hotspot, is not a good spot for engineering the interface because mutations seem to affect protein stability (compare with Figure S7A). Also of interest is that we detect both mutations that increase and disrupt binding (Figure S7B).

Assessment of rewiring of Ras-effector interaction on a systems level

Finally, we want to evaluate the behavior of the Ras effector system based on our predicted binding energies and based on the introduction of different branch pruning mutations. For this means, we went back to our mathematical model of the Ras-effector system that incorporated affinities and high-quality proteomics data in 29 human tissues [7]. Here, all 29 tissue systems were simulated at 20% and 90% GTP load on Ras. We are using exclusively the predicted binding energies for this (except RASSF8, see methods). Overall, the results for the systems without a branch pruning mutation are comparable to previous findings (Figure 7A; [7]). The Raf family members ARAF, BRAF and RAF1 dominate the binding profile in complex with Ras. Other effectors that are in high amount in complex with Ras in at least one of the 29 tissues are RGL2, RASSF7, RASSF5, RASIP1, RALGDS, PIK3CD, PIK3CA, and AFDN.

Expanding on this, we introduced our branch pruning mutations to the mathematical model. In total, there are 200 interface mutations that do not significantly affect overall protein stability but affect binding to at least one of the effectors. Some of the branch pruning mutations are able to dramatically change the system in all tissues, as can be seen from the example of D38A (Figure 7B). With a single interface mutation, almost all RAF binding is quenched, and other effectors start to compete for the binding. Interestingly, which effectors come up depends on the tissue.

Next, we explored to which extend, for a specific effector, it is possible to modulate its binding. For this, we visualized the possible changes to the relative amount of effector bound to Ras across all systems tested (Figure 7C). On the side of proteins that can be negatively influenced, mostly the high-affinity binders such as RAF family proteins show up, which is to be expected. Some effectors cannot be influenced by the branch pruning mutations, either because they are energetically not affected or because their concentrations in any of the 29 tissues do not leave them in a position to compete for binding. Examples for this would be RADIL or the TIAM family proteins. The effectors with the highest propensity to have their binding enhanced are AFDN, RADIL, SNX27, RASSF5.

Finally, we investigated whether there are recurring states that the modelled system assumes and whether these states are dependent on Ras-GTP load, the tissue, or interface mutation. To this end, we applied uniform manifold approximation and projection (UMAP) to

our systems to transform a high-dimensional space of absolute and relative effector binding into a 2D space. Then, we used OPTICS to identify areas of high density in this 2D plane and assigned them into 24 distinct clusters as well as outliers outside the clusters (Figure 8A). For each cluster, we picked three of the systems at random and visualized their relative effector binding (Figure S8). Each of these clusters belongs to a different "state" that the Ras-effector system can be rewired into, with systems belonging to a specific state showing similar trends in effector binding. Many of the systems are dominated by ARAF binding, as it would be expected, but even for these there are differences in the secondary effectors. To understand what the attributes of different "states" of the systems are, two of them were picked and investigated for Ras-GTP load, tissue composition, and interface mutation status (Figure 8BC). We find that there are different ways a distinct state of the system can come together: for the state analyzed in Figure 8B, we can see that it is composed of many different tissues, but only a handful of interface mutations, most of them D38 mutations. This indicates that this state can be reached from many different tissues by a specific, recurring set of mutations. In contrast, the state analyzed in Figure 8C is entirely composed of a single tissue (lymph node) with many different mutations, indicated that this state can only be reached by a specific tissue. Interestingly, both states analyzed are diverse in terms of Ras-GTP load. To conclude, we identify 24 distinct states of the Ras-effector systems and show that there are different mechanisms on how these states are formed.

In summary, based on a large set of Ras-effector models, we predicted Ras branchegetic mutations and evaluate their binding in a tissue-specific Ras competition model. The Ras mutations, once introduced into cells or tissues, can be used as a tool to probe the contribution of specific effector pathways to an output or cellular phenotype.

Discussion

In this work we have shown a complete structural reconstruction of Ras and the RBDs of its effectors based on state-of-the-art technology. We used these structural models to investigate the workings of the interface between Ras and its effectors, as well as to search for and identify potential branch pruning mutations on Ras that would alter the behavior of the underlying system. We analyzed the effects on the steady state of the Ras effector system.

Recent advances in structural modelling, mostly by the development and release of AlphaFold2 and similar algorithms have pushed the field of structural bioinformatics forward. This development was crucial for the quality of this study. While a normal homology modelling

pipeline would have been successful, the inclusion of both aspects of AlphaFold2 models was essential for the quality of the results. On the one hand, finding additional potential complex templates diversified the possible configurations of the interface that we could use to generate our models. On the other hand, having high quality templates of the RBDs enabled us to improve the structural predictions. Interestingly, with all the advancements that AlphaFold2 brought, this combined AlphaFold2/homology modelling approach yielded more consistent and better results for this particular question.

There is a growing body of literature about how Ras dimerizes or forms multimers, or interacts with the membrane to modulate the signaling. All these aspects have been deliberately left out for this approach. The essence of the interaction of Ras and its effectors is the binary protein-protein interaction between the Ras switch regions and the RBD. This common feature was the focus of this work, and we believe that other factors such as dimer/multimerization of Ras, the composition of the membrane, etc., are only modulators for this interaction.

By creating a complete structural system, we were able to investigate and understand the interactions of Ras with its effector proteins on a different level. Crucially, it enabled us to analyze how *in silico* mutations of the system could affect its behavior. This is an interesting approach for a lot of different systems, not just the Ras effector system. However, there are certainly challenges. The prediction and verification of a protein-protein interface can be very complicated, and the techniques for modelling protein-protein interactions are not sufficiently developed to easily translate that approach to a larger scale. Probably the main reason why it was possible to construct this structural system for the Ras-effector system was because it is a conserved domain-domain interaction between homologue proteins. Although the sequence identity of the RBD sequences is not well preserved anymore, the structural fold of these domains has been preserved. Additionally, there is also the preserved mode of binding by the formation of the intermolecular β-sheet. These factors were favorable for the construction of the structural system and would need to be addressed if this approach were to be taken to a higher scale. Recent publications are already able to work on system-wide structural prediction for interactions [32]. These approaches are very promising for the structural characterization of known protein complexes and can identify high-confidence novel interactions as well. However, for the exhaustive characterization of a full system, especially with transient interactions, the distinction between true and false positives seems to remain challenging.

Finally, we hope that our structural and systems analysis of the branch pruning interface mutations will enable interesting experimental setups that study different downstream pathways from Ras. Some of the more promising candidates are AFDN, RADIL, SNX27, RASSF5. The

downstream pathways of Ras that are best characterized are the RAF-MEK-ERK signaling pathway and the PI3K-AKT signaling pathway. However, some of the other proteins might play a role in a physiological or pathophysiological context as well. AFDN is essential for the organization of adhesion between cells [33], function that is often impaired in cancer [34]. RADIL is also linked to cell adhesion, and recent data showed that knockdown of was linked to decreased cell proliferation and invasion [35]. SNX27 is also part of signaling pathways that link to cell adhesion and barrier function [36]. RASSF5 is a tumor suppressor and has been shown to inhibit growth and invasion and to induce apoptosis [37]. Importantly, all branchegetic mutations were studied on a systems level using a tissues specific Ras competition model. These models can easily be adapted for specific cell systems of interest, provided that estimates for Ras and effector abundances are available. Altogether, this work contributed to a quantitative understanding of a key cellular hub protein – Ras.

Acknowledgements

The authors would like to thank Cian D'Arcy, Thomas Sevrin, Simona Catozzi, Swathi Ramachandra Upadhya, and Hiroaki Imoto for discussions and/or critical reading of the manuscript.

This work is part of the research program "Quantitative and systems analysis of (patho)physiological signaling networks" [16/FRL/3886], which is financed by Science Foundation Ireland (SFI) to C.K.

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number [16/FRL/3886]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

Conceptualization, P.J. and C.K.; Methodology, P.J. and C.K.; Software, P.J; Investigation, P.J.; Data Curation, P.J.; Writing – Original Draft, P.J. and C.K.; Writing – Review & Editing, P.J. and C.K.; Visualization, P.J. and C.K.; Supervision, C.K.; Funding Acquisition, C.K

Declaration of interests

The authors declare no competing interests.

STAR★Methods

Lead contact

Further information and requests for resources should be directed to Philipp Junk and will be fulfilled by the lead contact, Philipp Junk (philipp.junk@ucdconnect.ie)

Material availability

This study did not generate new unique reagents.

Data and Code availability

The code and generated data have been deposited on Zenodo: https://doi.org/10.5281/zenodo.7188727

METHOD DETAILS

Experimentally determined Ras-effector complex structures

Structures were downloaded from the PDB. In the case where multiple models were available, the best one by MolProbity score was selected [27 28]. The PDB files were processed using pdb-tools so that all GTP and Mg2+ annotations were in the expected format [38]. The list of used template structures can be found in Table S1.

Interface characterization

Hotspots residues were determined by FoldX *in silico* Alanine scan [39 31 20 40 41]. In detail, each residue on both RAS and the effector was mutated to alanine and the *in silico* change of binding energy ΔΔG was determined as such:

$$\Delta\Delta G\ alascan = \Delta G(mut) - \Delta G(wt)$$

A positive ΔΔG value indicates that the mutated residue is involved favorably in the interaction, whereas a negative ΔΔG value indicates that the mutation to alanine improved the interaction between the two proteins. The standard error for ΔΔG values in FoldX is around +- 0.8 kcal/mol. In order to identify the most important residues for the respective interaction, a ΔΔG cut-off of 1.6 kcal/mol was chosen for the investigation of crystal structures. Since the energies are systematically lower in the modelled complex structures, a cut-off of 1.2 kcal/mol was chosen for those.

Based on the alanine scan results, the contribution of the two major functional regions in the RAS interface [42], switch 1 (residues 20-42) and switch 2 (residues 56-76) to the interaction were determined. For each functional region and the remainder of Ras, the ΔΔG values were filtered by abs(ΔΔG) > 0.8 kcal/mol and then summed up. The definition of the switch regions used here is more generous than what is normally used in the literature, and it would be probably more correct to label them as "switch-influenced" regions. These residue ranges aim to capture the regions in the interface that are affected by the movement of switch 1 and 2 during the transition from the inactive GDP bound state to the active GTP bound state. The conserved intra-molecular β-sheet between the β2 sheet on Ras with the β2 sheet on the effector is evaluated by measuring the differences in inter-molecular distances between the β sheets in the experimentally solved complex structures and comparing them to the ones in a structure of interest. This measurement has been named Measured Alignment Error (MAE) in this manuscript:

$$MAE(res_{eff}, res_{ras}) = \frac{\sum_{i=1}^{n} \sqrt{(dist_{model} - dist_{ref})^2}}{n}$$

with $dist_{model}$ being the Euclidean distance between two residues in the model of interest; $dist_{ref}$ being the Euclidean distance between the residue on Ras and the corresponding residue in the reference structure, as determined by structural alignment using TMalign [43]; and n being the number of reference structures. The references used were the X-ray complex templates found in the PDB (Table S1). When the MAE is calculated for X-ray complex templates, the comparison with itself is removed from the calculation.

AlphaFold2 determined single structures for RBDs

For each protein containing one or more potential RBD sequences as identified in [6], the full structure was obtained from the AlphaFold Protein Structure database [14 15] (Release 1, accessed September 2021). The sequences of the RBDs were obtained from Pfam [44] and UNIPROT [45]. The part of the structures that correspond to RBD sequences was subsequently extracted from the AlphaFold2 structures. The list of used templates as well as information about the extracted RBD domains can be found in Table S2.

AlphaFold2 determined complex structures: generation and selection

AlphaFold2-multimer/ColabFold (version 1.3.0) was run on a local computer [46 14 47]. Multiple sequence alignments were generated from MS2seq [48 49 50 51]. Complex models were generated

with HRAS as interaction partner. For each target, five models were generated with additional template search and five without additional template search [46 52]. The resulting model were relaxed as per ColabFold defaults [46 53]. For each model, several metrices were evaluated. Firstly, AlphaFold2's Predicted Alignment Error (PAE) was taken into consideration. PAE is an estimation of the error of pairwise distances between residues, that can be used to assess how confident AlphaFold2 is in the inter-domain arrangement of its models. Secondly, FoldX [39 31 40 41] interaction energies were determined for the complexes as described above. Finally, the expected orientation of the inter-molecular β sheets was evaluated by MAE. The best model by MAE was selected for each target, and then all complex structures with a MAE > 1A were removed.

Homology modelling pipeline

Alignment generation, model generation and model evaluation were performed using the homelette homology modelling interface [22]. Inputs to the homology modelling pipeline were complex structures of Ras in complex with some effectors, either experimentally determined or selected from AlphaFold2 complex predictions, as well as AlphaFold2 models of all RBDs of interest. For each target, all combinations of the RBD template with all complex templates were used to generate 300 models of the target in complex with HRAS each. The different sources for the complex templates were run separately with slight differences in the modelling procedure. All sequence alignments for RAS were generated with M-Coffee [54,55]. For complex templates of experimental origin, TMalign [43] was used to generate a structure-based sequence alignment based on the RBD in the complex template and the single RBD of the target structure. Then, with those two templates as inputs, models were generated. For complex templates of *in silico* origin, structure-based sequence alignments were generated with TMalign [43] as described. As an additional template a HRAS single structure (5P21 [3]) was used in the modelling process since the AlphaFold2-generated complex templates, in contrast to the complex templates of experimental origin, do not have information about the important heteroatoms GTP and MG2+ in the Ras part of the structure. Models were generated using modeller [23 24] with the altMOD extension [25]. All models generated were evaluated using QMEAN [26 56], MolProbity, [27 28] and SOAP [29] potentials. A combined score was determined based on borda count as such:

$$combined\ score(X) = \sum_{i=1}^{m} n - rank_i(X)$$

For an observation X with i…m being a collection of evaluation criteria and n the total number of observations. A structure with high borda score is a structure that performs well across all metrices.

For each of the different sources of complex templates (experimental or *in silico*), 300 models were selected in a first selection step based on the combined score. These 600 models per target were then further analyzed using FoldX. *In silico* interaction energies were determined and *in silico* alanine scan was performed (see Interface Characterization).

In addition to generating models for unknown targets, a set of validation models based on the experimentally solved models were generated as well. For each of the seven PDB complex structures, 300 models were generated. Inputs were restricted so that the structure to be modelled would not be used as a complex template, but only the remaining six experimentally derived complex templates. For each validation target, 300 models were selected as described above.

Using the results from the analysis with FoldX, representative structures were selected based on an unsupervised learning workflow. As the clustering algorithm of choice, Ordering Points To Identify the Clustering Structure (OPTICS) [57] was used, as implemented in scikit-learn [58]. OPTICS is a density-based clustering method, which unlike the more popular k-means clustering does not require a manually set input of the number of clusters. Additionally, OPTICS is able to label data points as outliers. Several approaches to feature selection and/or feature engineering, hyperparameters of OPTICS, as well as methods for selecting representative structures from clusters were evaluated against the set of validation models (Table 3). To select the best combination of hyperparameters, for each combination, the *in silico* alanine scan results of the representative structures were correlated to the results of corresponding PDB structures, and the combination of hyperparameters with the most stable performance across all validation sets (by minimum z-score of the correlation against the PDB structure for all 7 validation sets) was chosen. The best combination of hyperparameters is highlighted in (Table 3).

Estimation of binding energies

Supervised learning based on a number of FoldX-derived features was used in order to estimate binding energy of complex models. The features consisted of the FoldX interaction energy, the energy contribution of switch 1, 2 and the remainder of the Ras protein interface (see Interface characterization), and the ΔΔG values for hotspot residues on Ras (cut off 1.2 kcal/mol). All

features were standardized and highly intercorrelated features were removed. Data preparation was performed in R.

The experimental measurements of the dissociation constant between Ras-effector complexes were collected from several publications (see Table S5, also available as supplementary data). Effectors that were experimentally determined as non-binders were removed from the data set due to uncertainty how to encode them with the varying technical limitations on detectable binding energies at the time of their publication. Also, the models generated for RASSF3 seem to be outliers with regards to FoldX interaction energy (see Figure S5B) and were therefore removed from the prediction. A test set was manually chosen from the available experimental measurements to cover the full spectrum of experimentally determined interaction energies. At the end, this gave us a training set with 51 observations (17 * 3 models) and a testing set with 12 observations (4 * 3 models).

Supervised learning was performed in scikit-learn [58] using different combinations of feature selection algorithms and regressors. Feature selection was performed using either f-regression or mutual information as implemented by scikit-learn. Regression was evaluated for different algorithms with various hyperparameter spaces (see Table S4). All combination of feature selection and regression were evaluated using Group-K-Fold cross validation within the training data, with k=5 and the groups corresponding to the three structural models chosen for each target. Models were evaluated using R2 score. The best performing combination was trained on the whole training data and evaluated against the test data. Finally, this model was used to predict the binding energies for the complex structures without experimental data.

Branch pruning

FoldX was used to evaluate the effect of interface mutation on the binding energies in the modelled complex structures. As previously described, [21] the FoldX command PSSM was used to evaluate the changes in binding energy, and the FoldX command PosScan was used to evaluate if mutations impacted the stability of RAS.

All mutations that were destabilizing HRAS in either of two structures (3TGP and 5P21) were removed from the analysis (cutoff 1.6 kcal/mol). Then, between the three models for each target structures, it was checked if a mutation had a noticeable impact (> +- 1kcal/mol) in at least two of the three structures. If so, the changes in binding energy for all models above/below the cutoff were averaged.

Systems analysis

A mathematical model of the Ras-effector system was set up as previously described [6 7]. The model is based on classic ligand-receptor kinetics according to the assumption of conservation of mass. A system of ordinary differential equations was set up and steady states were calculated as described. The reactions are expressed as such:

$$R + E_i \overset{(k_i, k_{-i})}{\longleftrightarrow} RE_i$$

With R representing the molar concentration of Ras, E_i the molar concentration of an effector of the set of $i \in (1, 2, \dots, 54)$ effectors (all modelled proteins, except for proteins from the Ubiquitin family) and RE_i the molar concentration of a Ras-effector complexes. The complex is formed at rate k_i and dismantled at rate k_{-i}. These rates define the dissociation constant:

$$K_{d_i} = \frac{k_{-i}}{k_i}$$

Due to the assumption of mass conservation,

$$R_{tot} = R + \sum_{1}^{54} RE_i$$

and

$$E_{tot_i} = E_i + RE_i$$

the system to solve therefore is:

$$R_{tot} = R + \sum_{1}^{54} RE_i$$

$$E_{tot_i} = E_i + RE_i$$

$$K_{d_i} = \frac{(R * E_i)}{RE_i}$$

For the set of $i = \{1,2, \dots, 54\}$ and can be numerically solved for RE_i. The system was solved using SciPy [59].

The concentrations for the species in the model were taken from the supplementary data of [7], in which molar concentrations were derived from high-quality proteomics data set of 29 different human tissues [60]. The concentration of Ras proteins (HRAS, NRAS, KRAS) were pooled together and then multiplied with a loading factor to take into account the balance between active (GTP bound) and inactive (GDP bound) Ras. This loading factor was set to 0.2 for a normal RAS system, and 0.9 for a system hyperactivated by an oncogenic hotspot mutation.

The binding affinities were taken from the predicted binding affinities (see Estimation of binding affinities), with the exception of RASSF3 for which we are not confident in our

structural predictions. An experimentally determined binding affinity for RASSF3 was substituted from [61]. Additional sets of binding affinities based on the branch pruning analysis were evaluated as well. For this, the predicted binding affinities were adapted by the ΔΔG values from the branch pruning analysis. For the case that a system with hyperactivated Ras was considered, it was made sure that common oncogenic hotspots such as G12, G13 and Q61 do not influence the binding energies for any effector.

To gain an overview over all solved system, absolute and relative effector bindings were transformed using UMAP [62] and visualized. OPTICS clustering [57] was performed on the UMAP transformed data (parameters: min_sample = 25, min_cluster_size=200).

Data analysis and visualization

All data analysis, unless otherwise noted was performed in R using the tidyverse environment [63,64]. Visualizations were generated using ggplot2 [65]. Visualizations of protein structures were generated using PyMol [66].

Figure legends

Figure 1. Noticeable features of the Ras-RBD interfaces

(A) Overview of interface features for the Ras-RAF1 interface. Highlighted are the intramolecular β-sheet alignment, the assembly of the Ras interface by the switch regions and energetic hotspots in the interface.

(B) β-sheet alignment for experimentally determined complexes (MAE). A lower value indicates a more similar alignment of the intramolecular β-sheets compared to other crystal structures.

(C) FoldX alanine scan hotspots on RAS for experimentally determined complex structures. The color scale is confined to the limits [-1.6, 1.6]. Positive ΔΔG values indicate a loss in binding energy upon alanine mutation, which reflects that the respective amino acid contributes to binding.

(D) Energetic contributions of functional regions based on alanine scan analysis.

Figure 2. Overview over the homology modelling pipeline

The principal homology modelling steps are shown from left to right with the steps of “Target Identification”, “Template Identification”, “Alignment Generation”, “Complex Model Generation”, “Evaluation & Selection 1”, and “Evaluation & Selection 2”. In the “Validation” process, the homology modelling pipeline is run on a set of 7 RBD domains from known x-ray

structures. The “Production” process describes the generation of 61 domains (RBD and Ubiquitin super-fold domains as controls).

Figure 3. Energy hotspots in modelled structures

(A) Heatmap of FoldX alanine scan averaged from the three representative structures on each target. The color scale is confined to the limits [-1.6, 1.6]. Hotspot residues with a ΔΔG >= 1.2 or <= -1.2 kcal/mol were marked. Positive and negative ΔΔG values indicate a destabilization and stabilization of the RAS effector models by alanine mutation at the indicated position, respectively.

C-F) Local neighborhood of D38 in structures where D38 is a favorable hotspot (C: RASSF1, D: RGL3) or an unfavorable hotspot (E: ARAP2, F: RAPGEF3). Ras and the effector structures are visualized in blue and green, respectively. Polar interactions between charged amino acids are indicated with dashes.

Figure 4. Results of the affinity prediction for all Ras-effector complex structures

Visualization of predicted binding affinities. The three representative structures for each target are visualized as black dots, with the averaged affinity (based on averaged energy) is visualized as a red cross.

Figure 5. Switch contributions for the summed-up energy contributions grouped by their predicted binding affinity

(A) Energy contributions were separately calculated for switch 1 switch or the rest of the Ras interface by summing up energy energies from the *in silico* alanine scan analysis. Complexes were grouped based on their predicted binding affinities into four groups.

(B – E) Examples of the switch contributions for each of the four groups. Visualized are BRAF (panel B), AFDN_1 (panel C), ARHGAP20 (panel D), and PIK3C2B (panel E).

Figure 6. Energetic characterization of RAS interface mutations that affect effector binding

The color scale is confined to the limits [-3.2, 3.2]. Hotspot residues with a ΔΔG >= 1 or <= -1 were marked. Negative ΔΔG values indicate an increase in binding energy compared to the WT interface, and positive ΔΔG values indicate a decrease in binding energy compared to the WT interface.

Figure 7. Ras-effector interaction rewiring at a systems level

(A-B) Effects of interface mutation on the Ras-effector system. Heatmap of effectors bound to Ras at 20% Ras-GTP load in 29 tissues in (panel A) WT interface and in (panel B) with the effects of a D38A mutation.

(C) Visualization of change in relative binding compared to WT system for each effector across all tested branch pruning mutations in all 29 tissues at 20 % or 90 % Ras-GTP load. Positive values indicate a relative increase in bound effector to Ras.

Figure 8. Ras-effector systems are rewired into distinct states

(A) UMAP transformation of all simulated systems. Similar systems were identified using OPTICS clustering. 24 identified clusters are colored in rainbow colors, with systems classified as outliers colored in grey.

(B) and (C) Characterization of two identified state clusters in terms of Ras-GTP load, tissues, interface mutations, and most representative effectors. For the tissue and mutation pie charts, all groups that were smaller than 5% of the total observations were collapsed into the “Other” group.

1. Gasper, R., and Wittinghofer, F. (2019). The Ras switch in structural and historical perspective. Biol Chem *401*, 143-163. 10.1515/hsz-2019-0330.
2. Pai, E.F., Kabsch, W., Krengel, U., Holmes, K.C., John, J., and Wittinghofer, A. (1989). Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation. Nature *341*, 209-214. 10.1038/341209a0.
3. Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W., and Wittinghofer, A. (1990). Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. Embo j *9*, 2351-2359. 10.1002/j.1460-2075.1990.tb07409.x.
4. Kiel, C., and Serrano, L. (2006). The ubiquitin domain superfold: structure-based sequence alignments and characterization of binding epitopes. J Mol Biol *355*, 821-844. 10.1016/j.jmb.2005.10.010.
5. Simanshu, D.K., Nissley, D.V., and McCormick, F. (2017). RAS Proteins and Their Regulators in Human Disease. Cell *170*, 17-33. 10.1016/j.cell.2017.06.009.
6. Ibáňez Gaspar, V., Catozzi, S., Ternet, C., Luthert, P.J., and Kiel, C. (2021). Analysis of Ras-effector interaction competition in large intestine and colorectal cancer context. Small GTPases *12*, 209-225. 10.1080/21541248.2020.1724596.
7. Catozzi, S., Halasz, M., and Kiel, C. (2021). Predicted 'wiring landscape' of Ras-effector interactions in 29 human tissues. NPJ Syst Biol Appl *7*, 10. 10.1038/s41540-021-00170-0.
8. Kiel, C., Matallanas, D., and Kolch, W. (2021). The Ins and Outs of RAS Effector Complexes. Biomolecules *11*. 10.3390/biom11020236.
9. Catozzi, S., Ternet, C., Gourrege, A., Wynne, K., Oliviero, G., and Kiel, C. (2022). Reconstruction and analysis of a large-scale binary Ras-effector signaling network. Cell Commun Signal *20*, 24. 10.1186/s12964-022-00823-5.
10. Kiel, C., Verschueren, E., Yang, J.S., and Serrano, L. (2013). Integration of protein abundance and structure data reveals competition in the ErbB signaling network. Sci Signal *6*, ra109. 10.1126/scisignal.2004560.
11. Kholodenko, B.N., Hoek, J.B., and Westerhoff, H.V. (2000). Why cytoplasmic signalling proteins should be recruited to cell membranes. Trends Cell Biol *10*, 173-178. 10.1016/s0962-8924(00)01741-4.
12. Kiel, C., Foglierini, M., Kuemmerer, N., Beltrao, P., and Serrano, L. (2007). A genome-wide Ras-effector interaction network. J Mol Biol *370*, 1020-1032. 10.1016/j.jmb.2007.05.015.
13. Kiel, C., and Serrano, L. (2007). Prediction of Ras-effector interactions using position energy matrices. Bioinformatics *23*, 2226-2230. 10.1093/bioinformatics/btm336.
14. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589. 10.1038/s41586-021-03819-2.
15. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res *50*, D439-d444. 10.1093/nar/gkab1061.
16. Gimple, R.C., and Wang, X. (2019). RAS: Striking at the Core of the Oncogenic Circuitry. Front Oncol *9*, 965. 10.3389/fonc.2019.00965.
17. Moore, A.R., Rosenberg, S.C., McCormick, F., and Malek, S. (2020). RAS-targeted therapies: is the undruggable drugged? Nat Rev Drug Discov *19*, 533-552. 10.1038/s41573-020-0068-6.
18. Stephen, A.G., Esposito, D., Bagni, R.K., and McCormick, F. (2014). Dragging ras back in the ring. Cancer Cell *25*, 272-281. 10.1016/j.ccr.2014.02.017.
19. Zhong, Q., Simonis, N., Li, Q.R., Charloteaux, B., Heuze, F., Klitgord, N., Tam, S., Yu, H., Venkatesan, K., Mou, D., et al. (2009). Edgetic perturbation models of human inherited disorders. Mol Syst Biol *5*, 321. 10.1038/msb.2009.80.

697 20. Kiel, C., and Serrano, L. (2014). Structure-energy-based predictions and network modelling of
698 RASopathy and cancer missense mutations. Mol Syst Biol *10*, 727. 10.1002/msb.20145092.
699 21. Junk, P., and Kiel, C. (2021). Engineering of Biological Pathways: Complex Formation and
700 Signal Transduction. Methods Mol Biol *2315*, 59-70. 10.1007/978-1-0716-1468-6_4.
701 22. Junk, P., and Kiel, C. (2021). HOMELETTE: a unified interface to homology modelling
702 software. Bioinformatics *38*, 1749-1751. 10.1093/bioinformatics/btab866.
703 23. Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial
704 restraints. J Mol Biol *234*, 779-815. 10.1006/jmbi.1993.1626.
705 24. Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. Curr
706 Protoc Bioinformatics *54*, 5.6.1-5.6.37. 10.1002/cpbi.3.
707 25. Janson, G., Grottesi, A., Pietrosanto, M., Ausiello, G., Guarguaglini, G., and Paiardini, A.
708 (2019). Revisiting the "satisfaction of spatial restraints" approach of MODELLER for protein
709 homology modeling. PLoS Comput Biol *15*, e1007219. 10.1371/journal.pcbi.1007219.
710 26. Benkert, P., Tosatto, S.C., and Schomburg, D. (2008). QMEAN: A comprehensive scoring
711 function for model quality assessment. Proteins *71*, 261-277. 10.1002/prot.21715.
712 27. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J.,
713 Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure
714 validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12-21.
715 10.1107/s0907444909042073.
716 28. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V.,
717 Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference
718 data for improved all-atom structure validation. Protein Sci *27*, 293-315. 10.1002/pro.3330.
719 29. Dong, G.Q., Fan, H., Schneidman-Duhovny, D., Webb, B., and Sali, A. (2013). Optimized
720 atomic statistical potentials: assessment of protein interfaces and loops. Bioinformatics *29*,
721 3158-3166. 10.1093/bioinformatics/btt560.
722 30. Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P.P., and Herschlag, D. (2020). How to measure
723 and evaluate binding affinities. Elife *9*. 10.7554/eLife.57264.
724 31. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX
725 web server: an online force field. Nucleic Acids Res *33*, W382-388. 10.1093/nar/gki387.
726 32. Humphreys, I.R., Pei, J., Baek, M., Krishnakumar, A., Anishchenko, I., Ovchinnikov, S., Zhang,
727 J., Ness, T.J., Banjade, S., Bagde, S.R., et al. (2021). Computed structures of core eukaryotic
728 protein complexes. Science *374*, eabm4805. 10.1126/science.abm4805.
729 33. Citalán-Madrid, A.F., García-Ponce, A., Vargas-Robles, H., Betanzos, A., and Schnoor, M.
730 (2013). Small GTPases of the Ras superfamily regulate intestinal epithelial homeostasis and
731 barrier function via common and unique mechanisms. Tissue Barriers *1*, e26938.
732 10.4161/tisb.26938.
733 34. Martin, T.A., and Jiang, W.G. (2009). Loss of tight junction barrier function and its role in
734 cancer metastasis. Biochim Biophys Acta *1788*, 872-891. 10.1016/j.bbamem.2008.11.005.
735 35. Choi, B.H., Kou, Z., Colon, T.M., Chen, C.H., Chen, Y., and Dai, W. (2021). Identification of
736 Radil as a Ras binding partner and putative activator. J Biol Chem *296*, 100314.
737 10.1016/j.jbc.2021.100314.
738 36. Zimmerman, S.P., Hueschen, C.L., Malide, D., Milgram, S.L., and Playford, M.P. (2013). Sorting
739 nexin 27 (SNX27) associates with zonula occludens-2 (ZO-2) and modulates the epithelial
740 tight junction. Biochem J *455*, 95-106. 10.1042/bj20121755.
741 37. Zhou, X.H., Yang, C.Q., Zhang, C.L., Gao, Y., Yuan, H.B., and Wang, C. (2014). RASSF5 inhibits
742 growth and invasion and induces apoptosis in osteosarcoma cells through activation of
743 MST1/LATS1 signaling. Oncol Rep *32*, 1505-1512. 10.3892/or.2014.3387.
744 38. Rodrigues, J., Teixeira, J.M.C., Trellet, M., and Bonvin, A. (2018). pdb-tools: a swiss army knife
745 for molecular structures. F1000Res *7*, 1961. 10.12688/f1000research.17456.1.
746 39. Guerois, R., Nielsen, J.E., and Serrano, L. (2002). Predicting changes in the stability of proteins
747 and protein complexes: a study of more than 1000 mutations. J Mol Biol *320*, 369-387.
748 10.1016/s0022-2836(02)00442-4.

40. Kiel, C., Serrano, L., and Herrmann, C. (2004). A detailed thermodynamic analysis of ras/effector complex interfaces. J Mol Biol *340*, 1039-1058. 10.1016/j.jmb.2004.05.050.
41. Kiel, C., and Serrano, L. (2009). Cell type-specific importance of ras-c-raf complex association rate constants for MAPK signaling. Sci Signal *2*, ra38. 10.1126/scisignal.2000397.
42. Vetter, I.R., and Wittinghofer, A. (2001). The guanine nucleotide-binding switch in three dimensions. Science *294*, 1299-1304. 10.1126/science.1062023.
43. Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res *33*, 2302-2309. 10.1093/nar/gki524.
44. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res *49*, D412-d419. 10.1093/nar/gkaa913.
45. UniProt, C. (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res *49*, D480-D489. 10.1093/nar/gkaa1100.
46. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat Methods *19*, 679-682. 10.1038/s41592-022-01488-1.
47. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., et al. (2022). Protein complex prediction with AlphaFold-Multimer. bioRxiv. 10.1101/2021.10.04.463034.
48. Mirdita, M., Steinegger, M., and Söding, J. (2019). MMseqs2 desktop and local web server app for fast, interactive sequence searches. Bioinformatics *35*, 2856-2858. 10.1093/bioinformatics/bty1057.
49. Mirdita, M., von den Driesch, L., Galiez, C., Martin, M.J., Söding, J., and Steinegger, M. (2017). Uniclust databases of clustered and deeply annotated protein sequences and alignments. Nucleic Acids Res *45*, D170-d176. 10.1093/nar/gkw1081.
50. Mitchell, A.L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M.R., Kale, V., Potter, S.C., Richardson, L.J., et al. (2020). MGnify: the microbiome analysis resource in 2020. Nucleic Acids Res *48*, D570-d578. 10.1093/nar/gkz1035.
51. Steinegger, M., Meier, M., Mirdita, M., Vöhringer, H., Haunsberger, S.J., and Söding, J. (2019). HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics *20*, 473. 10.1186/s12859-019-3019-7.
52. Berman, H., Henrick, K., and Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. Nat Struct Biol *10*, 980. 10.1038/nsb1203-980.
53. Eastman, P., Swails, J., Chodera, J.D., McGibbon, R.T., Zhao, Y., Beauchamp, K.A., Wang, L.P., Simmonett, A.C., Harrigan, M.P., Stern, C.D., et al. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol *13*, e1005659. 10.1371/journal.pcbi.1005659.
54. Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol *302*, 205-217. 10.1006/jmbi.2000.4042.
55. Wallace, I.M., O'Sullivan, O., Higgins, D.G., and Notredame, C. (2006). M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res *34*, 1692-1699. 10.1093/nar/gkl091.
56. Benkert, P., Biasini, M., and Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics *27*, 343-350. 10.1093/bioinformatics/btq662.
57. Ankerst, M., Breunig, M.M., Kriegel, H.-P., and Sander, J. (1999). Optics. ACM SIGMOD Record *28*, 49-60. 10.1145/304181.304187.
58. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-Learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

59. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods *17*, 261-272. 10.1038/s41592-019-0686-2.
60. Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D.P., Zecha, J., Asplund, A., Li, L.H., Meng, C., et al. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Mol Syst Biol *15*, e8503. 10.15252/msb.20188503.
61. Rezaei Adariani, S., Kazemein Jasemi, N.S., Bazgir, F., Wittich, C., Amin, E., Seidel, C.A.M., Dvorsky, R., and Ahmadian, M.R. (2021). A comprehensive analysis of RAS-effector interactions reveals interaction hotspots and new binding partners. J Biol Chem *296*, 100626. 10.1016/j.jbc.2021.100626.
62. Dorrity, M.W., Saunders, L.M., Queitsch, C., Fields, S., and Trapnell, C. (2020). Dimensionality reduction by UMAP to visualize physical and genetic interactions. Nat Commun *11*, 1537. 10.1038/s41467-020-15351-4.
63. Team, R.C. (2022). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
64. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. Journal of Open Source Software *4*. 10.21105/joss.01686.
65. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).
66. Schrodinger, LLC (2022). The PyMOL Molecular Graphics System, Version 2.5.

Figure 1

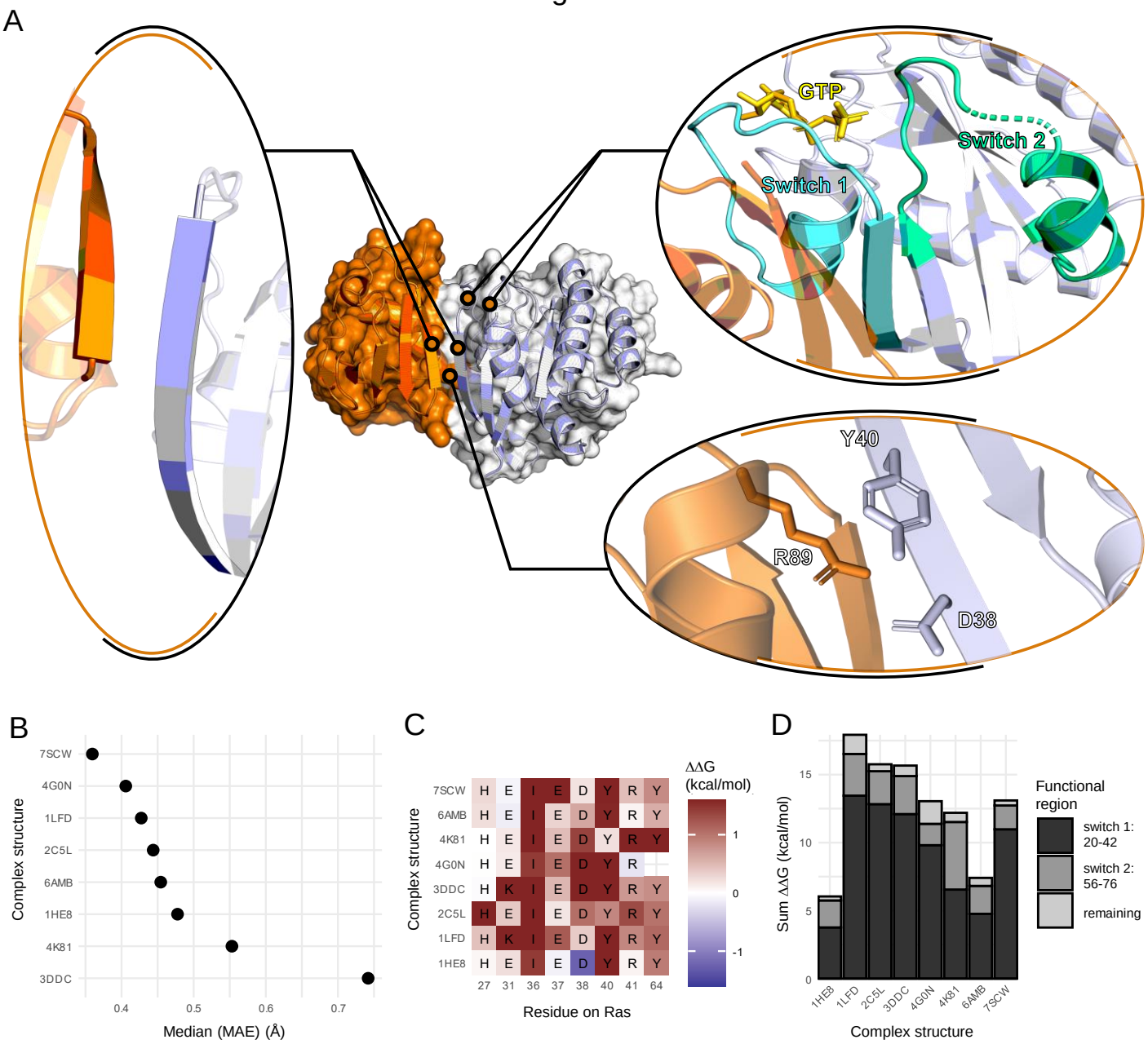

A
GTP
Switch 2
Switch 1
Y40
R89
D38
B
Complex structure
7SCW
4G0N
1LFD
2C5L
6AMB
1HE8
4K81
3DDC
0.4
0.5
0.6
0.7
Median (MAE) (Å)
C
Complex structure
7SCW H E I E D Y R Y
6AMB H E I E D Y R Y
4K81 H E I E D Y R Y
4G0N H E I E D Y R
3DDC H K I E D Y R Y
2C5L H E I E D Y R Y
1LFD H K I E D Y R Y
1HE8 H E I E D Y R Y
27 31 36 37 38 40 41 64
Residue on Ras
ΔΔG (kcal/mol)
1
0
-1
D
Sum ΔΔG (kcal/mol)
0
5
10
15
1HE8
1LFD
2C5L
3DDC
4G0N
4K81
6AMB
7SCW
Complex structure
Functional region
switch 1: 20-42
switch 2: 56-76
remaining

Figure 2

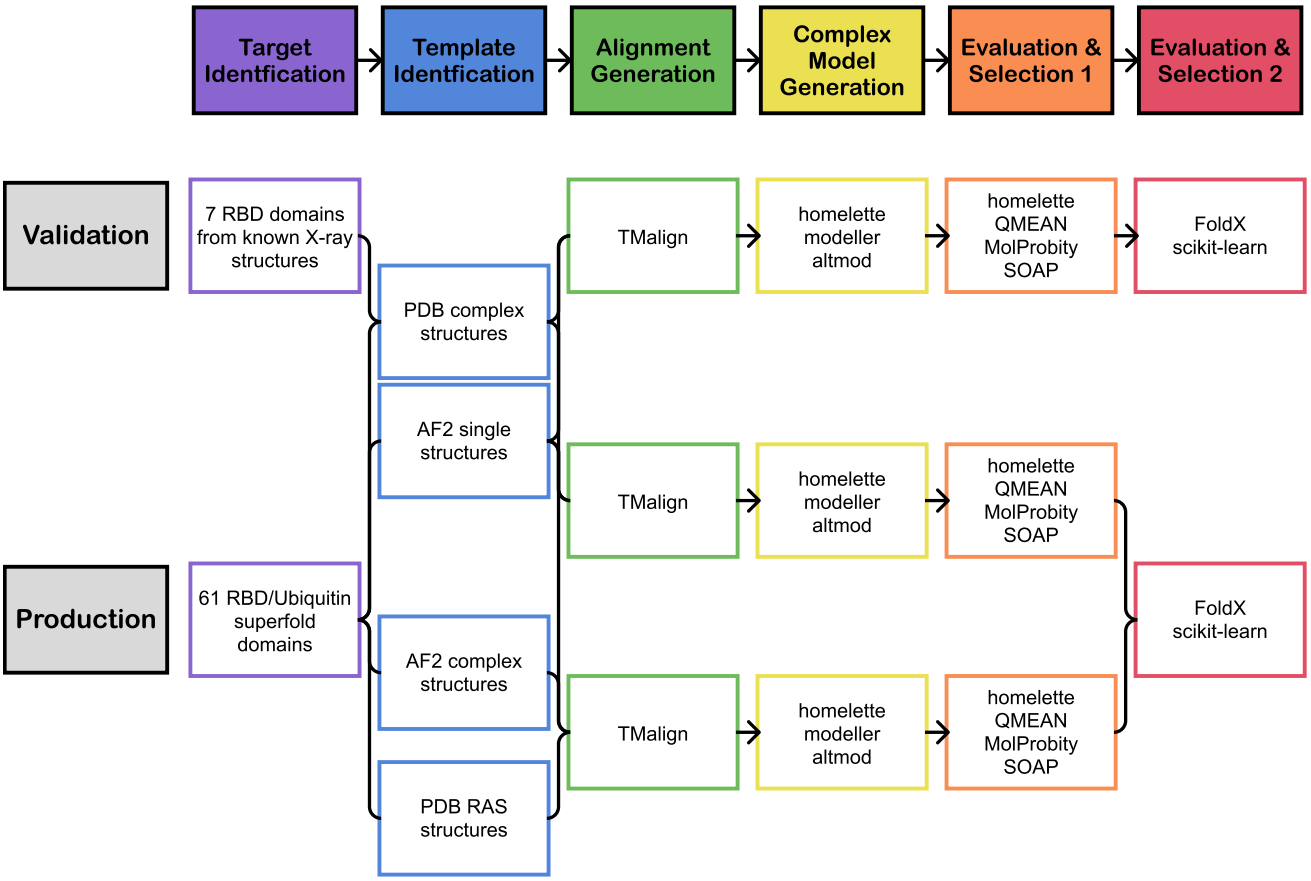

Target Identfication
Template Identfication
Alignment Generation
Complex Model Generation
Evaluation & Selection 1
Evaluation & Selection 2
Validation
Production
7 RBD domains from known X-ray structures
61 RBD/Ubiquitin superfold domains
PDB complex structures
AF2 single structures
AF2 complex structures
PDB RAS structures
TMalign
TMalign
TMalign
homelette modeller altmod
homelette modeller altmod
homelette modeller altmod
homelette QMEAN MolProbity SOAP
homelette QMEAN MolProbity SOAP
homelette QMEAN MolProbity SOAP
FoldX scikit-learn
FoldX scikit-learn

Figure 3

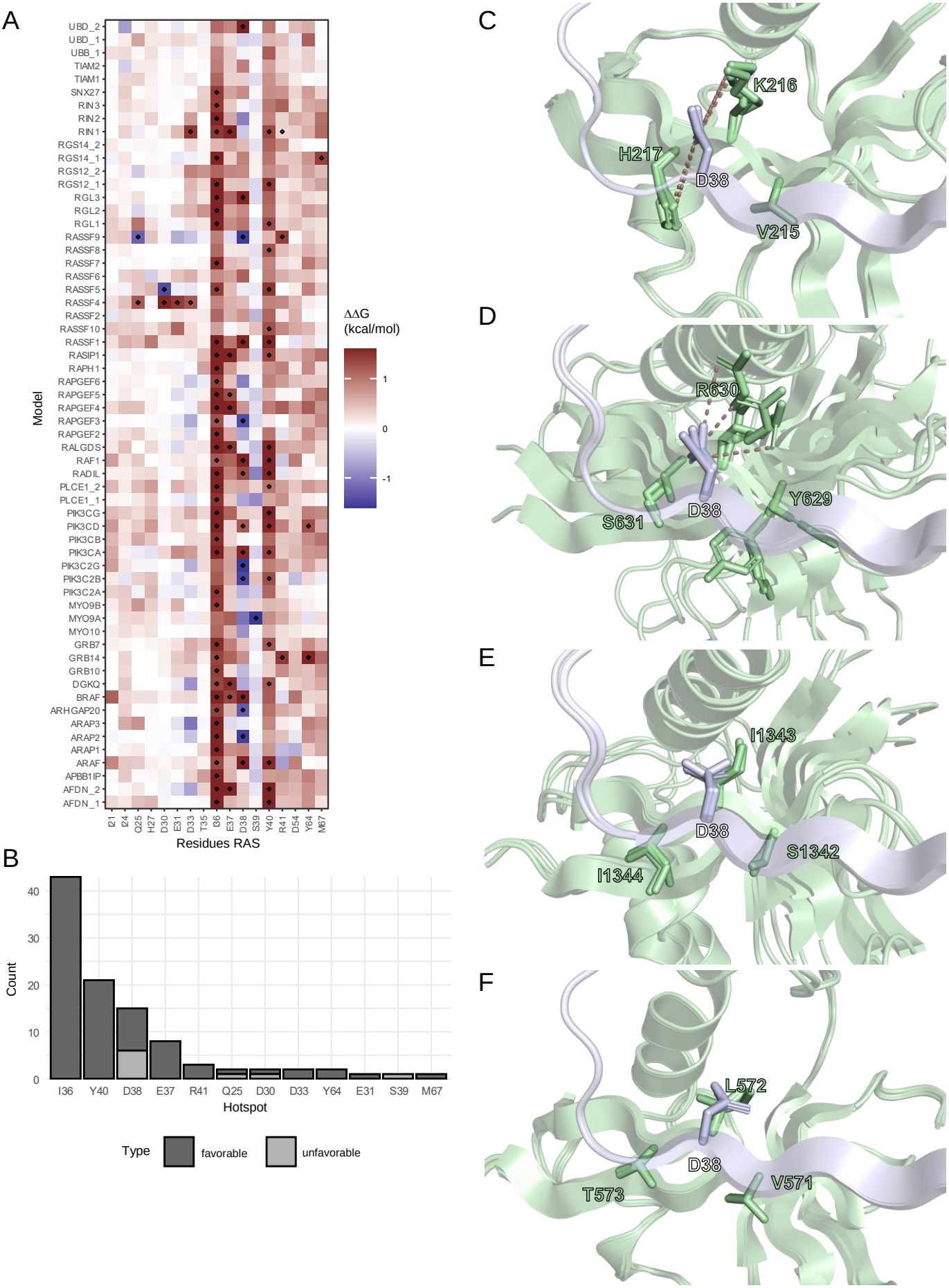

A
Model
UBD_2
UBD_1
UBB_1
TIAM2
TIAM1
SNX27
RIN3
RIN2
RIN1
RGS14_2
RGS14_1
RGS12_2
RGS12_1
RGL3
RGL2
RGL1
RASSF9
RASSF8
RASSF7
RASSF6
RASSF5
RASSF4
RASSF2
RASSF10
RASSF1
RASIP1
RAPH1
RAPGEF6
RAPGEF5
RAPGEF4
RAPGEF3
RAPGEF2
RALGDS
RAF1
RADIL
PLCE1_2
PLCE1_1
PIK3CG
PIK3CD
PIK3CB
PIK3CA
PIK3C2G
PIK3C2B
PIK3C2A
MYO9B
MYO9A
MYO10
GRB7
GRB14
GRB10
DGKQ
BRAF
ARHGAP20
ARAP3
ARAP2
ARAP1
ARAF
APBB1IP
AFDN_2
AFDN_1
I21
I24
Q25
H27
D30
E31
D33
T35
I36
E37
D38
S39
Y40
R41
D54
Y64
M67
Residues RAS
ΔΔG
(kcal/mol)
1
0
-1
B
Count
40
30
20
10
0
I36
Y40
D38
E37
R41
Q25
D30
D33
Y64
E31
S39
M67
Hotspot
Type
favorable
unfavorable
C
K216
H217
D38
V215
D
R630
D38
Y629
S631
E
I1343
D38
S1342
I1344
F
L572
D38
V571
T573

Figure 4

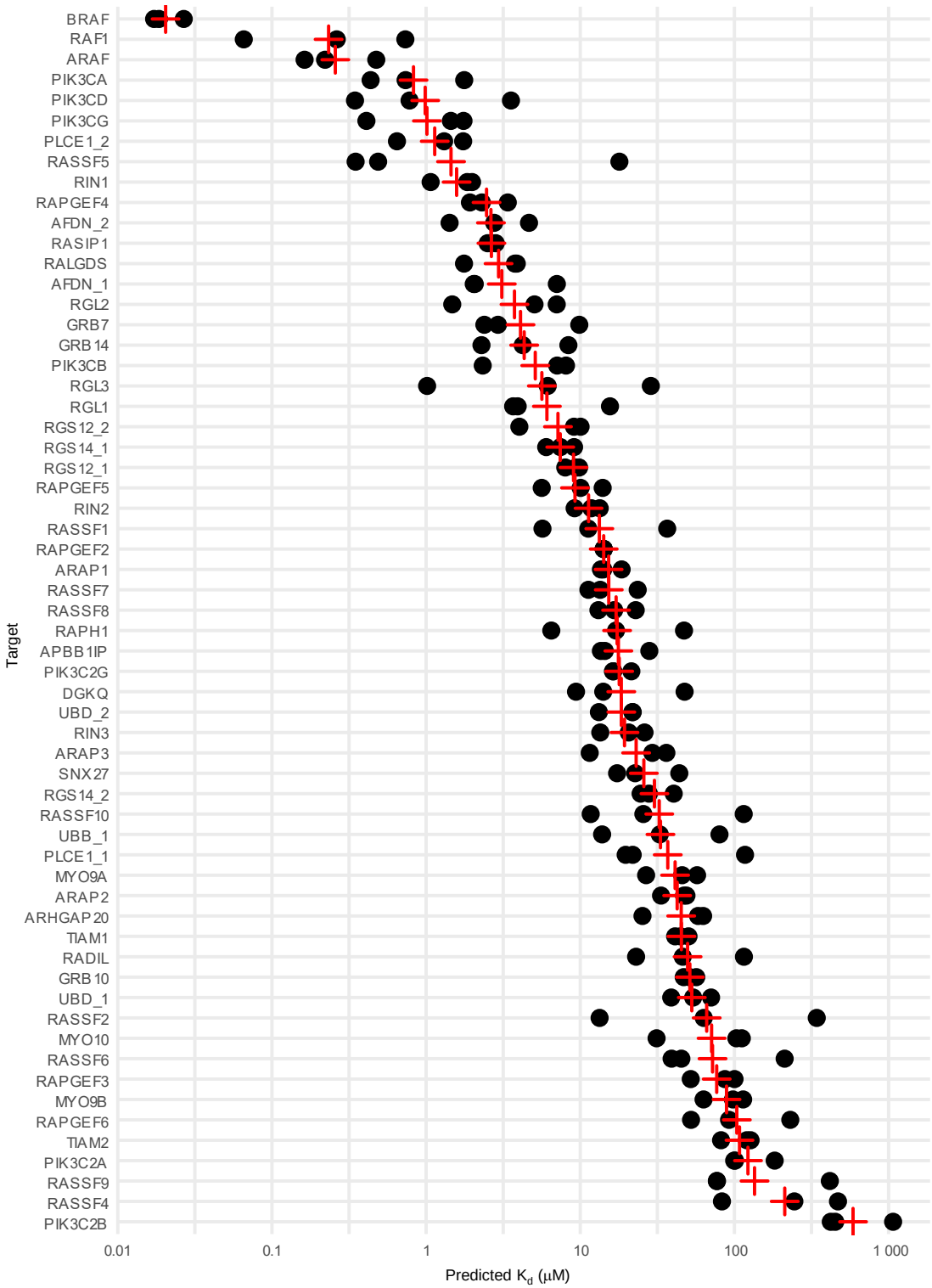

BRAF
RAF1
ARAF
PIK3CA
PIK3CD
PIK3CG
PLCE1_2
RASSF5
RIN1
RAPGEF4
AFDN_2
RASIP1
RALGDS
AFDN_1
RGL2
GRB7
GRB14
PIK3CB
RGL3
RGL1
RGS12_2
RGS14_1
RGS12_1
RAPGEF5
RIN2
RASSF1
RAPGEF2
ARAP1
RASSF7
RASSF8
RAPH1
APBB1IP
PIK3C2G
DGKQ
UBD_2
RIN3
ARAP3
SNX27
RGS14_2
RASSF10
UBB_1
PLCE1_1
MYO9A
ARAP2
ARHGAP20
TIAM1
RADIL
GRB10
UBD_1
RASSF2
MYO10
RASSF6
RAPGEF3
MYO9B
RAPGEF6
TIAM2
PIK3C2A
RASSF9
RASSF4
PIK3C2B
Target
0.01
0.1
1
10
100
1 000
Predicted K_d (μM)

Figure 5

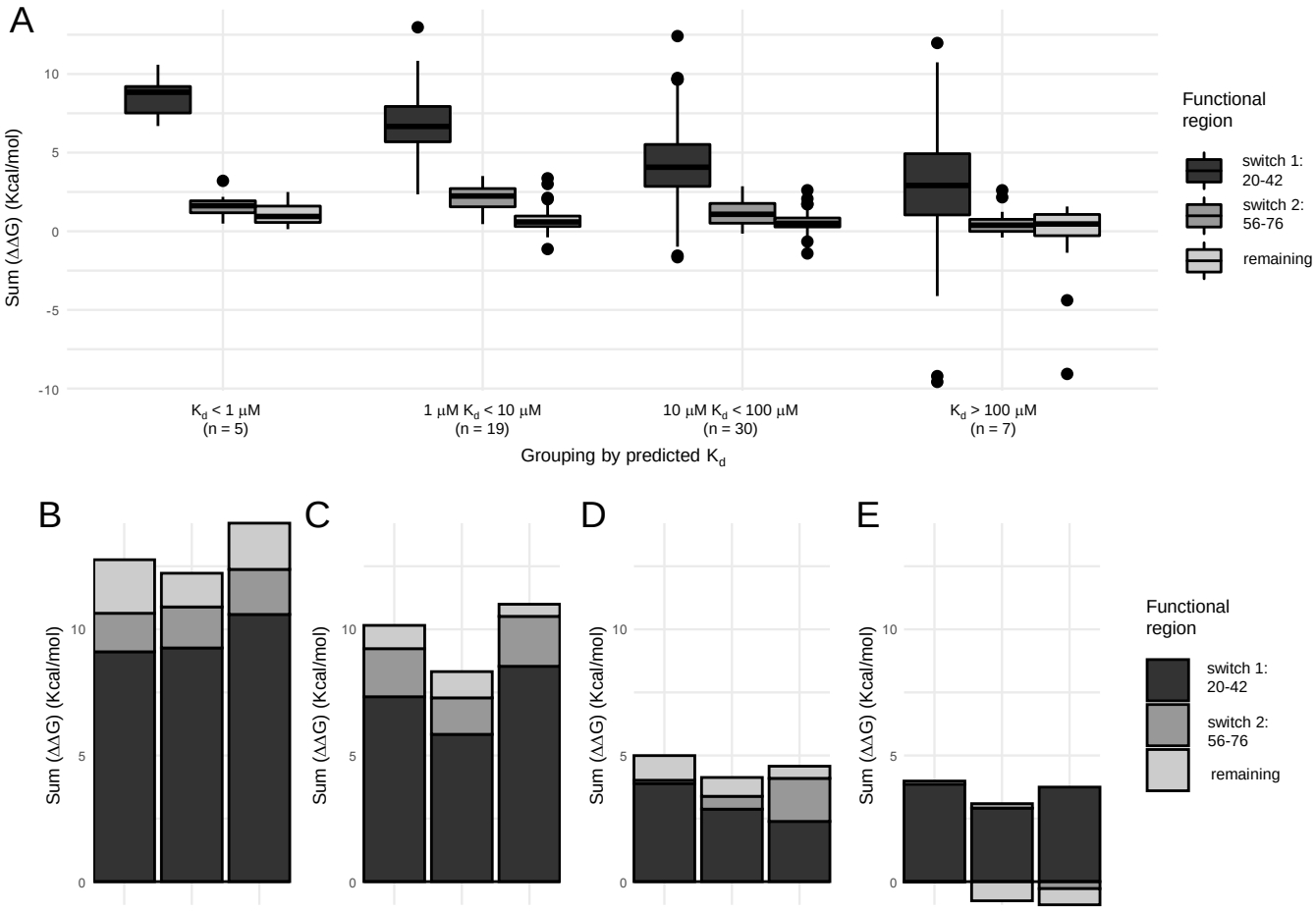

A
Sum (ΔΔG) (Kcal/mol)
10
5
0
-5
-10
Kd < 1 μM
(n = 5)
1 μM Kd < 10 μM
(n = 19)
10 μM Kd < 100 μM
(n = 30)
Kd > 100 μM
(n = 7)
Grouping by predicted Kd
Functional region
switch 1: 20-42
switch 2: 56-76
remaining
B
C
D
E
Sum (ΔΔG) (Kcal/mol)
10
5
0
Functional region
switch 1: 20-42
switch 2: 56-76
remaining

Figure 6

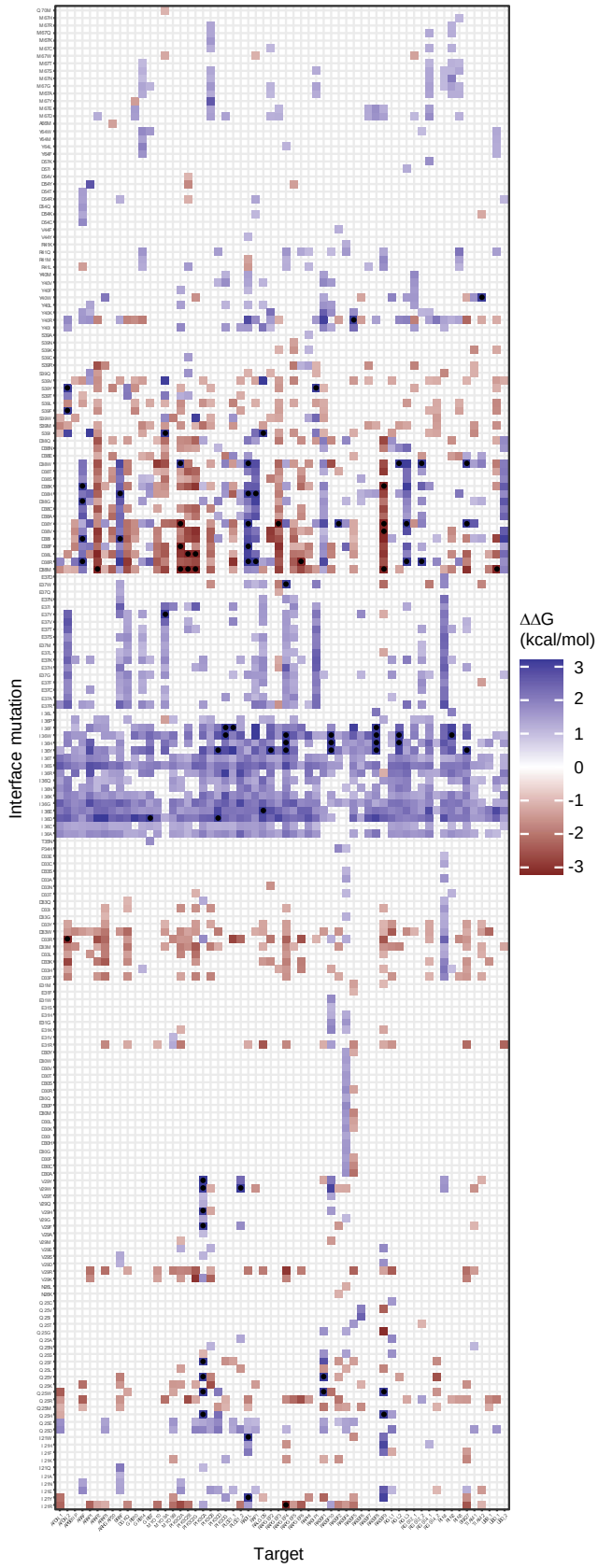


Figure 7

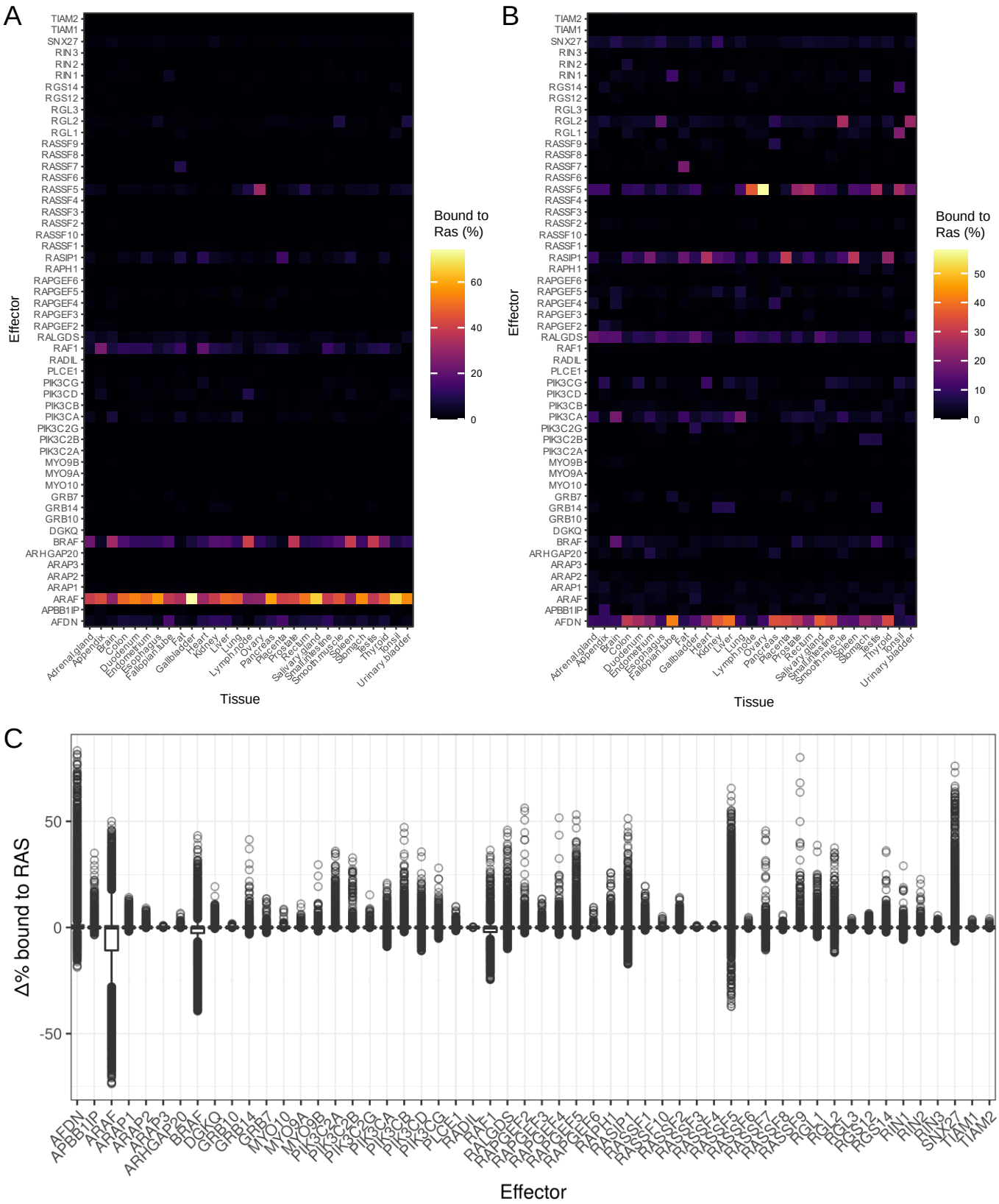
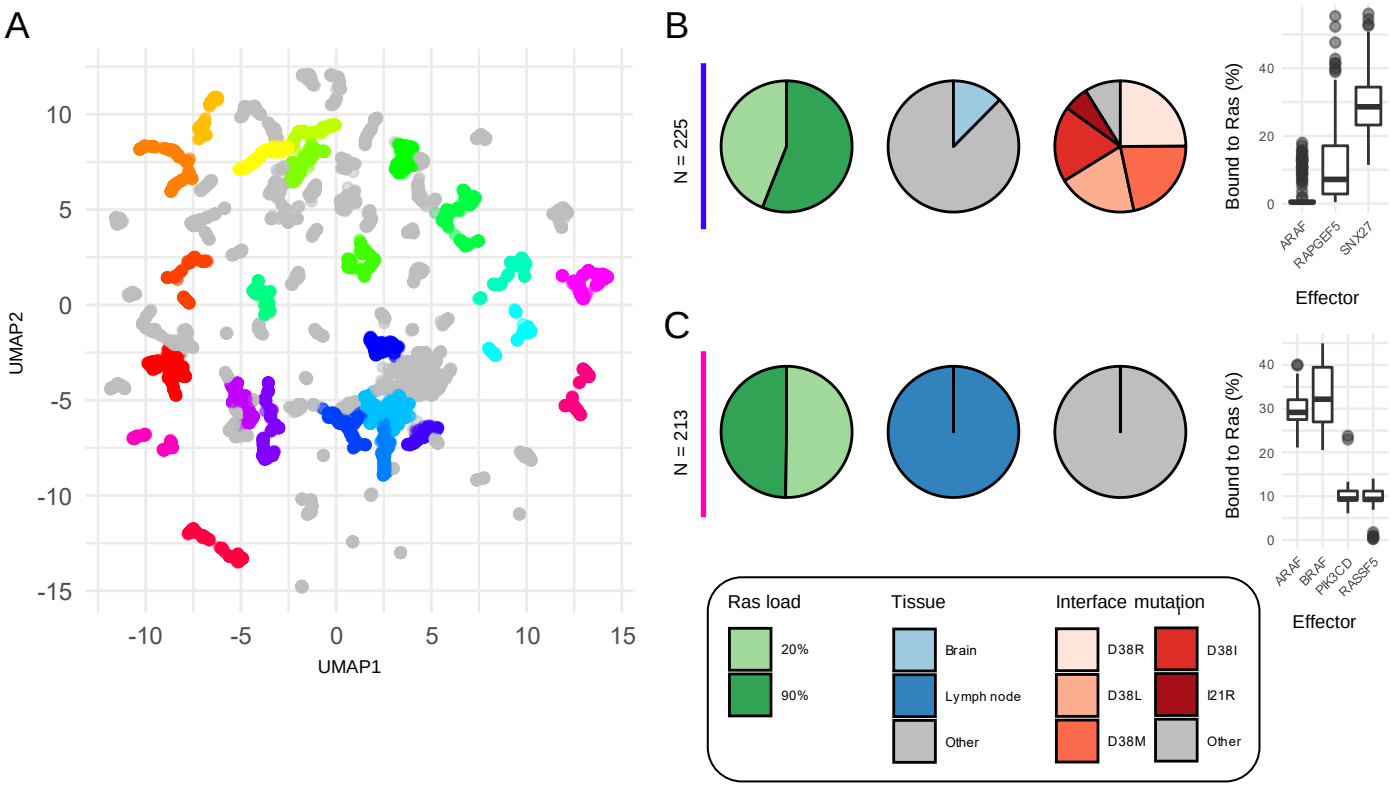

A
B
C
Effector
Tissue
Bound to Ras (%)
Δ% bound to RAS
TIAM2
TIAM1
SNX27
RIN3
RIN2
RIN1
RGS14
RGS12
RGL3
RGL2
RGL1
RASSF9
RASSF8
RASSF7
RASSF6
RASSF5
RASSF4
RASSF3
RASSF2
RASSF10
RASSF1
RASIP1
RAPH1
RAPGEF6
RAPGEF5
RAPGEF4
RAPGEF3
RAPGEF2
RALGDS
RAF1
RADIL
PLCE1
PIK3CG
PIK3CD
PIK3CB
PIK3CA
PIK3C2G
PIK3C2B
PIK3C2A
MYO9B
MYO9A
MYO10
GRB7
GRB14
GRB10
DGKQ
BRAF
ARHGAP20
ARAP3
ARAP2
ARAP1
ARAF
APBB1IP
AFDN
Adrenal gland
Appendix
Brain
Colon
Duodenum
Endometrium
Esophagus
Fallopian tube
Fat
Gallbladder
Heart
Kidney
Liver
Lung
Lymph node
Ovary
Pancreas
Placenta
Prostate
Rectum
Salivary gland
Small intestine
Smooth muscle
Spleen
Stomach
Testis
Thyroid
Tonsil
Urinary bladder

Figure 8


A
UMAP2
UMAP1
10
5
0
-5
-10
-15
-10
-5
0
5
10
15
B
N = 225
Bound to Ras (%)
40
20
0
ARAF
RAPGEF5
SNX27
Effector
C
N = 213
Bound to Ras (%)
40
30
20
10
0
ARAF
BRAF
PIK3CD
RASSF5
Effector
Ras load
20%
90%
Tissue
Brain
Lymph node
Other
Interface mutation
D38R
D38L
D38M
D38I
I21R
Other