The genetic architecture of G-protein coupled receptor signaling

Taylor L. Mighell[1] and Ben Lehner[1, 2, 3, 4, *]

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain

[2]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, United Kingdom

[3] Universitat Pompeu Fabra (UPF), Barcelona, Spain

[4] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

*e-mail: bl11@sanger.ac.uk

Abstract

G-protein coupled receptors (GPCRs) are the most abundant class of human receptors and drug targets. The vast majority of GPCR drugs bind the conserved orthosteric pocket, which can lead to non-specificity and toxicity. Precision modalities such as biased signaling and allosteric modulation could lead to GPCR therapeutics with improved efficacy and safety profiles. The rational design of precision therapies is, however, limited by the lack of high-resolution functional understanding of GPCRs. Here we present a general approach to chart high-resolution functional maps of GPCRs and apply the approach to the β2 adrenergic receptor. Across ~150,000 measurements of five different phenotypes, we construct an unprecedented picture of signaling properties of this model receptor. A contiguous core activation network of residues required for signaling links the ligand binding site to the transducer binding site. Residues controlling signaling parameters connect to this core network, with determinants of signaling bias located more peripherally. Dose-response curves for nearly all possible amino acid substitutions, a total of >7,500 different proteins, reveal the relationship between expression level and potency; and the spatial distribution of potency changing residues. The multi-modal maps rationalize previous serendipitously discovered allosteric sites and prospectively annotate novel sites to target. Application of this approach across the superfamily of GPCRs will unlock deep understanding of these critical molecules and enable rational design of precision GPCR therapeutics.

Introduction

G-protein coupled receptors (GPCRs) are the largest family of transmembrane proteins that receive and transmit signals from the outside to the inside of cells. These signals are very diverse and include neurotransmitters, metabolites, immune system molecules, pH, and photons of light[1]. The topology of GPCRs is highly conserved and consists of seven transmembrane (TM) helices connected by alternating intra- and extracellular loops (ECLs and ICLs). Binding of an agonist to the extracellular portion of the receptor results in an allosteric conformational change to the intracellular portion of the receptor[2], which enables binding and activation of one or more members of the family of heterotrimeric G-proteins, which subsequently activate diverse downstream signaling pathways[3]. Receptor activation also results in phosphorylation of the receptor C-terminal tail and recruitment of β-arrestin, which terminates G-protein signaling by internalizing the receptor, and also scaffolds G-protein independent signaling[4]. Because of their role as master signal transducers and their availability on the cell surface, GPCRs are the most important class of drug targets, with one third of clinical therapeutics targeting GPCRs[5,6].

The vast majority of established GPCR drugs act by binding to the orthosteric site and activating or inactivating the receptor[7]. However, alternative modalities for modulating GPCRs exist, such as biased agonism[8] (also called "functional selectivity") and allosteric modulation[9]. Biased agonists are defined in opposition to balanced agonists; a biased agonist preferentially signals through one downstream pathway over another (e.g. G-protein over β-arrestin), compared to balanced agonists. In cases where signaling through one pathway

leads to an unwanted side effect, biased agonists may achieve improved therapeutic efficacy. A therapeutic benefit for biased agonism has been proposed for at least thirty GPCRs[10].

Allosteric drugs are those that bind the receptor at a spatially distinct site from the orthosteric site. The high structural conservation of orthosteric sites makes it difficult to target one receptor without also engaging multiple family members, frequently causing toxicity and other side-effects that result in clinical failure[7]. In contrast, allosteric drugs targeting non-conserved sites can have high specificity and reduced toxicity. Further, since allosteric drugs modulate responses to physiological ligands, their effects temporally and spatially match physiological signaling processes[11]. Due to these major advantages, increasing effort is being put toward development of biased and allosteric modulators[12]. However, with our current understanding of GPCR function, these molecules are typically only found by blind screening or iteration upon previously discovered molecules[13,14].

In order to design better GPCR drugs, an improved understanding of the rich functional and allosteric architecture linking agonist binding with intracellular signaling is required. Much work has focused on using the structures of active and inactive states of receptors to try and understand this allosteric mechanism[15,16]. This has provided important mechanistic insights, but these structures are only the starting point for assigning a functional role for the hundreds of individual amino acids in each highly dynamic receptor.

A powerful complementary approach for understanding how proteins work is the use of mutagenesis followed by functional characterization[17]. Serial experiments have measured the effects of hundreds of alanine substitutions on

GPCR signaling through one[18] or two[19,20] pathways, revealing a rich diversity of signaling responses to individual mutations. Massively multiplexed approaches have been used to characterize surface expression[21–24], ligand binding[23,25], and signaling of human GPCRs[26–28]. However, these approaches to quantifying signaling relied on linking transcriptional reporters to endogenous signaling pathways, which can be non-linear and subject to feedback circuits, adding noise to measurements. Additionally, these systems have only been established for G-protein mediated signaling pathways. Finally, these multiplexed studies were not extended to formal dose-response analysis.

We construct and deploy here GPCR-MAPS (GPCR multiplexed assays with pathway specificity), an experimental and computational framework for generating comprehensive functional maps of GPCRs. We used the beta-2 adrenergic receptor (β2AR) as a model receptor. This extensively studied receptor is expressed in the cardiovascular system where it modulates smooth muscle relaxation and bronchodilation; β2AR agonists are therefore used for treating asthma as well as chronic obstructive pulmonary disease[29]. First, we measure the surface expression level and signaling response of all possible variants in the receptor to saturating concentration of agonist through G-protein and β-arrestin pathways. We model the relationship between expression and signaling to account for indirect effects of changes in expression on signaling and identify a structurally contiguous network of residues linking the ligand binding site to the transducer binding site. By directly comparing G-protein and β-arrestin signaling, we then identify residues involved in biased signaling enriched at the extracellular loops (also known as the “extracellular vestibule”) and near the conserved NPxxY motif in the 7th transmembrane (TM) helix. We

then extend the platform to enable, for the first time, massively multiplexed dose-response analysis for GPCRs. A 12-point dose-response experiment, encompassing nearly 100,000 measurements, enables highly accurate quantification of the dose-response curves of 5,588 different receptor variants. The data provide numerous functional insights, including identifying intracellular loop 3 as a hotspot for mutations that increase potency. Finally, we use the rich data generated in this study to explain the allosteric activity of known modulators and prospectively annotate new structural pockets to therapeutically target. Application of the GPCR-MAPS platform across the important superfamily of GPCRs will advance our understanding of these proteins and enable rational, precision drug design.

Results

GPCR-MAPS: multiplexed, multimodal assay with pathway specificity

An ideal platform for measuring GPCR signaling activity at scale would enable the readout of the two canonical signaling pathways (G-protein- and β-arrestin-mediated) and would rely on a signal detector proximal to the receptor, thereby avoiding feedback or non-linearity of intracellular signaling pathways. A previously described experimental platform[30] uses three genetic constructs to report on GPCR signaling: 1) a GPCR of interest with a fusion of the exogenous transcription factor Gal4, and linked by sequence encoding an evolved light, oxygen, and voltage sensitive (eLOV) domain as well as tobacco etch virus protease cleavage site (TEVcs), 2) a GPCR binding partner (e.g. β-arrestin) fused to NanoLuciferase (NLuc) and tobacco etch virus protease (TEV), and 3)

the fluorescent protein Citrine under the control of the UAS promoter. When co-expressed in cells, agonist-mediated activation of the GPCR causes recruitment of the transducer construct to the receptor. In the absence of light, the eLOV domain cages the TEVcs and prevents cleavage; however, co-addition of the NLuc substrate furimazine results in bioluminescence, and when the transducer brings NLuc into close proximity with the eLOV domain, the steric hindrance is relieved and TEVp can cleave at TEVcs. This cleavage results in liberation of Gal4, which can move to the nucleus and drive transcription of the UAS-Citrine reporter construct.

While this assay faithfully reports on GPCR activation, it is limited to serial experimental designs and has only been demonstrated with one downstream signaling partner. Therefore, we extensively re-engineered the system to enable massively multiplexed experiments. We combined the GPCR construct and the reporter construct onto a single plasmid and added a DNA barcode (26 nucleotides) in the 3' untranslated region (UTR) of the Citrine reporter so that the GPCR variant on any given plasmid can be identified (Figure 1a and Supplementary Figure 1d). We also added the attB recombination site to enable integration into HEK293T landing pad cells[31], which ensures copy number of exactly one in each cell. We used the active-state-specific nanobody Nb80[32] as the binding partner to optimize the assay. While the original assay used transient transfection to introduce transgenic material and took place over the course of minutes, we found that substantially increased times were necessary, presumably due to the reduction of copy number of the receptor and reporter constructs (24 hours, Supplementary Figure 1a). Further, we tested evolved variants of both TEV and the LOV

domain, uTEV1Δ[33] and hLOV1[34]; the combination that yielded the highest signal to noise ratio was uTEV1Δ and eLOV (Supplementary Figure 1b). Next, we also validated the use of β-arrestin and miniG_s[35] as the binding partner in this system (Supplementary Figure 1b). Finally, in order to account for differences in barcode abundance in the library, as well as barcode-specific effects on RNA abundance, we also include a condition in which Gal4 is expressed alone (Supplementary Figure 1d-e); this condition should maximally stimulate all barcodes and provides normalization for the experimental conditions[26]. While this platform enables multiplexed measurement of GPCR signaling, it is expected that many variants that exhibit decreased signaling will do so because of a lack of surface expression[27]. We therefore implemented in parallel a FACS-based method to measure the surface expression of library members[36] (Figure 1a, Materials and Methods).

Surface expression and signaling through both canonical pathways

We generated a saturation mutagenesis library of β2AR in a vector also containing: an N-terminal HA epitope tag and C-terminal fusions of eLOV, TEVcs, and Gal4, and additionally the UAS-Citrine module (attB-HA-b2AR-eLOV-TEVcs-Gal4-UAS-Citrine, Supplementary Figure 1d) including all missense and nonsense substitutions, synonymous mutations where possible (i.e. the encoded amino acid can be encoded by alternate codons), and single residue deletions. We then introduced short degenerate nucleotide barcodes into the 3' UTR of Citrine and used long read sequencing to make 212,840 β2AR genotype-barcode links (Supplementary Figure 1c, Materials and

Methods). These represented 8,970 out of an intended 9,046 unique genotypes (99.2%).

We introduced this library into HEK293T landing pad cells[31], ensuring a single variant is expressed per cell. Then, we performed a FACS-based multiplexed surface expression experiment. Namely, cells were immunostained with anti-HA antibody, then sorted into four bins based on fluorescence intensity (Supplementary Figure 1d). DNA from the cells sorted into each bin was purified and converted into sequencing libraries, and the barcodes were sequenced with short-read sequencing (Figure 1a, Materials and Methods).

All counts from barcodes representing the same variant were tallied and combined with the fluorescence value of each FACS bin to arrive at surface expression estimates. Variants with low frequency in the library were filtered to yield 8,740 high confidence surface expression estimates (Materials and Methods, Supplementary Table 1). These measurements were highly reproducible across three replicates (Pearson r=0.94-0.96, Figure 1c). Scores were normalized between the median of premature stop codons, set to 0, and the wild-type surface expression level, set to 1. A heatmap representation of the data shows nonsense mutations are uniformly damaging, except those occurring within intracellular loop 1 (ICL1) and those in the C-terminus, suggesting the receptor can be stably expressed as either a single-pass transmembrane protein, or as a C-terminal tail truncation. The transmembrane helices are more mutation sensitive than extra- or intra-cellular loops, consistent with previous work[27,36] (Figure 1d). Synonymous wild-type and nonsense variants have well separated expression distributions, with missense variants spanning the full range between those two classes (Figure 1e).

Next, we performed the GPCR-MAPS signaling assay using Nb80 (as a surrogate for G-protein, as Nb80 returns higher signal to noise than miniG_s (Supplementary Figure 1b)) as well as β-arrestin as readouts for the two canonical GPCR signaling pathways. Namely, we transiently transfected one of the transducer constructs (pUC19-NLuc-[Nb80/βArr]-uTEV1Δ) or the Gal4 normalization construct (pUC19-CMV-Gal4, Supplementary Figure 1e) into cells bearing the β2AR receptor construct. The next day, saturating concentration of isoprenaline (10 μM) was added to the transducer samples to induce signaling, and 24 hours later, those cells were collected (Figure 1b, Supplementary Figure 1f). The Gal4 maximum induction samples were collected 30 hours after Gal4 transfection. RNA was collected from all samples, reverse transcribed, converted to short-read sequencing libraries and sequenced (Supplementary Figure 1f-g, Materials and Methods).

Signaling scores were calculated using the variant frequency in the experimental samples compared to variant frequency in the Gal4 condition and re-scaled with 0 representing no signaling activity and 1 representing wild-type signaling. Low frequency variants were filtered to yield 7,850 high confidence Nb80 and β-arrestin signaling scores (Materials and Methods, Supplementary Table 1). Replicates were well correlated (Pearson's r=0.9-0.91 for Nb80 replicates, r=0.88-0.89 for β-arrestin replicates, Figure 1c). Heatmap representations of the data reveal similar overall trends between Nb80 and β-arrestin (as well as with the surface expression data) defined by transmembrane regions being most sensitive to mutation (Figure 1 f, h). Synonymous wild-type variants are again well separated from nonsense variants, and missense variants again span the whole distribution from

nonsense to wild-type-like (Figure 1 g, i). The Nb80 and β-arrestin signaling scores are correlated with each other (Spearman's ρ = 0.87), as well as with the surface expression scores (ρ=0.55, 0.6, respectively, Supplementary Figure 1h), consistent with changes in expression being a very important cause of changes in signaling.

While the overall patterns of mutational tolerance are similar between surface expression and signaling datasets, there are positions with stark differences. For example, missense mutations at the eight positions making up the isoprenaline binding pocket have minimal effect on stability (Figure 1j), while these same mutations are strongly deleterious for signaling through both binding partners (Figure 1k-l).

Residues contributing to signaling beyond expression changes

While the effects of mutations substantially manifest through expression changes, we sought to identify mutation effects that specifically affect signaling. So, we directly compared signaling and surface expression scores of missense variants. These demonstrate a non-linear, correlated relationship (ρ = 0.55, 0.6 for surface expression compared with Nb80 and β-arrestin, respectively, Figure 2a-b). We fit a LOWESS regression to the surface expression and signaling data, which acts as a null expectation of signaling level produced by a given surface expression level. This null expectation explains a significant amount of variance (R^2 = 0.37 and 0.42 for Nb80 and β-arrestin, respectively).

We next calculated signaling residuals and subsequently z-scores of all variants to the global model (i.e. residual divided by standard error of the mean for each variant, Figure 2c-e). Comparing the group of z-scores at each position

with the z-scores from all other positions with a two-sided Mann-Whitney U allows us to identify residues where mutations have greater or lesser signaling activity than would be expected based on their surface expression level (Figure 2e, Supplementary Table 2).

For both Nb80 and β-arrestin, more residues confer loss-of-function (LoF) than gain-of-function (GoF) when mutated, and the magnitude of loss is greater than the magnitude of gain (Figure 2 f-g). In total, 49 and 73 residues are LoF for Nb80 and β-arrestin, respectively (false discovery rate, FDR = 0.1, Supplementary Table 2). Of these, 47 are shared (Figure 2h). An example of a strong LoF residue is D113^{3x32} (Figure 2i, GPCRdb numbering[37] in superscript), which participates directly in ligand binding. There are 31 and 53 GoF residues for Nb80 and β-arrestin (FDR=0.1), respectively, of which 20 are shared. E268^{6x30} is the most significant GoF (Figure 2j), consistent with previous work demonstrating constitutively activating mutations at this site[38]. Finally, we extended this analysis to highlight residues with highly variable effects on signaling. Namely, after excluding the LoF and GoF residues, we compared the interquartile range (IQR) of z-scores at all positions (Supplementary Figure 2a-b). We categorized those residues with IQR above the 90th percentile as change-of-function (CoF): 34 and 31 residues for Nb80 and β-arrestin, respectively, with 8 shared (Figure 2h). An example of a strong CoF residue is S120^{3x39} (Figure 2k). This residue repacks upon β2AR activation[18] and is thought to modulate the conformation of helix 7[39].

We prioritized for further analysis residues that were identified in the analyses of both transducers. We note that average position z-scores and position IQRs are well correlated (r = 0.73 and 0.56, respectively,

Supplementary Figure 2c-d), indicating that many residues contribute similarly to signaling through both pathways. Plotting the residues identified for both transducers onto the structure of isoprenaline-stimulated β2AR (PDB ID: 7DHR) reveals that LoF residues are predominantly within the core of the receptor (Figure 2l), whereas GoF and CoF residues are dispersed on the periphery (Figure 2m-n).

Structural basis of loss-, gain-, and change-of-function residues

Next, we analyzed in more detail the identified functional residues. We examined the LoF residues in an inactive (PDB ID: 2RH1) and an active (PDB ID: 3SN6) structure of β2AR. Strikingly, the majority of these residues (39 out of 47) form an interconnected path linking the ligand binding site all the way to the transducer binding site in the active state (Figure 3a, hereafter referred to as the "core activation network"). All 8 isoprenaline-binding residues are identified (D113^{3x32}, V117^{3x36}, F193^{45x52}, S203^{5x43}, S207^{5x461}, F290^{6x52}, N293^{6x55}, and N312^{7x38}) along with many residues in the second shell of the ligand binding pocket (i.e. residues that contact residues that contact isoprenaline, Figure 3b). These residues include W286^{6x48} of the CWxP motif and I121^{3x40} of the PIF motif[40,41]. The core activation network then connects to N322^{7x49} of the critical NPxxY motif[40] via two pathways: first, W286^{6x48} contacts N318^{7x45} which then connects to N322^{7x49}. Second, a pathway from A119^{3x38} goes through A78^{2x49} and D79^{2x50} of the allosteric sodium binding pocket which then contacts N322^{7x49}. Of the NPxxY motif, N322^{7x49}, I325^{7x52}, Y326^{7x53} are identified as LoF. Finally, in the active state structure, Y326^{7x53} of the NPxxY motif then makes contact with R131^{3x50} (of the DRY motif) and I127^{3x46} of helix 3. Y326^{7x53} is also

thought to make an important water-mediated hydrogen bond with Y219$^{5.58}$,[42] which is also identified as LoF and makes a direct contact with R131^{3x50}.

Of the eight LoF residues not connected to the core activation network, three likely directly disrupt transducer binding[32,43] (P138^{34x50}, F139^{34x51}, and Y141^{34x53}), while three more take part in critical disulfide bonds between ECL2 and TM3 (C106^{3x25}, C190^{ECL2}, C191^{45x50}). F193^{45x52} is a disconnected LoF residue that makes direct contact with bound isoprenaline. Finally, P88^{2x58} in TM2 is the last disconnected residue.

We then overlaid the GoF residues onto the inactive and active structures (Supplementary Figure 3a). Two of the GoF residues, D130^{3x49} and E268^{6x30}, participate in the highly conserved ionic lock (with R131^{3x50} of the DRY motif[44]). While R131^{3x50} of the DRY motif is a LoF residue that contacts Nb80[32] and is predicted to contact β-arrestin,[43] mutations at D130^{3x49} of the DRY motif have been shown to constitutively activate the receptor[45]. The other GoF residues are predominantly in two clusters: one consisting of TMs 3, 5, and 6, near the NPxxY and DRY motifs (Supplementary Figure 3b). L124 sits in the center of the receptor, near the I of PIF motif, DRY motif, and NPxxY. Mutations at L124^{3x43} are known to cause constitutive activity[46]. The other cluster is near the ligand binding site and in the extracellular vestibule (Supplementary Figure 3c). E187 and T189 are in ECL2, where they may affect ligand binding or activation[47]. W313^{7x39} and V87^{2x57} form two van der waals interactions in the inactive (PDB:2RH1), but only one in the active state (PDB:3SN6, GetContacts); our data suggest that disrupting this interaction leads to increased signaling. Finally, there are 7 GoF residues in the C-terminal tail (Supplementary Figure 3g), consistent with previous findings[26].

Plotting the CoF residues onto the inactive and active state structures (Supplementary Figure 3d) reveals that they cluster near important LoF residues that make structural changes during activation: for instance, S120^{3x39} and E122^{3x41} are adjacent to I121^{3x40} of the dynamic PIF motif. Likewise, F321^{7x48} is adjacent to N322^{7x49} of the dynamic NPxxY motif and identified LoF, while F332^{8x50} is nearby (Supplementary Figure 3e). This proximity to essential switch residues suggests that CoF residues may specifically stabilize or destabilize the active or inactive receptor state depending on the amino acid substitution.

Genetic basis of biased signaling

Next, we sought to directly compare mutational effects on Nb80 and β-arrestin signaling to understand the molecular basis of biased signaling (Figure 4a). Missense mutational effects on signaling through Nb80 and β-arrestin are well correlated (Pearson's r=0.89, Figure 4b). However, single mutants previously shown to bias the receptor toward G-protein signaling (F193A[48], Y219A[49], and Y199A[50]) are indeed biased towards Nb80 in our experiments (Figure 4c). To find residues with stronger effects on one pathway compared to the other, we implemented a similar strategy as before: we fit a LOWESS regression curve to the data, and calculated residuals and subsequently z-scores (Figure 4c) for all variants in the dataset. Then, we compared z-scores of mutations at each position with z-scores of mutations at all other positions. This identifies 19 residues where mutations have significantly greater effects on β-arrestin signaling and 7 positions where mutations have significantly greater effects on Nb80 signaling (FDR=0.1, Figure 4d, Supplementary Table 3). As

before, we then compared the IQR of all remaining residues; here, to more closely match the number of biased residues, we prioritized the top 5% of residues as bias switches (Supplementary Figure 4a).

Examples of strongly biased residues are shown (Figure 4e-f) as well as a strong bias switching residue (Figure 4g). One expectation is that mutations at residues making direct contacts with either Nb80 or β-arrestin could result in biased signaling effects. Nb80 and β-arrestin have overlapping but distinct binding interactions with the intracellular portion of β2AR. Nb80 forms direct contacts with 16 residues in TMs 3, 5, 6, and 7[32]. Since there is no experimentally determined structure of β2AR bound by β-arrestin, we defined this interface from a simulation study[43]. The β-arrestin binding interface is predicted to be larger, encompassing 28 residues across ICLs 1, 2, 3, and TMs 6 and 7, with 10 residues overlapping the Nb80 interface.

Three out of the seven residues biased toward β-arrestin make direct binding interactions with both Nb80 in the solved structure and with β-arrestin in the simulation (P138^{34x50}, F139^{34x51}, and Y326^{7x53}), suggesting a more critical role for these residues in binding Nb80. The other four β-arrestin-biased residues are either in ICL3 (G252 and G255) or the C-terminal tail (E373 and C406, Supplementary Figure 4b). The only residue biased toward Nb80 within either binding interface is P330^{8x48}, which is predicted to interact with Nb80 but not β-arrestin. There is also one bias switch residue, R328^{7x55}, that is predicted to bind β-arrestin. This relative sparsity of bias effects suggests that few residues at the transducer interface contribute significantly more to signaling through either pathway, at least for these two transducers (Supplementary Figure 4c).

In contrast, there are many residues biased toward Nb80 located in the extracellular vestibule: W99^{23x50}, F101^{23x52}, G102^{3x21}, W173^{ECL2}, A181^{ECL2}, C184^{ECL2}, Y185^{ECL2}, N196^{5x36}, and E306^{7x32} (Figure 4h, k). The most strongly Nb80 biased residue, W99^{23x50}, was previously identified as particularly mutationally intolerant[26]; however only by measuring both canonical signaling pathways could we uncover its strongly biased effects (Figure 4e). Considered together, the extracellular loops (ECLs) are enriched in Nb80 biased residues compared to the whole receptor (OR=6.4, $p=1.0x10^{-3}$, one-sided Fisher exact test), and in particular, ECL2 is enriched for Nb80 biased residues (OR=5.6, p=0.014, one-sided Fisher exact test).

The importance of the extracellular vestibule for G-protein biased signaling revealed in our mutational data is also consistent with structural data. Salmeterol is a G-protein biased β2AR agonist that binds the orthosteric site in a similar way as the endogenous ligand (epinephrine) except that it has a long aryloxyalkyl tail that extends into the extracellular vestibule, where it interacts extensively with ECL2[51]. Also, comparative structural analyses of other GPCRs have suggested that the extracellular vestibule adopts a more constricted conformation in the β-arrestin-bound state compared to the G-protein-bound state.[52] Taken together, these data point to the extracellular vestibule, and in particular ECL2, as hotspots for G-protein-biased agonism at the β2AR.

Among the bias switching residues, the second strongest is the vestibule residue, Y308^{7x34} (Supplementary Figure 3d). This residue has been proposed to be necessary for G-protein signaling[53], as well as forming a hydrophobic gate with F193^{45x52} that interacts with and dehydrates ligand molecules as they traverse the vestibule and approach the orthosteric binding site[54]. In the active

state, these two residues form a “lid” over the ligand, with F193^{45x52} making a direct contact with isoprenaline (Figure 4i). Based on the alanine mutation, F193^{45x52} was postulated to coordinate β-arrestin signaling[48]. Interestingly, while this single variant does have a stronger effect on β-arrestin signaling, other variants at the same site have the opposite effect (Supplementary Figure 4e). We examined the effects of mutations at these two sites and found that signaling bias is correlated with mutant hydrophobicity for Y308^{7x34} ($\rho = 0.67$, $p = 4.3 \times 10^{-3}$, Figure 4j), but not for F193^{45x52} ($\rho = -0.36$, $p = 0.16$, Supplementary Figure 4f). This suggests that the role of Y308^{7x34} in determining bias is directly tied to its hydrophobicity, while the relationship of F193^{45x52} to bias is more nuanced.

Besides the vestibule, there are multiple bias-functional residues (Nb80- or β-arrestin-biased or bias switches) in or near the NPxxY motif of TM7(Figure 4l). In fact, TM7 is enriched for bias-functional residues compared to the whole receptor (OR=3.7, p=0.011, one-sided Fisher exact test, Supplementary Figure 4g). P323^{7x50} is not predicted to contact directly β-arrestin[43], suggesting that the Nb80 bias of the residue cannot be explained simply by disrupting a direct contact. N318^{7x45}, F332^{8x50}, and R328^{7x55} are all bias switches in close proximity to the NPxxY motif (Figure 4k). A role for TM7 in determining bias is consistent with prior nuclear magnetic resonance experiments demonstrating changes in TM7 conformation upon binding of biased agonists[55].

Quantifying >7,500 dose-response curves

The data generated thus far represents β2AR signaling with saturating levels of isoprenaline, which convolutes the contribution of different

pharmacological parameters, such as potency (the amount of agonist needed to elicit a certain response) and efficacy (the maximum signaling response elicited by saturating concentrations of agonist). Therefore, we built a platform for measuring full dose-response curves for thousands of β2AR variants in parallel. First, we performed the GPCR-MAPS signaling assay with 12 serially diluted concentrations of isoprenaline, ranging from 10 pM to 10 μM (Figure 5a). In order to compare the magnitude of signaling across experimental conditions, following RNA isolation, we performed quantitative PCR (qPCR) to estimate the total amount of reporter transcription (Figure 5b, Supplementary Figure 5a, Materials and Methods). Then, by deep sequencing the barcodes and normalizing the frequencies by the estimated total reporter RNA, we arrived at normalized scores across the 12 concentrations (Figure 5c and Supplementary Figure 5b). This allowed us to quantify dose-response relationships for 7,680 different proteins, increasing data of this type for β2AR or any GPCR >10-fold[18–20].

Heatmap and correlation representations of the data reveals induction of signaling across the ~92,000 measurements (Figure 5d and Supplementary Figure 5c). Clustering the induction dataset revealed three predominant clusters- these can be assigned as wild-type-like, moderately damaging, and strong loss-of-function (Figure 5e).

Next, we fit a four-parameter Hill curve to the wild-type genotype, normalizing E_{min} and E_{max} to 0 and 1, respectively, and deriving a half maximal response estimate of 26.2 nM and slope (Hill coefficient) of 0.84 (Figure 5f). Then we fit curves to all variants; after quality control (Supplementary Figure 5e-f, Materials and Methods), this resulted in 5,588 high quality dose response

curves with well-estimated parameters (Figure 5. g-l). We note that variants with reduced signaling level are more poorly fit and therefore likelier to be filtered.

Genetic determinants of potency

We sought to use the fitted dose response curves to understand the contribution of residues to potency, quantified as the amount of agonist required to reach a half maximal response (EC_{50}, Figure 6a). To validate our EC_{50} measurements, we compared them to a set of individually assayed alanine substitutions[18]; these two datasets are well correlated (Pearson r=0.74, n=317 variants, Figure 6b). Next, we examined correlations between the parameters of the dose-response curves, as well as their dependence upon surface expression. Expression level has been predicted to be an important influence on signaling parameters in theoretical work[56–58], and by experimental work on a plant hormone receptor[59]. The strongest relationships we found were between slope and E_{max} (ρ = -0.46) and between surface expression and EC_{50} (ρ = -0.41, Figure 6c-d). Surface expression level is thus an important influence on potency in the β2AR, consistent with theoretical predictions[56–58]. In order to isolate mutational effects on EC_{50} beyond those caused by changes in expression we again used LOWESS regression to model the relationship between surface expression and EC_{50} (Supplementary Figure 6a) and used residuals to that model to identify positions with consistently higher or lower residuals than expected. 29 residues were identified as GoF (i.e. they had EC_{50} lower than expected) and 59 were identified as LoF (i.e. higher EC_{50} than expected, FDR=0.1, Figure 6e). There is substantial overlap between these sets and the sets identified from the saturating concentration experiment (23/59 LoF shared,

4/29 GoF shared). Finally, the interquartile range was used again to identify 29 EC_{50} CoF residues (one shared with saturating concentration experiment, Supplementary Figure 5b, Supplementary Table 4).

In principle, mutations could affect potency beyond changes in expression by changing the affinity of isoprenaline binding, or by affecting the probability of isoprenaline-bound receptor to adopt the active conformation[60]. While mutations at residues making direct contact with isoprenaline would be expected to reduce affinity, mutations at these sites mostly totally abolish signaling; dose-response curves could only be fit to 32 out of 132 missense variants at these sites. However, these 32 variants do have over an order of magnitude reduced potency (median EC_{50}=477 nM, compared to wild-type 26.2 nM). Interestingly, EC_{50} LoF residues are enriched in the second shell (15/59 LoF sites (25.4%, OR=4.12, $p=1.9x10^{-4}$, Fisher exact test). While not statistically enriched, a higher fraction of EC_{50} CoF sites (5/24, 20.8%) are in the second shell compared to only 1 out of 29 GoF residues (3.4%) in the second shell (Figure 6i). This shows that whereas mutations in residues directly contacting the ligand have very strong effects on affinity and so completely inactivate signaling, mutations in second shell residues frequently have more moderate effects and modulate the response. These effects are, by definition, allosteric, either indirectly altering ligand binding affinity or altering the consequences of ligand engagement. This also suggests that mutations in the second shell are likelier to reduce affinity than increase it.

Beyond changes in ligand affinity mutations can also affect signaling by altering the likelihood of ligand-bound receptor adopting the active conformation.[60] One way to disrupt the activation pathway could be for

mutations to interfere with the core activation network identified in the maximum signaling experiment. We therefore quantified the distance of EC_{50}-affecting residues to this network and found that LoF and CoF residues are closer in 3D space than GoF residues ($p=7.4x10^{-4}$, $1.3x10^{-3}$, respectively, Mann Whitney U test, Figure 6j). Taken together, as with the maximum signaling analysis, these data suggest that mutations that reduce potency do so by perturbing the core activation pathway whereas the mechanism-of-action of GoF mutations is likely distinct.

Three of the four most significant EC_{50} GoF residues are $E268^{6x30}$ (Figure 6g), $D130^{3x49}$, and $L272^{6x34}$. $E268^{6x30}$ and $D130^{3x49}$ were already identified in the maximum signaling analysis and make up the conserved ionic lock while $L272^{6x34}$ is in close proximity to the ionic lock, where mutations are also known to destabilize the inactive state.[60] Nearby $E268^{6x30}$ in intracellular loop 3 (ICL3), there are nine EC_{50} GoF residues, which is enriched compared to the whole receptor (OR=3.38, $p=7.3x10^{-3}$, one-sided Fisher exact test). This loop connects TM5 and TM6, which make dramatic movements upon receptor activation (Figure 6k). These GoF residues could act in a similar way as $D130^{3x49}$, $E268^{6x30}$, and $L272^{6x34}$, weakening the interaction between TM3 and TM6 or other helices. Alternatively, previous studies suggest that ICL3 may have an autoinhibitory role that sterically occludes transducer binding in the inactive state, and that some mutations can abrogate this autoinhibitory effect[61]. Further experiments would be necessary to determine the exact mechanism of the EC_{50} GoF phenotypes in ICL3. Additionally, there are two GoF residues in the N-terminus that coincide with the two N-glycosylation sites[62] (N5 and N15,

Supplementary Figure 6d) suggesting that loss of these glycosylation modifications can increase receptor potency.

F282^{6x44} of the PIF motif is the strongest EC_{50} CoF residue consistent with previous data showing multiple constitutive activating mutations and at least one attenuating mutation.[63] Likewise, S120^{3x39} and L212^{5x51} are identified as CoF and are directly adjacent to the other two members of the PIF motif (I121^{3x40} and P211^{5x50}) reinforcing that some perturbations to this key switch motif can result in gains or loss of signaling. There are also several EC_{50} CoF residues in TMs 5 and 6 (4 in each, Supplementary Figure 6f), reflecting the key role of these helices in activation.

Rationalization of known allosteric modulators and prospective pocket prioritization

We next brought to bear our rich datasets to rationalize the activity of known allosteric modulators and prospectively prioritize new pockets for therapeutic discovery. Cmpd-15 is a negative allosteric modulator (NAM) but its mechanism is likely simple steric interference with transducer binding.[64] AS408 is a NAM that binds away from the orthosteric and transducer binding sites, between TMs 3 and 5[65] (Figure 7a). Across the three analyses (maximum signaling, biased signaling, and EC_{50}), the 7 residues making up this pocket have 7 total annotations, compared to 96 annotations for all 220 structurally resolved surface exposed residues (>10% relative solvent accessible surface area in 2RH1), indicating a functionally important pocket. In particular, V129^{3x48} is LoF for maximum signaling as well as EC_{50}, while E122^{3x41} is CoF for maximum signaling as well as EC_{50} (Figure 7b-c). In addition, C125^{3x44} is CoF

for EC_{50} and I214^{5x53} is LoF for EC_{50}. Mutations at P211^{5x50} are biased toward Nb80 signaling, though AS408 is not thought to be a biased NAM.

While the AS408 pocket is replete with functional residues, Cmpd-6, a positive allosteric modulator[66] (PAM), binds in a pocket formed by 9 residues between TMs 2, 3, 4, and ICL2 (Figure 7d) with only 4 annotations. However, one of the residues is D130^{3x49}, where mutations cause strong GoF for both maximum signaling and EC_{50} (Figure 7e-f), potentially explaining the PAM activity of Cmpd-6. Additionally, while Cpmd-6 mostly forms hydrophobic interactions with β2AR, it does form one hydrogen bond with K149^{4x41}, which is an EC_{50} CoF residue (Figure 7e-f). Taken together, our functional annotations are consistent with the activities of the two known β2AR allosteric modulators.

To explore prospective functional annotation, we used Fpocket[67] to predict pockets in the inactive β2AR (PDB ID:2RH1). One additional pocket, between TMs 5 and 6 (Figure 7g), has 8 functional annotations among 10 residues, with most being maximum signaling or EC_{50} LoF or CoF, but also one GoF residue (Figure 7h-i), suggesting a highly active pocket. Indeed, an allosteric agonist was recently found to bind and activate another GPCR, GPR88, at this same site[68]. On the other hand, a nearby and partially overlapping pocket also between TMs 5 and 6 (Supplementary Figure 7a) has only two annotations among 9 residues (Supplementary Figure 7b-c), deprioritizing this pocket as a therapeutic target.

Discussion

GPCRs are the most successful class of drug targets, yet nearly all GPCR drugs act orthosterically, which often comes with severe drawbacks related to specificity and toxicity[7]. Precision modalities such as biased signaling and allosteric modulation could lead to GPCR drugs with improved efficacy and safety, but these drugs have historically only been found by serendipity. Despite a revolution in GPCR structural biology, we lack sufficient functional understanding of GPCR signaling to design biased or allosteric drugs. To address that shortcoming, we have developed and implemented GPCR-MAPS, a new framework for high resolution functional mapping of GPCRs.

Our method makes multiple technical advancements over existing technologies for multiplexed GPCR assessment[26–28]. First, our approach detects receptor activation via protein-protein interaction, instead of via endogenous signaling pathways. This reduces experimental noise and enables more flexible experimental design. Second, GPCR-MAPS is highly modular - we demonstrate here the successful use of a nanobody, a mini-G protein, and β-arrestin as signal transducers. The receptor-proximal readout also facilitates comparisons across these signal transducers, which can otherwise be confounded by differential signal amplification and crosstalk[69]. Validation of additional transducers will continue to enrich the functional readouts possible with the method. To demonstrate generalizability, we also show the vasopressin 2 receptor signaling through miniGs and β-arrestin in response to the endogenous ligand arginine vasopressin (Supplementary Figure 6e). Finally, we extended the platform to enable massively multiplexed GPCR dose-response modeling for the first time (Supplementary Figure 7).

Across >150,000 measurements, we have charted here the first complete map of how mutations affect diverse signaling properties of a human GPCR. After normalizing for expression, residues that specifically disrupt signaling form a contiguously connected core activation network of residues linking the ligand binding site to the transducer coupling site (Figure 8a). Residues important for biased signaling are enriched in the extracellular vestibule, particularly ECL2, as well as TM7 near the NPxxY motif (Figure 8b). >7,500 dose response curves enabled accurate EC_{50} estimation of >5,500 unique receptors. EC_{50} is related to expression level, validating theoretical models[56–58] and agreeing with recent work on plant hormone receptors[59]. Adjusting for expression level, there is a strong enrichment of residues decreasing potency in the second shell of isoprenaline binding, and residues increasing potency in ICL3 (Figure 8c). Residues that decrease potency are in closer proximity to the core activation network and so likely change potency by directly affecting the core activation network. In contrast, residues increasing potency are far from the core activation network, likely affecting potency via mechanisms independent of the core activation network. This distinction in the location and likely mechanism of LoF and GoF variants parallels observations in the Src kinase, where mutations that decrease kinase activity are enriched close to the active site but mutations that increase it are not[70].

We believe that GPCR-MAPS – and related approaches[18–20,27,28] – will provide the foundations for designing new classes of precision GPCR drugs. Residues biased toward Nb80 signaling are enriched in ECL2, which forms part of the exosite of salmeterol, a β2AR agonist strongly biased toward G-protein signaling, suggesting mutant behavior is related to drug activity binding at the

same sites. Likewise, known allosteric modulators bind in pockets with high impact mutations that match their activities. AS408, a NAM, binds in a pocket with loss- and change- of function residues, while Cmpd-6FA, a PAM, binds in a pocket with gain- and change- of function residues. In addition to these known allosteric sites, GPCR-MAPS prioritizes an additional – and, to our knowledge, novel – allosteric pocket to target in β2AR and deprioritizes one where mutations only have limited effects.

The GPCR-MAPS approach is modular and scalable, and its application across the GPCR superfamily will enable a better understanding of the largest family of human receptors. To date GPCR drug discovery has proceeded from an era of blind high-throughput screening to an era of structure-based drug discovery. We believe we are now entering a third era of GPCR drug discovery, where high-resolution functional maps empower the rational design of highly specific biased drugs and allosteric modulators.

Materials and Methods

Plasmids

Plasmids encoding SPARK2 components were gifts from Christina Kim and Alice Ting and were obtained from Addgene (Addgene plasmid # 125228, 104839, 104840, 104841). From these plasmids we generated new plasmids using Gibson cloning. attB-HA-b2AR-eLOV-TEVcs-Gal4-UAS-Citrine encodes the receptor fusion followed by the bGH poly(A) signal, then the UAS sequence followed by Citrine fluorescent protein. A 26 nucleotide barcode was inserted in the 3' UTR of the Citrine sequence. To make cloning more efficient, we transferred the fusion coding sequence from the NanoLuc-15aa linker-βarrestin2-HA-GS linker-TEVp plasmid from the pAAV vector (which contains GC-rich inverted terminal repeats) to the pUC19 vector. We then used site directed mutagenesis to introduce the S153N variant to TEV, generating uTEV1Δ). We then made versions of this plasmid with Nb80 and miniG_s sequence instead of β-Arrestin. We also modified the miniG_s vector so that uTEV1Δ was fused N-terminally (NanoLuc-15aa linker-GS linker-TEVp-miniG_s). pCAG-NLS-HA-Bxb1 was a gift from Kenny Matreyek and Doug Fowler.

Saturation mutagenesis library creation

We designed six pools of mutation-bearing oligonucleotides, each pool representing mutated coding sequence for 66-70 amino acids of the full-length (413 amino acids) b2AR with flanking wild-type sequence for amplification from the pool (using primers dialout_tile[1-6]_[F/R], Supplementary Table 5). For each pool we designed primers to amplify the whole plasmid except for the mutated segment (using primers b2AR_satmut_tile[1-6]_[F/R]), then used Gibson assembly to introduce the mutagenic oligos. Oligos were designed to

encode all possible single amino acid changes (using the human optimized codon for each alternative amino acid), all premature stop codons, and all synonymous codons when possible (using the most optimal codon, excluding the wild-type codon).

Barcode integration and variant-barcode association

10 µg of the attB-HA-b2AR-eLOV-TEVcs-Gal4-UAS-Citrine saturation mutagenesis plasmid was digested with 40 units PpuMI (New England Biolabs) overnight at 37°, then in the morning 10 units more PpuMI were added and 30 minutes more incubation at 37°. Then the cut band was gel purified. The barcode construct (oTLM_257) was made double stranded by PCR amplifying with oTLM_258 and oTLM_259 (Supplementary Table 5). Double stranded barcode was incorporated into the cut plasmid backbone by Gibson assembly and transformed into electrocompetent bacteria. Following recovery in SOC media, we plated 0.1% of the transformation to estimate the number of clones, and inoculated the rest into LB media with ampicillin for overnight growth. Estimated number of clones was 300,000. In the morning the bacteria was pelleted and plasmid was midi-prepped (Qiagen Plasmid Plus Midi Kit).

We used PacBio long read sequencing to associate each variant with the barcode. We digested the plasmid pool with BsiWI (New England Biolabs) to release the coding sequence and barcode and SMRT bell adapters were ligated according to manufacturer's instructions. The library was sequenced on a PacBio Sequel IIe. We used alignparse[71] to align reads, quality filter, and call sequences for each barcode (Supplementary Figure 1).

Landing pad integration

HEK 293T LLP-iCasp9-Blast[31] cells were a gift from Kenny Matreyek. Cells were grown in DMEM with GlutaMAX (Gibco), with Tetracyline-free fetal bovine serum. 14 million cells were plated onto a T225 flask. The next day, transfection was done with Lipofectamine 3000 (Invitrogen) with 17.5 µg of pCAG-NLS-HA-Bxb1 and 17.5 µg of the mutated and barcoded attB-HA-b2AR-eLOV-TEVcs-Gal4-UAS-Citrine. The cells were allowed to incubate for two days. Then, the media was removed and replaced with media containing 2 µg/mL doxycycline and incubated for one day. Then, the media was changed for media containing 2 µg/mL doxycycline and 10 nM rimiducid (MedChemExpress) and incubated for one day. Then media was changed to include just 2 µg/mL doxycycline and outgrowth of cells was allowed for ~one week, passaging cells when they reached 90% confluency.

GPCR-MAPS signaling experiment

10 million cells were plated onto each T175 flask and allowed to incubate overnight. The next day, the cells were transfected with 36 µg of transducer plasmids (pUC19-NanoLuc-β-arrestin-uTEVp1 or pUC19-NanoLuc-Nb80-uTEVp1), using Lipofectamine 3000 and allowed to incubate overnight. The next morning, cells were treated with media containing furimazine and either drug or DMSO: [13.05 mL DMEM, 126.56 µL furimazine, 270 µL 1M HEPES, and 13.5 µL isoproterenol or DMSO]. Furimazine was obtained from the Nano-Glo Live Cell Assay System (Promega). Isoprenaline was obtained from Sigma (catalog number 15627). Cells were incubated in the dark (covered with aluminum foil) at 37° for 24 hours. Then, cells were lifted with Trypsin-EDTA, pelleted and stored in -80 until RNA extraction. For the Gal4 maximum stimulation condition, after 10 million cells were plated and grown overnight, 36

µg of pUC19-CMV-Gal4 was transfected with Lipofectamine 3000. The cells were allowed to incubate for 30 hours, then cells were pelleted and stored at -80°.

RNA isolation and sequencing library preparation

For the first set of experiments (Nb80 and BArr signaling with 10 µM isoprenaline), cell pellets were thawed and resuspended in buffer RLT with added beta-mercaptoethanol. Then, they were homogenized by centrifuging through QIAshredder columns. RNA was then purified from the homogenate with RNeasy Midi-prep kit (Qiagen) and eluted with 400 µL of water. Between 120 and 160 µg of total RNA was used as input for reverse transcription (RT) reactions (6-8 µg per reaction, 20 reactions per sample) using SuperScript III reverse transcriptase (Invitrogen). For the dose response experiment, Dynabeads mRNA direct purification kit (ThermoFisher Scientific) was used according to manufacturers instructions and mRNA was eluted in 80 µL of Tris-HCl. Total mRNA yields were between 11.5 µg and 16 µg. Between 5.75 and 8 µg mRNA was used as input for RT (1.15-1.6 µg per reaction, 10 reactions per sample). For both sets of reactions, RT was carried out with 2 pmol of primer (oTLM_266, Supplementary Table 5) and incubated at 55° for one hour then 75° for 15 minutes. Following this, RNase A was added to a final concentration of 100 µg/mL and reactions incubated at 37° for 30 minutes. Then, reactions were cleaned up with Nucleospin PCR cleanup columns (Macherey Nagel) and eluted in 50 µL of elution buffer. 0.5 µL of the cDNA eluate was used in a qPCR reaction (Q5 High-Fidelity polymerase, New England Biolabs, and EvaGreen fluorescent dye, Biotium) to determine the number of cycles required to reach near-plateau phase. Then, 40 µL of cDNA was amplified across 4x 50 µL PCR

reactions with the program 98° for 30s, followed by 13-16 cycles of [98° for 15s, 64° for 30s, 72° for 30s].

Reactions were column purified and 4% of this was used as input for a second round PCR, which added the rest of the Illumina adapter as well as unique index sequences. The second round PCR was 98° for 30s, followed by 5 cycles of [98° for 15s, 64° for 30s, 72° for 30s]. Frameshifting primers were used to introduce variability in the first few cycles of the sequencing reaction (oTLM_277-282, Supplementary Table 5). Reactions were then gel purified and analyzed on TapeStation (Agilent Technologies). Sequencing was done with Illumina NovaSeq with 2x50 paired end reads.

An additional step was taken for the dose response experiment. Following RT, we performed triplicate qPCR reactions with each cDNA sample, along with a set of serial dilutions of the original library plasmid, in order to estimate the relative quantity of Citrine RNA produced from each experiment. We used primers oTLM_021 and oTLM_024, which amplify the 5' UTR of Citrine. We chose to amplify this region instead of the barcode (as was done for the library prep) to avoid idiosyncratic effects of amplification efficiency caused by diverse and potentially different barcode identities across different samples. Following qPCR we fit a standard curve (Supplementary Fig. 5b) then used this to estimate RNA abundance for the dose response samples (Figure 5b).

Surface expression experiment cell culture and FACS

Cells were cultured in 245x245mm cell culture flasks until ~80% confluent. Then, cells were lifted with Trypsin-EDTA, pelleted, supernatant removed, and resuspended in blocking buffer (1% bovine serum albumin in phosphate buffered saline). Cells were counted and 50 M cells were transferred

to a 5 mL conical tube with 3.5 mL of blocking buffer, and incubated on ice, covered from light, with rocking for 30 minutes. Then, anti-HA antibody (Cell Signaling HA-Tag 6E2 Alexa Fluor 647 Conjugate) was added at dilution of 1:100 and incubated on ice, covered from light, with rocking for 60 more minutes. Then, cells were pelleted, supernatant removed, and resuspended in blocking buffer. Sorting was done on the BD FACSAria instrument. Bins were drawn so that each represented ~25% of cells. In total, ~20M cells were sorted for replicate 1, and ~30M cells were sorted for replicates 2 and 3. 30-45M 2x50 basepair reads were collected on an Illumina NovaSeq instrument for each replicate.

Surface expression sequencing library preparation

Sorted cells were pelleted and DNA extracted using Qiagen DNeasy Blood & Tissue kit, using 2 columns for each sample. 200 µL elution buffer was used to elute from each column, then eluates for each sample were combined. Then, 320 µL of each sample was was used as input for 16x 50 µL PCRs with primers adjacent to the barcode that include partial Illumina adapters and also varying lengths of degenerate bases to increase nucleotide complexity in the first few cycles of the sequencing run (oTLM277-oTLM282, Supplementary Table 5). Cycling was 98° for 30s, followed by 25 cycles of [98° for 15s, 64° for 30s, 72° for 30s]. Reactions were column purified and 4% of this was used as input for a second round PCR, which added the rest of the Illumina adapter as well as unique index sequences. The second round PCR was 98° for 30s, followed by 5 cycles of [98° for 15s, 64° for 30s, 72° for 30s]. The products of the second PCR were gel purified and analyzed on TapeStation, then sequenced on Illumina NovaSeq with 2x50 paired end reads.

Sequencing data processing

Paired-end sequencing reads were first merged with vsearch[72] and then adapters were trimmed with cutadapt[73]. This leaves each read representing the barcode sequence, which was then matched with barcodes identified in the variant-barcode association with a custom script.

Calculation of surface expression scores

Surface expression scores were calculated as previously described[36]. Briefly, the frequency of each variant in each bin is enumerated by combining counts of all barcodes associated with that variant. The frequency, along with the total number of cells sorted into each bin, is used to estimate the number of cells in each bin. Then, the log10 transformed geometric mean fluorescence value of all cells sorted into each bin is combined with the estimated cell counts to arrive at the raw surface expression estimate. These estimates were then normalized between 0 (represented by the median of presumed loss-of-function variants: e.g. premature stop codons before the 300th amino acid) and 1, the WT score. Surface expression scores were considered high confidence if the estimated number of cells sorted was >50.

Calculation of signaling scores

First, barcode read counts for all barcodes representing the same variant were combined and then the frequency of that variant was calculated by dividing that number by the total number of barcode counts in the sample. Then, to arrive at a raw score, we divided the variant frequency in the experimental sample by the variant frequency in the Gal4 sample, and take the $\log_2$ of this. The Gal4 sample represents a maximum stimulation condition in which the barcode frequency in the library, as well as any idiosyncratic effects

of the barcode sequence on RNA stability should be captured. Raw scores are then scaled between 0 and 1, with the median of nonsense mutations set as 0 and the wild-type score set as 1. Signaling scores were considered high confidence if the frequency of that variant in the Gal4 condition was $>2x10^{-5}$.

LOWESS modeling and identifying outlier positions

The python package statsmodels.nonparametric.smoothers_lowess[74] was used with the following parameter values: frac=0.5, it=3, delta=0.0. Following curve fitting, residuals were calculated as the distance in the y-axis between the data point and the line, and z-scores were calculated as the residual / standard error of the mean. In the case of Nb80 versus β-arrestin, the average standard error of the mean between the two measurements was used.

Structural analysis

Residue contacts were defined by GetContacts (https://getcontacts.github.io/index.html). The binding interface of β2AR with Nb80 was defined by GetContacts analysis of structure. The predicted inactive and active structures in Figure 6k were taken from GPCRdb and were produced with AlphaFold-Multistate[37]. These structures were used to enable visualization of the entire ICL3, which is not resolved in experimentally solved structures.

Dose response curve fitting

In the package neutcurve[75], we used the neutcurve.hillcurve.HillCurve function to fit curves with the following options: infectivity_or_neutralized='neutralized', fix_slope_first=True, fs_stderr=None, fixbottom=False, fixtop=False, fixslope=False, fitlogc=False, use_stderr_for_fit=False, no_curve_fit_first=False. Following the fitting procedure, we then filtered curves based on heuristics: $R^2 > 0.9$, EC50 within

the measured range (i.e. between 10pM and 10uM), and E_{max} < 1.5x the maximum measured signaling value. See Supplementary Figure 5e for the number of variants filtered.

Fpocket pocket prediction

We used PDB ID 2RH1 as the structural template for pocket prediction and used the Fpocket web server with default settings to predict pockets. We chose the two pockets to illustrate simply based on having many or few annotated residues.

Acknowledgments

We thank Christina Kim for advice regarding the SPARK2 system. We thank the CRG/UPF Flow Cytometry Unit for technical assistance. We thank Jana Selent and lab members for helpful discussion. We thank all members of the Lehner lab for helpful discussion.

Data availability

Files needed to reproduce analyses can be found at zenodo (https://zenodo.org/records/15363336). Raw sequencing reads can be found at Sequence Read Archive (accession number PRJNA1260170).

Code availability

Custom code to reproduce analyses can be found at github (https://github.com/lehner-lab/GPCR-MAPS).

Competing interests

CRG and the Wellcome Sanger Institute have filed a patent application for the use of the methods described in this study for mapping allostery in GPCRs; T.M. and B.L. are listed as inventors. B.L. is a founder and shareholder of ALLOX.

Figure 1. GPCR-MAPs

a Schematic of the experimental framework. A mutagenized GPCR (in this case β2AR) is encoded in a plasmid that also includes the attB recombination site, an N-terminal HA epitope tag, C-terminal fusions of a light, oxygen, and voltage sensitive (LOV) domain, the cleavage site of TEV protease (TEVcs), and Gal4 transcription factor. Downstream, on the same plasmid, is the fluorescent protein Citrine, driven by Gal4-responsive UAS promoter, with a degenerate DNA barcode included in the 3' untranslated region. This construct is recombined into HEK 293T LLP-iCasp9-Blast landing pad cells. Then, two assays are performed to measure the surface expression (FACS-based) and signaling activity (transcriptional reporter) of variants *en masse***.** The outputs of these experiments (DNA for FACS and RNA for transcriptional reporter) are sequenced with short read technology. Estimated surface expression and signaling scores are then modeled to understand the contribution of each of these to the variant effects. **b** Detailed schematic of the signaling assay. The receptor construct is expressed from landing pad cells while the transducer construct is transiently transfected. Addition of agonist (isoprenaline in this case) and furimazine results in recruitment of transducer to the receptor fusion and subsequent cleavage of the TEVcs, liberating Gal4 which moves to the nucleus and drives transcription of the UAS-Citrine-barcode construct. **c** Replicate Pearson correlations for the surface expression, Nb80 signaling, and βArr signaling. $p<10^{-100}$ in each case. **d** Heatmap representation of the surface expression data. β2AR positions on the x-axis, mutation on the y-axis. Gray cells indicate expression similar to wild-type while blue and red cells indicate reduced and increased expression, respectively. **e** Histogram of surface

expression variant scores; nonsense variants in red, synonymous wild-type in green, and missense in grey. **f** Heatmap representation of the Nb80 signaling data. Axes and color scale as in (d). **g** Histogram of Nb80 signaling scores; colors as in (e). **h** Heatmap representation of β-Arrestin signaling scores. Axes and color scale as in (d). **i** Histogram of β-Arrestin signaling scores; colors as in (e). **j** Surface expression scores for residues that directly contact isoprenaline in the solved structure (PDB ID: 7DHR). **k** Nb80 signaling scores for residues that directly contact isoprenaline. **l** β-Arrestin signaling scores for residues that directly contact isoprenaline.

Figure 2. Deconvoluting expression and signaling effects

a Surface expression versus Nb80 signaling scores for missense variants. $p<10^{-100}$. **b** Surface expression versus β-Arrestin signaling scores for missense variants. $p<10^{-100}$. **c** Schematic of the method for identifying residues contributing to signaling beyond what can be explained by surface expression. First, the signaling data is related to surface expression with a LOWESS regression. Then, for each variant, residuals to this fit are calculated. Then, residues with significantly biased residuals are identified as loss-of-function or gain-of-function via Mann Whitney U test. Finally, positions where variants have a wide range of effects, quantified using the interquartile range of variant z-scores, are identified as change-of-function. **d** LOWESS fit of surface expression versus Nb80 signaling, with variant points colored by residual to the fit. **d** Same as (e) but for β-Arrestin signaling. **f** Volcano plot showing the position-average z-score and the minus log10 transformed p-value of all residues for Nb80 signaling. Residues with significantly biased residuals are

colored according to the average z-score. **g** Same as in (f) but for β-Arrestin signaling. **h** Venn diagrams showing the number of residues called loss-, gain-, and chage-of-function for Nb80 and β-Arrestin signaling. **i** Surface expression and signaling effects of variants at D113^{3x32}, plotted over the population of variant effects, for Nb80 and β-Arrestin versus surface expression. Error bars represent standard error of the mean. **j** Same as (i) but with variants at position E268^{6x30}. **k** Same as (i) and (j) but with variants at position S120^{3x39}. **l** Loss-of-function residues highlighted in purple on the isoprenaline-bound β2AR structure (PDB ID: 7DHR, isoprenaline in yellow). **m** Gain-of-function residues highlighted in green on β2AR. **n** Change-of-function residues highlighted in red on β2AR

Figure 3. A core activation network links ligands to transducers

a Inactive and active state structures of β2AR with LoF residues colored purple. Those LoF residues that make a direct contact with other LoF residues are visualized as pseudobonds connecting them, while disconnected LoF residues are shown with spheres. **b** Enlarged view of the core activation network.

Figure 4. Genetic basis of biased signaling

a Schematic showing β2AR signaling through β-Arrestin and G-protein (Nb80) mediated pathways in response to isoprenaline. **b** Nb80 versus β-Arrestin signaling for missense variants. $p<10^{-100}$, n=6,791. **c** Nb80 versus β-Arrestin signaling with a LOWESS fit and variant scores colored by residual to the fit line. Individual variant scores, with error bars representing standard error of the mean, shown for three variants previously reported to cause biased signaling. **d**

Volcano plot showing the position-average z-score and the minus log10 transformed p-value of all residues for Nb80 versus β-Arrestin signaling. Residues with significantly biased residuals are colored according to the average z-score. **e** Nb80 and β-Arrestin signaling effects of variants at W99^{23x50}, plotted over the population of variant effects, error bars represent standard error of the mean. **f** Same as in (e) but with variants at P138^{34x50}. **g** Same as in (e) but with variants at F332^{8x50}. **h** β-Arrestin biased, Nb80 biased, or bias switch residues highlighted on the solved structure, 7DHR. Isoprenaline show in yellow. **i** F193^{45x52} and Y318^{7x34} form a "lid" over isoprenaline in the solved structure (7DHR). **j** Hydrophobicity (Kyte-Doolittle scale) of the mutant amino acid compared with the Nb80 versus β-Arrestin signaling bias. **k** Enlarged view of the extracellular vestibule. Colored residues are significantly Nb80-biased (orange), β-Arrestin-biased (blue), or bias switches (magenta). **l** Enlarged view of NPxxY motif and TM7, colored residues same as in (k).

Figure 5. Quantifying >7,500 dose-response relationships

a Experimental design: serial dilutions of isoprenaline were used to stimulate landing pad cells expressing the β2AR library. Then, the GPCR-MAPS signaling assay was performed and RNA was collected from the stimulated cells. **b** The RNA abundance from each sample was first quantified with qPCR, then the variant frequencies were estimated with short-read sequencing. **c** Per-variant normalized scores were then calculated by normalizing variant frequencies in each condition by the RNA abundance of each condition. **d** Heatmap representations of the 8 highest concentration drug treatment conditions. β2AR positions on the x-axis, mutation identities on the y-axis. One y-axis is blown up;

the rest of the heatmaps follow this same scheme. Grey cells indicate wild-type-like signaling, blue and red indicate reduced and increased signaling, respectively. One colorbar is blown up; the rest of the heatmaps follow this same scheme. **e** Clustered heatmap of signaling intensities. Colorbar on top. Three main clusters correspond to wild-type-like, moderately damaging, and complete loss-of-function curves. **f** The curve fit to the wild-type data, with E_{max}, E_{min}**,** EC_{50}, and slope indicated. **g** Fit curves from the wild-type-like cluster in black, wild-type curve in red. **h** Fit curves from the moderately damaging cluster in black, wild-type curve in red. **i** Fit curves from the loss-of-function cluster in black, wild-type curve in red.

Figure 6. Genetic determinants of potency

a Schematic showing example curves for increased and decreased potency. **b** Comparison of variant EC_{50} estimates from this study with previously published estimates[18]. Red dashed line is x=y line. **c** Spearman correlations of curve fitting parameter estimates and surface expression scores. **d** Surface expression scores compared with log-transformed EC_{50} estimates. $p<10^{-100}$. **e** Volcano plot showing the position-average z-score and the minus log10 transformed p-value of all residues for log EC_{50} versus surface expression. **f** Log-transformed EC_{50} versus surface expression for variants at $R328^{7x55}$, plotted over the population of variant effects, error bars represent standard error of the mean. **g** Same as in (f) but with variants at $E268^{6x30}$. **h** Same as in (f) but with variants at $F282^{6x44}$. **i** EC_{50} loss-of-function, gain-of-function, or change-of-function residues highlighted on the solved structure, 7DHR. Isoprenaline show in yellow. **j** Violin plots showing minimum distance between LoF, GoF, or CoF

residues from the core activation network in AlphaFold-Multistate active state model. Distances capped at 20Å. **k** Portions of TMs 3, 5, and 6, as predicted by AlphaFold-Multistate and downloaded from GPCRdb[37]. LoF residues colored aquamarine, GoF residues colored brown, and CoF residues colored salmon.

Figure 7. Rationalization and prioritization of allosteric pockets

a The binding pocket of known negative allosteric modulator, AS408, shown in yellow. PDB ID: 6OBA. Residues making direct contact with the NAM are colored beige. **b** Expression normalized maximum signaling through Nb80, signaling bias, and expression normalized potency of the mutations at sites making up the binding pocket of AS408. Shown below the heatmaps are the categorical annotations of significant residues: purple, green, and red indicate LoF, GoF, and CoF residues for maximum signaling. Orange, blue, and magenta indicate Nb80 biased, β-Arrestin, biased, or bias switches. Aquamarine, brown, and salmon indicate LoF, GoF, and CoF for potency. **c** The binding pocket of AS408 with annotated residues colored. For residues annotated in both maximum signaling and potency analyses, the color from maximum signaling was used. **d, e, f** Same as in (a, b, c) but for the positive allosteric modulator Cmpd-6FA, with structure 6N48. **g, h, i** Same as in (a, b, c) but for a predicted active pocket in structure 2RH1. **j, k, l** Same as in (a, b, c) but for a predicted inactive pocket in structure 2RH1.

Figure 8. Summary of findings.

a Selected findings from the maximum signaling analysis. Core activation network shown as purple lines, residues or topological domains colored

according to annotation. **b** Selected findings from biased signaling analysis. NPxxY motif collectively illustrated as “bias switch” as there are multiple bias functional residues in and near this motif. **c** Selected findings from the potency analysis. Ligand binding and second shell residues illustrated as triangle.

References

1. Bockaert, J. & Philippe Pin, J. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J.* **18**, 1723–1729 (1999).
2. Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L. & Khorana, H. G. Requirement of Rigid-Body Motion of Transmembrane Helices for Light Activation of Rhodopsin. *Science* **274**, 768–770 (1996).
3. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
4. Jean-Charles, P.-Y., Kaur, S. & Shenoy, S. K. GPCR signaling via β-arrestin-dependent mechanisms. *J. Cardiovasc. Pharmacol.* **70**, 142–158 (2017).
5. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
6. Sriram, K. & Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol. Pharmacol.* **93**, 251–258 (2018).
7. Wold, E. A., Chen, J., Cunningham, K. A. & Zhou, J. Allosteric Modulation of Class A GPCRs: Targets, Agents, and Emerging Concepts. *J. Med. Chem.* **62**, 88–127 (2019).
8. Wootten, D., Christopoulos, A., Marti-Solano, M., Babu, M. M. & Sexton, P. M. Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nat. Rev. Mol. Cell Biol.* **19**, 638–653 (2018).
9. May, L. T., Leach, K., Sexton, P. M. & Christopoulos, A. Allosteric modulation of G protein-coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* **47**, 1–51 (2007).
10. Kenakin, T. Biased Receptor Signaling in Drug Discovery. *Pharmacol. Rev.* **71**, 267–315 (2019).

1063 11. Wootten, D., Christopoulos, A. & Sexton, P. M. Emerging paradigms in GPCR
1064 allostery: implications for drug discovery. *Nat. Rev. Drug Discov.* **12**, 630–644
1065 (2013).
1066 12. Lorente, J. S. *et al.* GPCR drug discovery: new agents, targets and indications.
1067 *Nat. Rev. Drug Discov.* 1–22 (2025) doi:10.1038/s41573-025-01139-y.
1068 13. Bermudez, M., Nguyen, T. N., Omieczynski, C. & Wolber, G. Strategies for the
1069 discovery of biased GPCR ligands. *Drug Discov. Today* **24**, 1031–1037 (2019).
1070 14. Allosteric Modulator Discovery: From Serendipity to Structure-Based Design. *J.*
1071 *Med. Chem.* **62**, 6405–6421 (2019).
1072 15. Venkatakrishnan, A. J. *et al.* Diverse activation pathways in class A GPCRs
1073 converge near the G-protein-coupling region. *Nature* **536**, 484–487 (2016).
1074 16. Zhou, Q. *et al.* Common activation mechanism of class A GPCRs. *eLife* **8**, e50279
1075 (2019).
1076 17. Ward, W. H. J., Timms, D. & Fersht, A. R. Protein engineering and the study of
1077 structure—function relationships in receptors. *Trends Pharmacol. Sci.* **11**, 280–284
1078 (1990).
1079 18. Heydenreich, F. M. *et al.* Molecular determinants of ligand efficacy and potency
1080 in GPCR signaling. *Science* **382**, eadh1859 (2023).
1081 19. Morales-Pastor, A. *et al.* Multiple triggers converge to preferential effector
1082 coupling in the CB2R through a complex allosteric communication network.
1083 2024.02.27.582279 Preprint at https://doi.org/10.1101/2024.02.27.582279 (2024).
1084 20. Cao, Y. *et al.* Unraveling allostery within the angiotensin II type 1 receptor for
1085 Gαq and β-arrestin coupling. *Sci. Signal.* **16**, eadf2173 (2023).

1086 21. Penn, W. D. *et al.* Probing biophysical sequence constraints within the
1087 transmembrane domains of rhodopsin by deep mutational scanning. *Sci. Adv.* **6**,
1088 eaay7505 (2020).
1089 22. McKee, A. G. *et al.* Systematic profiling of temperature- and retinal-sensitive
1090 rhodopsin variants by deep mutational scanning. *J. Biol. Chem.* **297**, 101359 (2021).
1091 23. Heredia, J. D. *et al.* Mapping Interaction Sites on Human Chemokine Receptors
1092 by Deep Mutational Scanning. *J. Immunol.* **200**, 3825–3839 (2018).
1093 24. Park, J. *et al.* Structural architecture of a dimeric class C GPCR based on co-
1094 trafficking of sweet taste receptor subunits. *J. Biol. Chem.* **294**, 4759–4774 (2019).
1095 25. Yoo, J. I., Daugherty, P. S. & O'Malley, M. A. Bridging non-overlapping reads
1096 illuminates high-order epistasis between distal protein sites in a GPCR. *Nat.*
1097 *Commun.* **11**, 690 (2020).
1098 26. Jones, E. M. *et al.* Structural and functional characterization of G protein–
1099 coupled receptors with deep mutational scanning. *eLife* **9**, e54895 (2020).
1100 27. Howard, M. K. *et al.* Molecular basis of proton sensing by G protein-coupled
1101 receptors. *Cell* **188**, 671-687.e20 (2025).
1102 28. Howard, C. J. *et al.* High resolution deep mutational scanning of the
1103 melanocortin-4 receptor enables target characterization for drug discovery. *eLife*
1104 **13**, (2024).
1105 29. Ippolito, M. & Benovic, J. L. Biased agonism at β-adrenergic receptors. *Cell.*
1106 *Signal.* **80**, 109905 (2021).
1107 30. Kim, C. K., Cho, K. F., Kim, M. W. & Ting, A. Y. Luciferase-LOV BRET enables
1108 versatile and specific transcriptional readout of cellular protein-protein interactions.
1109 *eLife* **8**, e43826 (2019).

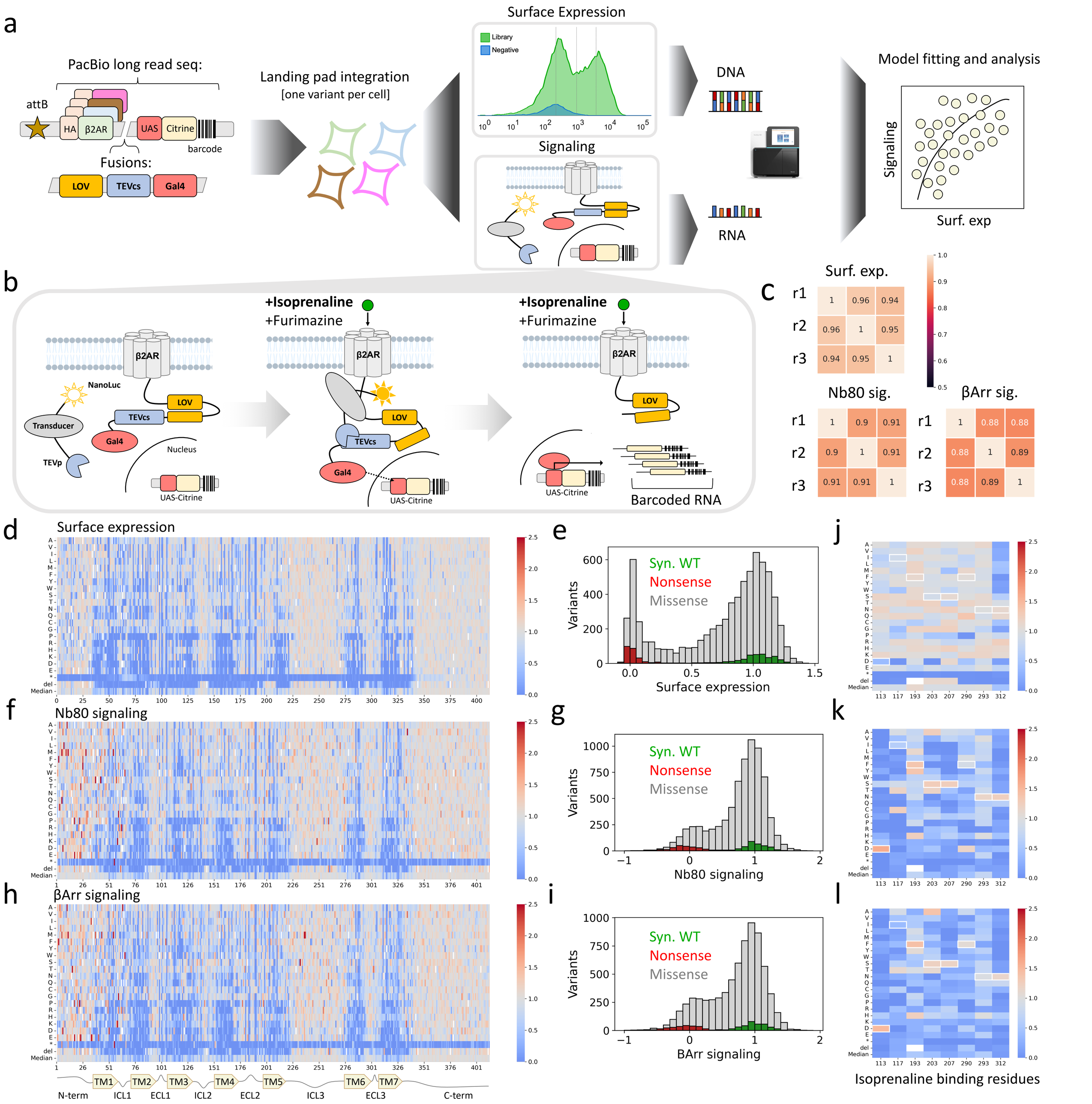
1110 31. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An
1111 improved platform for functional assessment of large protein libraries in
1112 mammalian cells. *Nucleic Acids Res.* **48**, e1 (2020).
1113 32. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the
1114 β2 adrenoceptor. *Nature* **469**, 175–180 (2011).
1115 33. Sanchez, M. I. & Ting, A. Y. Directed evolution improves the catalytic efficiency
1116 of TEV protease. *Nat. Methods* **17**, 167–174 (2020).
1117 34. Kim, M. W. *et al.* Time-gated detection of protein-protein interactions with
1118 transcriptional readout. *eLife* **6**, e30233 (2017).
1119 35. Carpenter, B. & Tate, C. G. Engineering a minimal G protein to facilitate
1120 crystallisation of G protein-coupled receptors in their active conformation. *Protein*
1121 *Eng. Des. Sel.* **29**, 583–594 (2016).
1122 36. Mighell, T. L. & Lehner, B. A pharmacological chaperone stabilizer rescues the
1123 expression of the vast majority of pathogenic variants in a G protein-coupled
1124 receptor. 2024.11.28.625821 Preprint at
1125 https://doi.org/10.1101/2024.11.28.625821 (2024).
1126 37. Pándy-Szekeres, G. *et al.* GPCRdb in 2023: state-specific structure models using
1127 AlphaFold2 and new ligand resources. *Nucleic Acids Res.* **51**, D395–D402 (2023).
1128 38. Valentin-Hansen, L. *et al.* The Arginine of the DRY Motif in Transmembrane
1129 Segment III Functions as a Balancing Micro-switch in the Activation of the β2-
1130 Adrenergic Receptor. *J. Biol. Chem.* **287**, 31973–31982 (2012).
1131 39. Sanchez-Reyes, O. B. *et al.* G Protein-Coupled Receptors Contain Two
1132 Conserved Packing Clusters. *Biophys. J.* **112**, 2315–2326 (2017).

1133 40. Nygaard, R., Frimurer, T. M., Holst, B., Rosenkilde, M. M. & Schwartz, T. W.
1134 Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol.*
1135 *Sci.* **30**, 249–259 (2009).
1136 41. Yang, D. *et al.* G protein-coupled receptors: structure- and function-based drug
1137 discovery. *Signal Transduct. Target. Ther.* **6**, 1–27 (2021).
1138 42. Ragnarsson, L., Andersson, Å., Thomas, W. G. & Lewis, R. J. Mutations in the
1139 NPxxY motif stabilize pharmacologically distinct conformational states of the α1B-
1140 and β2-adrenoceptors. *Sci. Signal.* **12**, eaas9485 (2019).
1141 43. Zhao, L., He, X., Jiang, H. & Cheng, X. Computational characterization of
1142 transducer recognition of β2 adrenergic receptor. *Biochem. Biophys. Res. Commun.*
1143 **592**, 67–73 (2022).
1144 44. Rovati, G. E., Capra, V. & Neubig, R. R. The highly conserved DRY motif of class
1145 A G protein-coupled receptors: beyond the ground state. *Mol. Pharmacol.* **71**, 959–
1146 964 (2007).
1147 45. Rasmussen, S. G. F. *et al.* Mutation of a Highly Conserved Aspartic Acid in the
1148 *β*2 Adrenergic Receptor: Constitutive Activation, Structural Instability, and
1149 Conformational Rearrangement of Transmembrane Segment 6. *Mol. Pharmacol.* **56**,
1150 175–184 (1999).
1151 46. Tao, Y. X., Abell, A. N., Liu, X., Nakamura, K. & Segaloff, D. L. Constitutive
1152 activation of G protein-coupled receptors as a result of selective substitution of a
1153 conserved leucine residue in transmembrane helix III. *Mol. Endocrinol. Baltim. Md*
1154 **14**, 1272–1282 (2000).
1155 47. Wheatley, M. *et al.* Lifting the lid on GPCRs: the role of extracellular loops. *Br. J.*
1156 *Pharmacol.* **165**, 1688–1703 (2012).

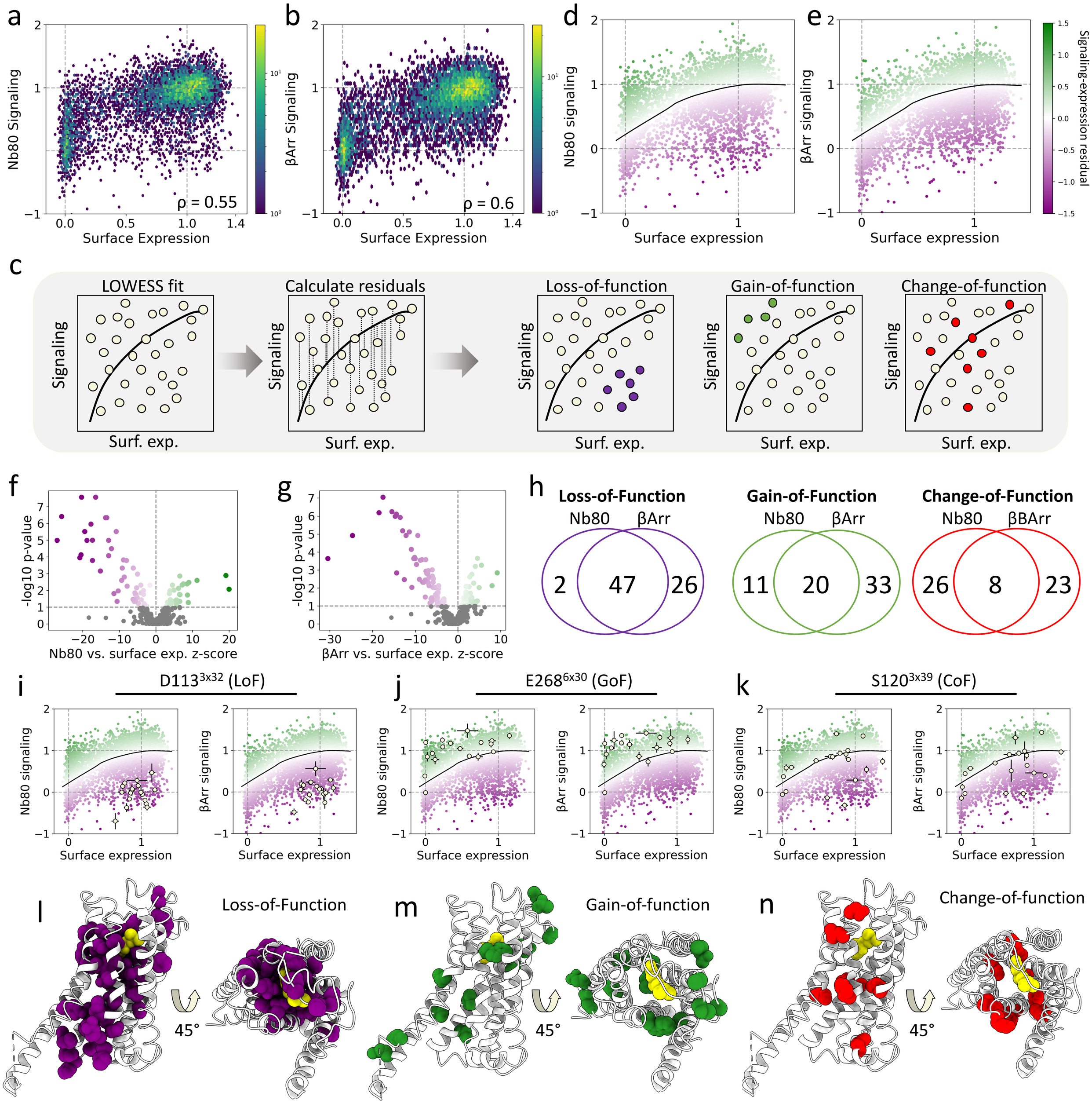
1157 48. Ippolito, M., De Pascali, F., Inoue, A. & Benovic, J. L. Phenylalanine 193 in
1158 Extracellular Loop 2 of the β2-Adrenergic Receptor Coordinates β-Arrestin
1159 Interaction. *Mol. Pharmacol.* **101**, 87–94 (2022).
1160 49. Choi, M. *et al.* G protein–coupled receptor kinases (GRKs) orchestrate biased
1161 agonism at the β2-adrenergic receptor. *Sci. Signal.* **11**, eaar7084 (2018).
1162 50. Sanchez-Soto, M. *et al.* A structural basis for how ligand binding site changes
1163 can allosterically regulate GPCR signaling and engender functional selectivity. *Sci.*
1164 *Signal.* **13**, eaaw5885 (2020).
1165 51. Masureel, M. *et al.* Structural insights into binding specificity, efficacy and bias
1166 of a β2AR partial agonist. *Nat. Chem. Biol.* **14**, 1059–1066 (2018).
1167 52. Bock, A. & Bermudez, M. Allosteric coupling and biased agonism in G protein-
1168 coupled receptors. *FEBS J.* **288**, 2513–2528 (2021).
1169 53. Woo, A. Y.-H. *et al.* Tyrosine 308 is necessary for ligand-directed Gs protein-
1170 biased signaling of β2-adrenoceptor. *J. Biol. Chem.* **289**, 19351–19363 (2014).
1171 54. Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled
1172 receptors. *Proc. Natl. Acad. Sci.* **108**, 13118–13123 (2011).
1173 55. Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. & Wüthrich, K. Biased Signaling
1174 Pathways in β2-Adrenergic Receptor Characterized by 19F-NMR. *Science* **335**, 1106–
1175 1110 (2012).
1176 56. Chure, G. *et al.* Predictive shifts in free energy couple mutations to their
1177 phenotypic consequences. *Proc. Natl. Acad. Sci.* **116**, 18275–18284 (2019).
1178 57. Razo-Mejia, M. *et al.* Tuning Transcriptional Regulation through Signaling: A
1179 Predictive Theory of Allosteric Induction. *Cell Syst.* **6**, 456-469.e10 (2018).

1180 58. Steiner, P. J. *et al.* A Closed Form Model for Molecular Ratchet-Type Chemically
1181 Induced Dimerization Modules. *Biochemistry* **62**, 281–291 (2023).
1182 59. Maximilian R Stammnitz & Ben Lehner. The genetic architecture of an allosteric
1183 hormone receptor. *bioRxiv* (2025).
1184 60. Weis, W. I. & Kobilka, B. K. The Molecular Basis of G Protein–Coupled Receptor
1185 Activation. *Annu. Rev. Biochem.* **87**, 897–919 (2018).
1186 61. Sadler, F. *et al.* Autoregulation of GPCR signalling through the third intracellular
1187 loop. *Nature* **615**, 734–741 (2023).
1188 62. Li, X., Zhou, M., Huang, W. & Yang, H. N-glycosylation of the β2 adrenergic
1189 receptor regulates receptor function by modulating dimerization. *FEBS J.* **284**, 2004–
1190 2018 (2017).
1191 63. Chen, S. *et al.* Mutation of a Single TMVI Residue, Phe282, in the β2-Adrenergic
1192 Receptor Results in Structurally Distinct Activated Receptor Conformations.
1193 *Biochemistry* **41**, 6045–6053 (2002).
1194 64. Liu, X. *et al.* Mechanism of intracellular allosteric β2AR antagonist revealed by
1195 X-ray crystal structure. *Nature* **548**, 480–484 (2017).
1196 65. Liu, X. *et al.* An allosteric modulator binds to a conformational hub in the β2
1197 adrenergic receptor. *Nat. Chem. Biol.* **16**, 749–755 (2020).
1198 66. Liu, X. *et al.* Mechanism of β2AR regulation by an intracellular positive allosteric
1199 modulator. *Science* **364**, 1283–1287 (2019).
1200 67. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An open source platform
1201 for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).
1202 68. Chen, G. *et al.* Activation and allosteric regulation of the orphan GPR88-Gi1
1203 signaling complex. *Nat. Commun.* **13**, 2375 (2022).

1204 69. Onaran, H. O. *et al.* Systematic errors in detecting biased agonism: Analysis of
1205 current methods and development of a new model-free approach. *Sci. Rep.* **7**,
1206 44247 (2017).
1207 70. Beltran, A., Faure, A. J. & Lehner, B. The allosteric landscape of the Src kinase.
1208 2024.04.26.591297 Preprint at https://doi.org/10.1101/2024.04.26.591297 (2024).
1209 71. Crawford, K. H. d & Bloom, J. D. alignparse: A Python package for parsing
1210 complex features from high-throughput long-read sequencing. *J. Open Source*
1211 *Softw.* **4**, 1915 (2019).
1212 72. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile
1213 open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
1214 73. Martin, M. Cutadapt removes adapter sequences from high-throughput
1215 sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
1216 74. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling
1217 with Python. *Proc. 9th Python Sci. Conf.* 92–96 (2010) doi:10.25080/Majora-
1218 92bf1922-011.
1219 75. Loes, A. N. *et al.* High-throughput sequencing-based neutralization assay
1220 reveals how repeated vaccinations impact titers to recent human H1N1 influenza
1221 strains. *J. Virol.* **98**, e00689-24 (2024).
1222



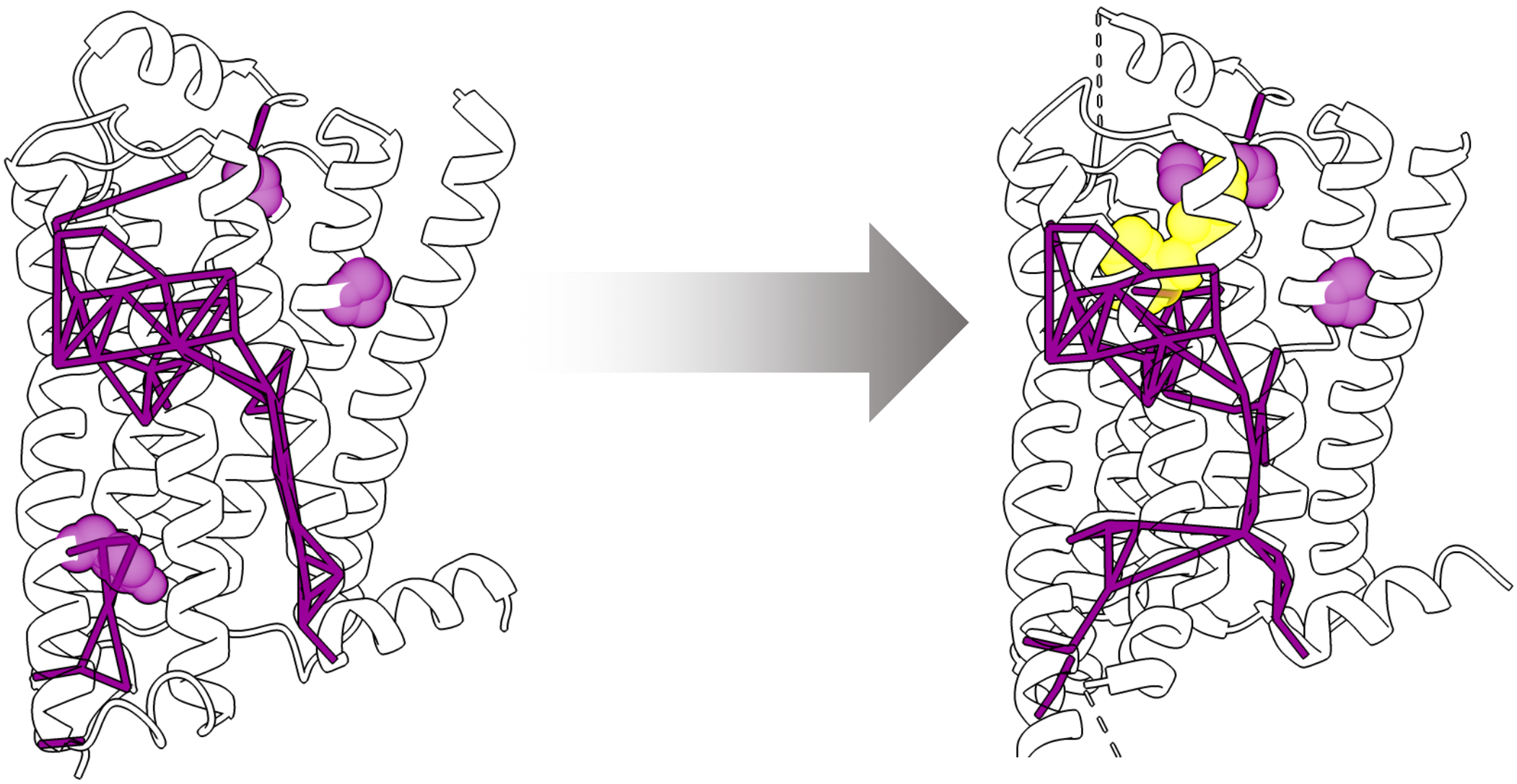
a
PacBio long read seq:
attB
HA
β2AR
UAS
Citrine
barcode
Fusions:
LOV
TEVcs
Gal4
Landing pad integration
[one variant per cell]
Surface Expression
Library
Negative
Signaling
DNA
RNA
Model fitting and analysis
Signaling
Surf. exp
b
+Isoprenaline
+Furimazine
β2AR
NanoLuc
Transducer
TEVp
LOV
TEVcs
Gal4
Nucleus
UAS-Citrine
Barcoded RNA
c
Surf. exp.
r1 1 0.96 0.94
r2 0.96 1 0.95
r3 0.94 0.95 1
Nb80 sig.
r1 1 0.9 0.91
r2 0.9 1 0.91
r3 0.91 0.91 1
βArr sig.
r1 1 0.88 0.88
r2 0.88 1 0.89
r3 0.88 0.89 1
d
Surface expression
e
Syn. WT
Nonsense
Missense
Variants
Surface expression
j
f
Nb80 signaling
g
Nb80 signaling
k
h
βArr signaling
i
BArr signaling
l
Isoprenaline binding residues
TM1
TM2
TM3
TM4
TM5
TM6
TM7
N-term
ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
C-term


a
Nb80 Signaling
Surface Expression
ρ = 0.55
b
βArr Signaling
Surface Expression
ρ = 0.6
d
Nb80 signaling
Surface expression
e
βArr signaling
Surface expression
Signaling-expression residual
c
LOWESS fit
Calculate residuals
Loss-of-function
Gain-of-function
Change-of-function
Signaling
Surf. exp.
f
-log10 p-value
Nb80 vs. surface exp. z-score
g
βArr vs. surface exp. z-score
h
Loss-of-Function
Nb80
βArr
2
47
26
Gain-of-Function
Nb80
βArr
11
20
33
Change-of-Function
Nb80
βBArr
26
8
23
i
D113^3x32 (LoF)
j
E268^6x30 (GoF)
k
S120^3x39 (CoF)
Nb80 signaling
βArr signaling
Surface expression
l
Loss-of-Function
45°
m
Gain-of-function
45°
n
Change-of-function
45°

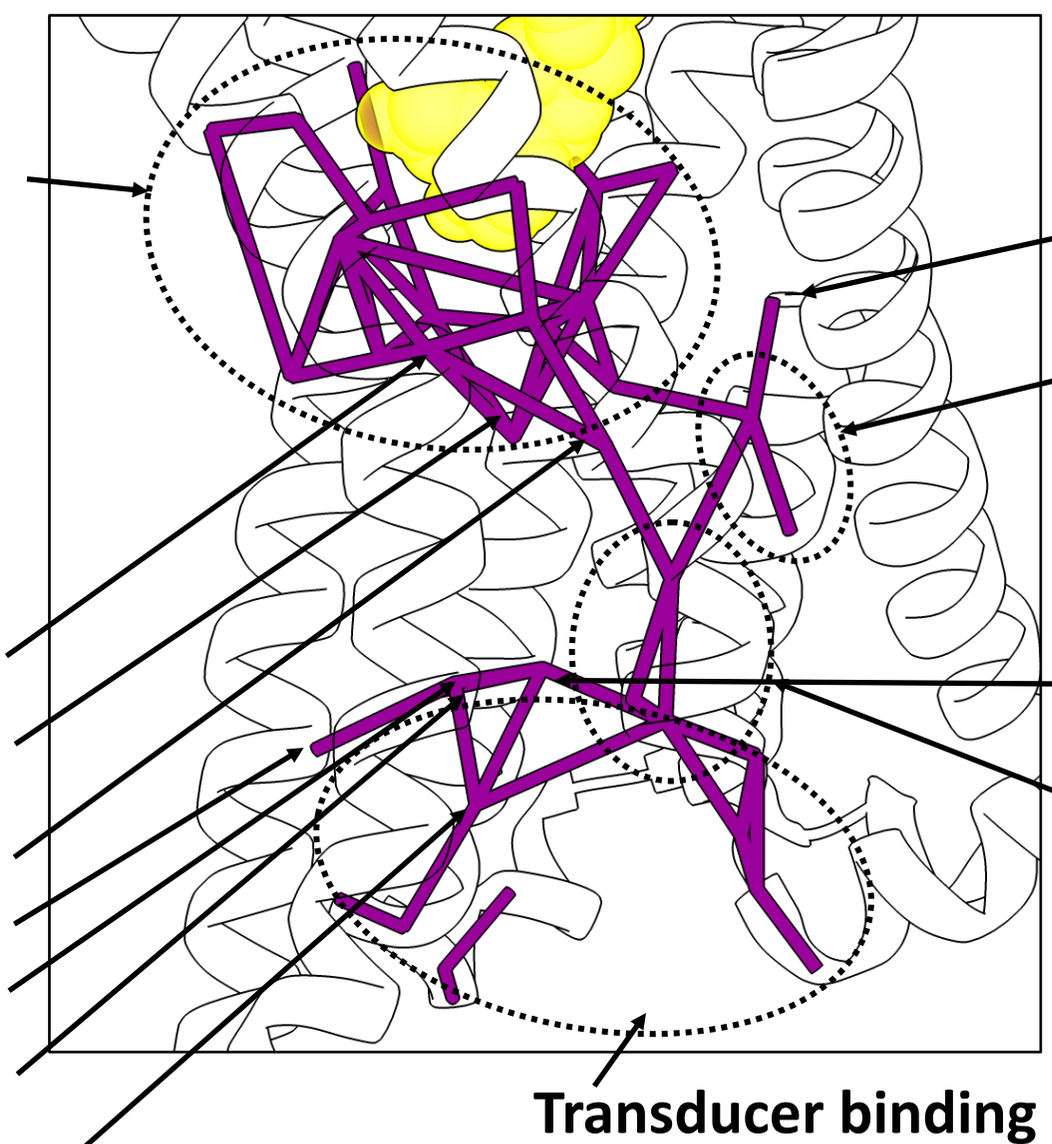
a

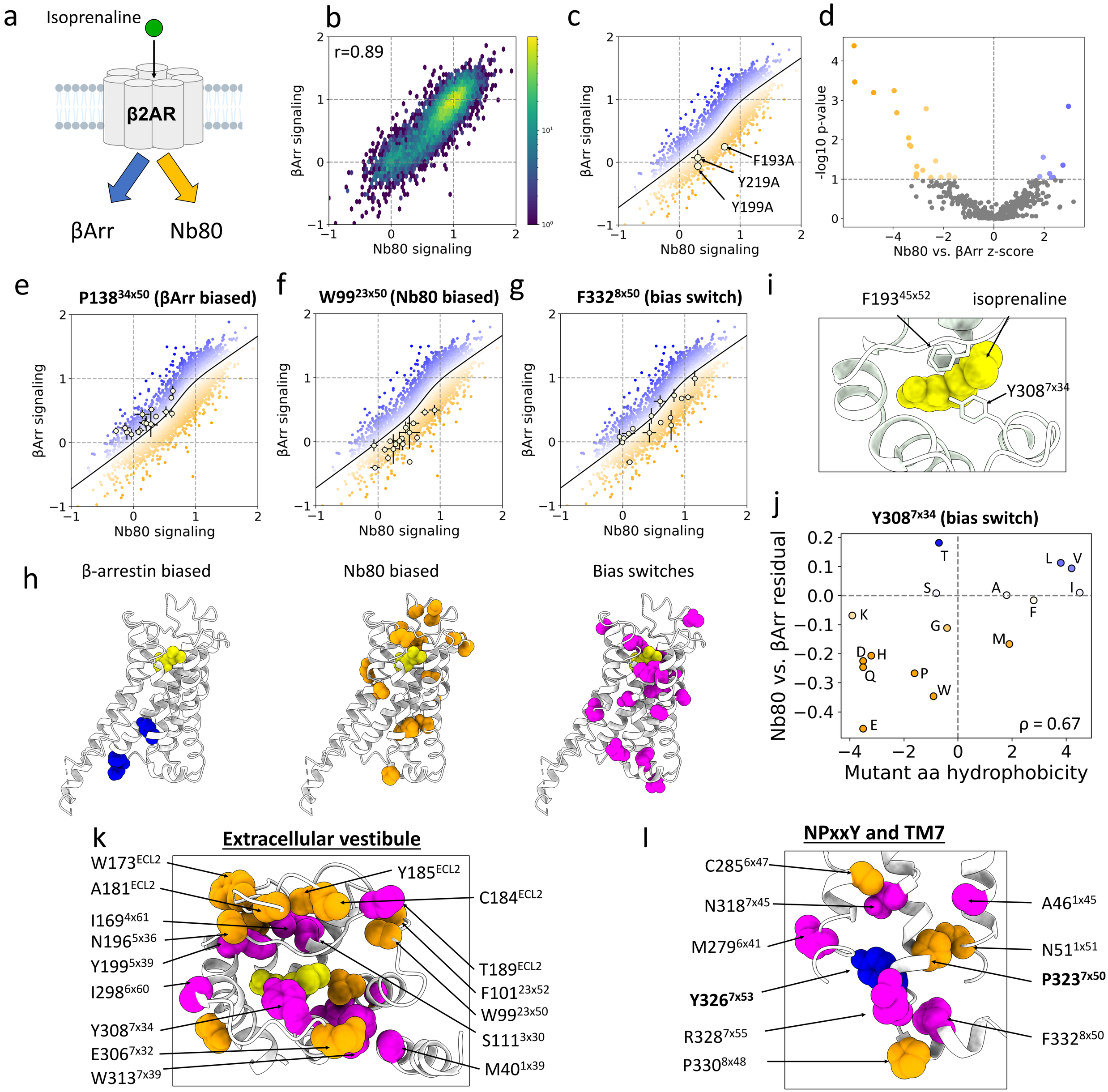
Inactive (2RH1)

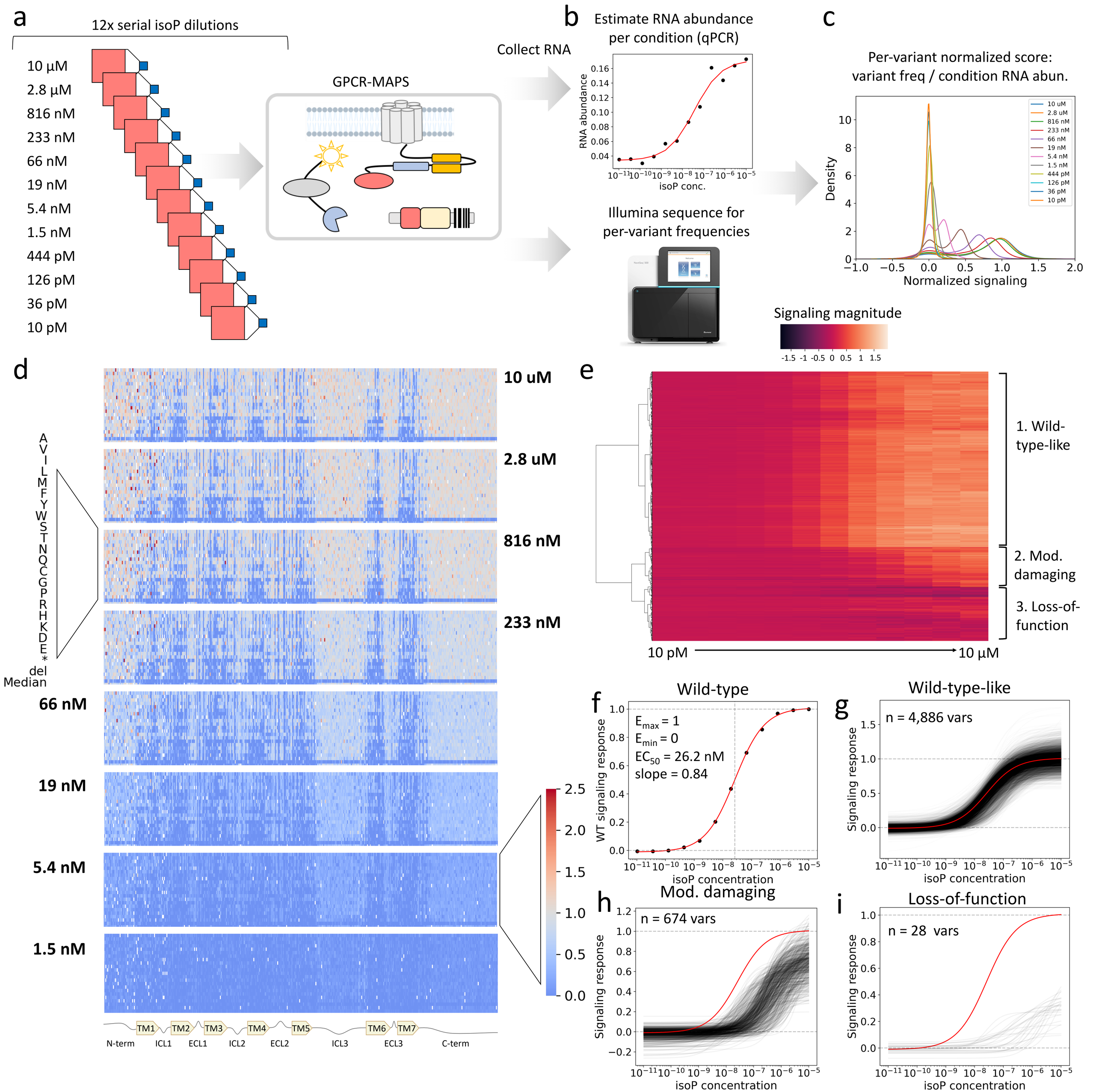
Active (3SN6)

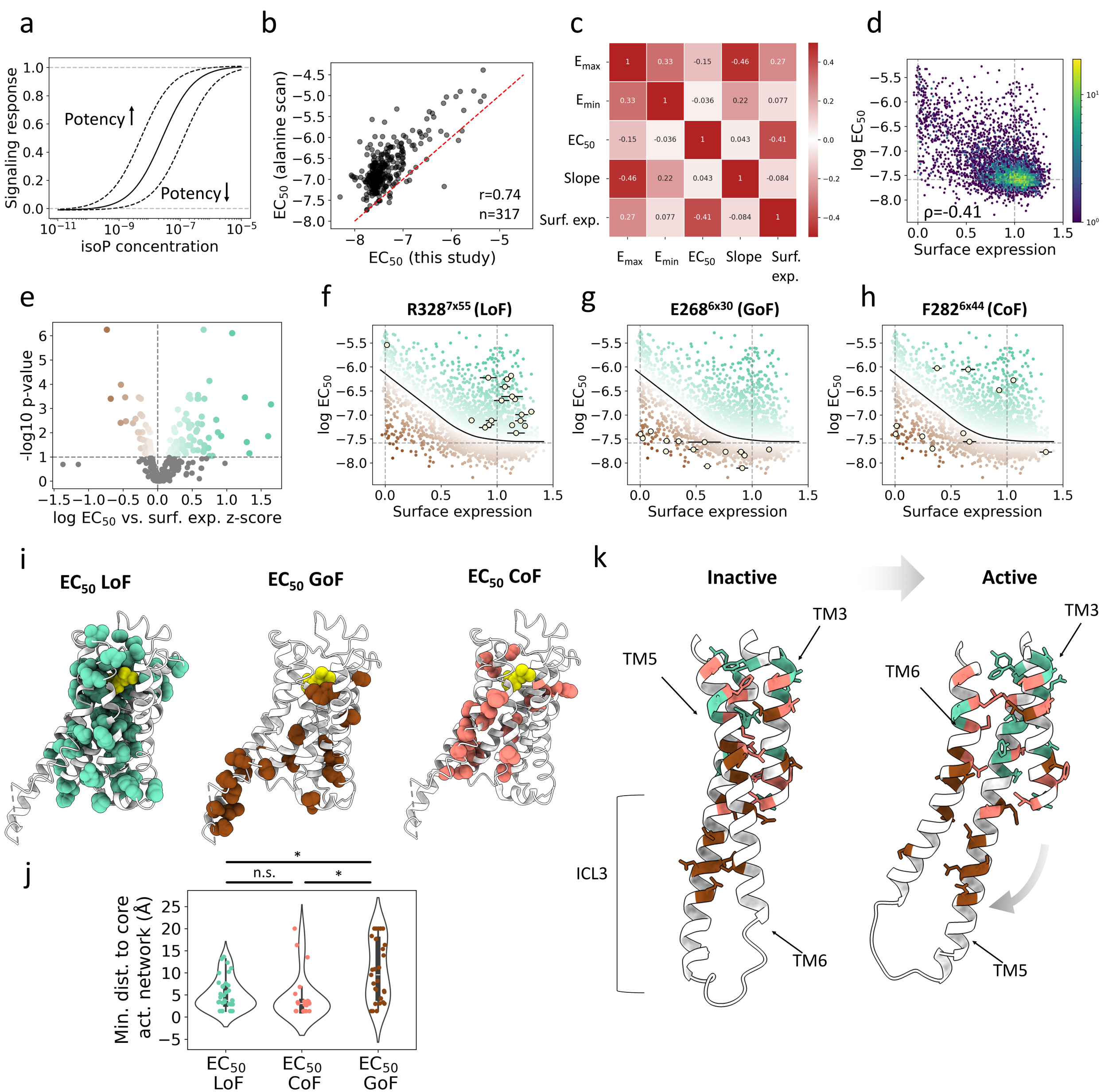


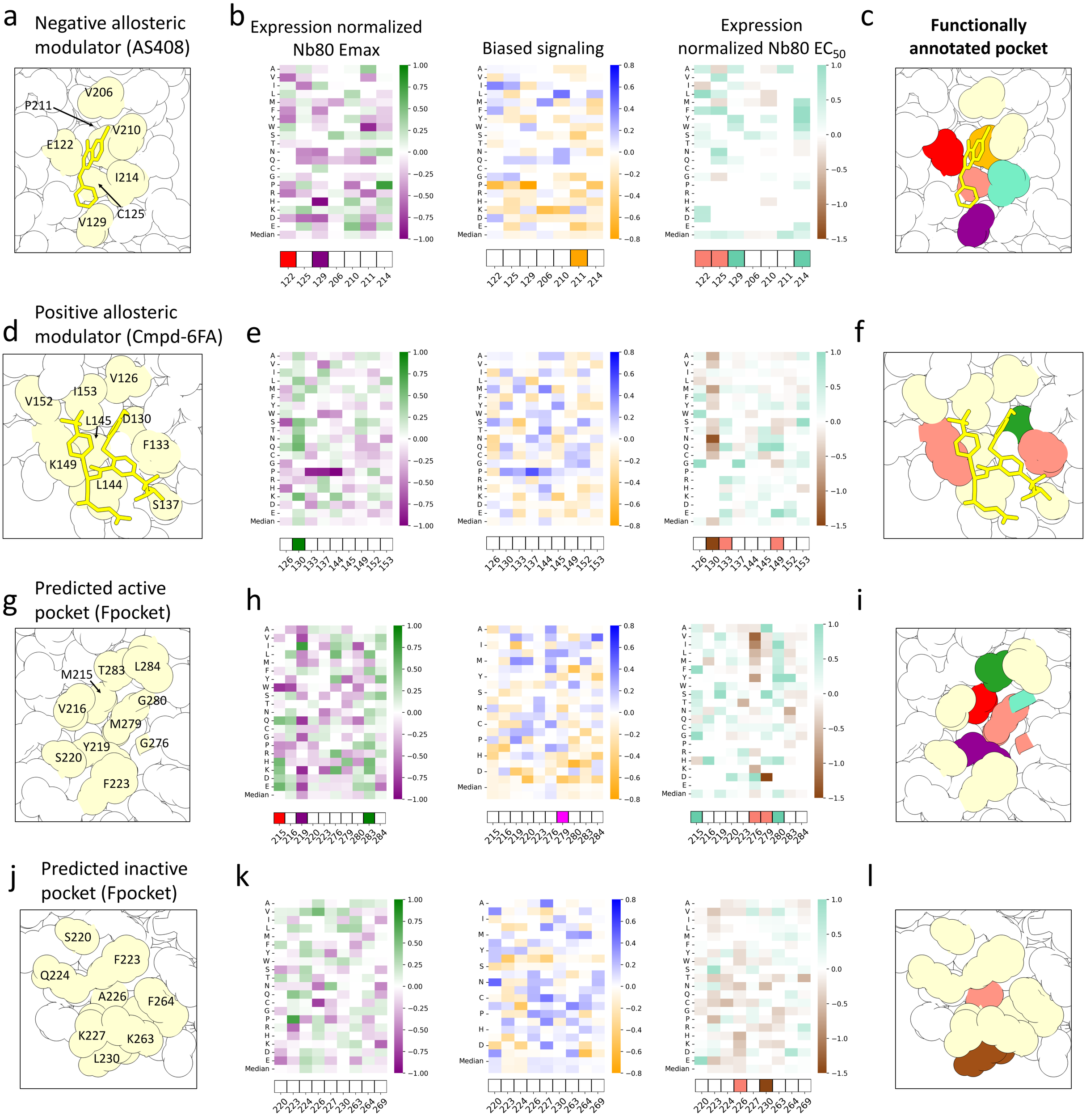
b

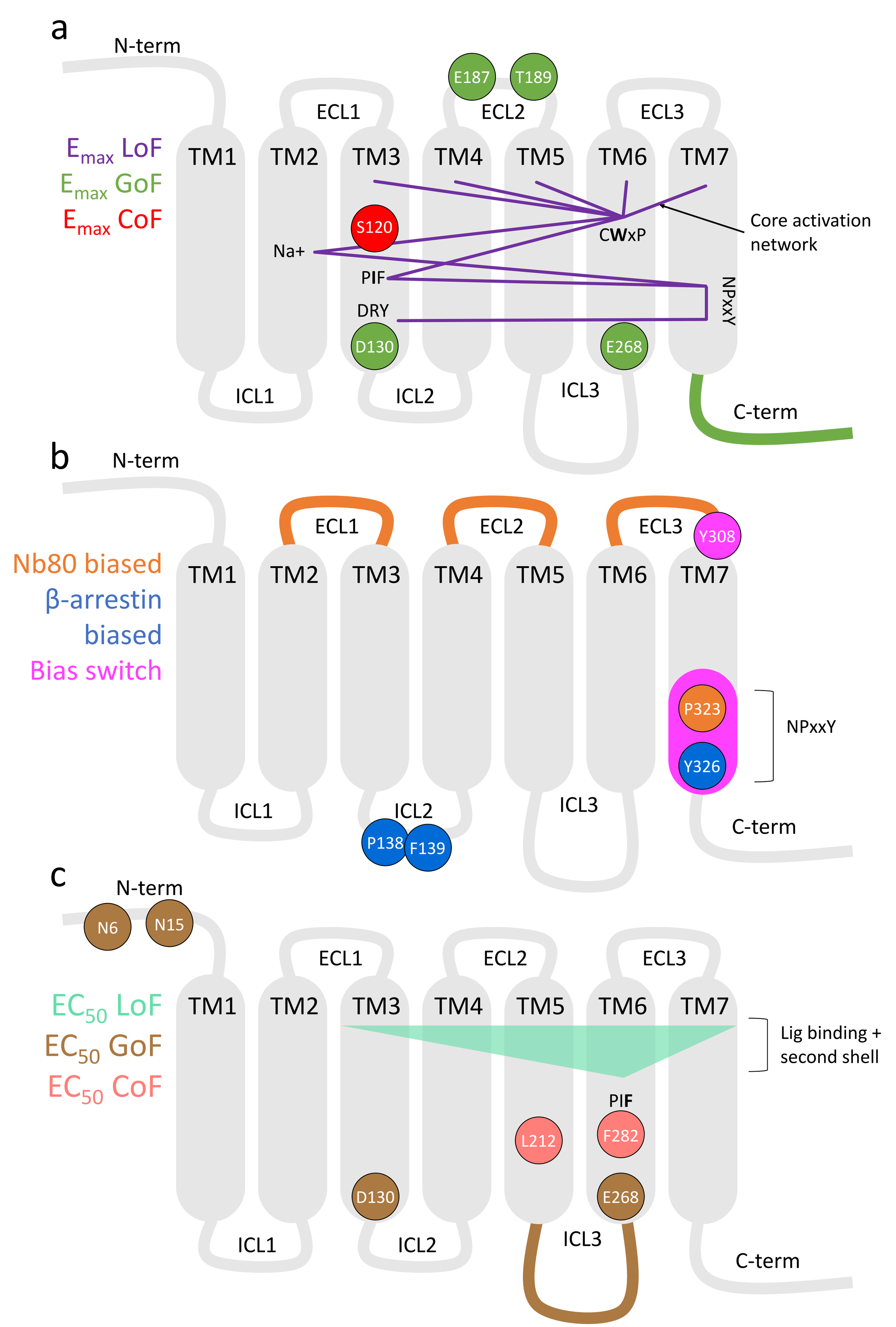



a
Isoprenaline
β2AR
βArr
Nb80
b
r=0.89
βArr signaling
Nb80 signaling
c
F193A
Y219A
Y199A
βArr signaling
Nb80 signaling
d
-log10 p-value
Nb80 vs. βArr z-score
e
P138^34x50 (βArr biased)
βArr signaling
Nb80 signaling
f
W99^23x50 (Nb80 biased)
βArr signaling
Nb80 signaling
g
F332^8x50 (bias switch)
βArr signaling
Nb80 signaling
i
F193^45x52
isoprenaline
Y308^7x34
j
Y308^7x34 (bias switch)
Nb80 vs. βArr residual
Mutant aa hydrophobicity
ρ = 0.67
h
β-arrestin biased
Nb80 biased
Bias switches
k
Extracellular vestibule
W173^ECL2
A181^ECL2
I169^4x61
N196^5x36
Y199^5x39
I298^6x60
Y308^7x34
E306^7x32
W313^7x39
Y185^ECL2
C184^ECL2
T189^ECL2
F101^23x52
W99^23x50
S111^3x30
M40^1x39
l
NPxxY and TM7
C285^6x47
N318^7x45
M279^6x41
Y326^7x53
R328^7x55
P330^8x48
A46^1x45
N51^1x51
P323^7x50
F332^8x50


a
12x serial isoP dilutions
10 µM
2.8 µM
816 nM
233 nM
66 nM
19 nM
5.4 nM
1.5 nM
444 pM
126 pM
36 pM
10 pM
GPCR-MAPS
Collect RNA
b
Estimate RNA abundance per condition (qPCR)
RNA abundance
isoP conc.
Illumina sequence for per-variant frequencies
c
Per-variant normalized score: variant freq / condition RNA abun.
Density
Normalized signaling
Signaling magnitude
d
10 uM
2.8 uM
816 nM
233 nM
66 nM
19 nM
5.4 nM
1.5 nM
A V I L M F Y W S T N Q C G P R H K D E * del Median
TM1 TM2 TM3 TM4 TM5 TM6 TM7
N-term ICL1 ECL1 ICL2 ECL2 ICL3 ECL3 C-term
e
1. Wild-type-like
2. Mod. damaging
3. Loss-of-function
10 pM
10 µM
f
Wild-type
E_{max} = 1
E_{min} = 0
EC_{50} = 26.2 nM
slope = 0.84
WT signaling response
isoP concentration
g
Wild-type-like
n = 4,886 vars
Signaling response
isoP concentration
h
Mod. damaging
n = 674 vars
Signaling response
isoP concentration
i
Loss-of-function
n = 28 vars
Signaling response
isoP concentration


a
Signaling response
Potency
Potency
isoP concentration
b
EC50 (alanine scan)
EC50 (this study)
r=0.74
n=317
c
Emax
Emin
EC50
Slope
Surf. exp.
d
log EC50
Surface expression
ρ=-0.41
e
-log10 p-value
log EC50 vs. surf. exp. z-score
f
R328 7x55 (LoF)
g
E268 6x30 (GoF)
h
F282 6x44 (CoF)
i
EC50 LoF
EC50 GoF
EC50 CoF
j
Min. dist. to core act. network (Å)
n.s.
*
*
EC50 LoF
EC50 CoF
EC50 GoF
k
Inactive
Active
TM3
TM5
TM6
ICL3
TM6
TM3
TM5


a
Negative allosteric modulator (AS408)
P211
V206
V210
E122
I214
C125
V129
b
Expression normalized Nb80 Emax
Biased signaling
Expression normalized Nb80 EC50
c
Functionally annotated pocket
d
Positive allosteric modulator (Cmpd-6FA)
V126
I153
V152
L145
D130
F133
K149
L144
S137
e
f
g
Predicted active pocket (Fpocket)
M215
T283
L284
V216
G280
M279
Y219
G276
S220
F223
h
i
j
Predicted inactive pocket (Fpocket)
S220
F223
Q224
A226
F264
K227
K263
L230
k
l


a
N-term
E187
T189
ECL1
ECL2
ECL3
Emax LoF
Emax GoF
Emax CoF
TM1
TM2
TM3
TM4
TM5
TM6
TM7
S120
Na+
CWxP
Core activation network
PIF
DRY
NPxxY
D130
E268
ICL1
ICL2
ICL3
C-term
b
N-term
ECL1
ECL2
ECL3
Y308
Nb80 biased
β-arrestin biased
Bias switch
TM1
TM2
TM3
TM4
TM5
TM6
TM7
P323
Y326
NPxxY
ICL1
ICL2
ICL3
P138
F139
C-term
c
N-term
N6
N15
ECL1
ECL2
ECL3
EC50 LoF
EC50 GoF
EC50 CoF
TM1
TM2
TM3
TM4
TM5
TM6
TM7
Lig binding + second shell
PIF
L212
F282
D130
E268
ICL1
ICL2
ICL3
C-term