GPCR-MAPS: high-resolution functional and allosteric mapping of G protein-coupled receptor activation and bias

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 largest family of human receptors and drug targets, but how the signaling properties of receptors are encoded in their sequence is incompletely understood. Most GPCR drugs bind a conserved orthosteric pocket, which can result in non-specificity, toxicity and clinical failure. Precision modalities including 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 annotation of GPCR structures. Here we present a fast and general approach to rapidly build complete, high-resolution, multi-modal functional and allosteric maps of receptors. The approach, GPCR-MAPS, quantifies direct recruitment of proteins to a receptor and deconvolves the substantial effects of mutations on receptor expression. Applying GPCR-MAPS to the β2 adrenergic receptor generates >150,000 phenotypic measurements, including the full activation functions for all possible amino acid substitutions (>7,500 unique variants) in a single experiment. The multi-modal maps provide numerous mechanistic insights and reveal a modular receptor architecture, with a core activation network surrounded by residues controlling quantitative parameters and bias. The maps also identify multiple allosteric surface pockets, including pockets bound by serendipitously discovered allosteric ligands and a novel pocket with no known ligands. The application of this approach across the superfamily of GPCRs will provide comparative maps of receptor mechanisms and a functional framework for the 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, including in the 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 current 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. This non-specificity can result in 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

96 GPCR signaling through one[18] or two[19,20] pathways, revealing a rich diversity of
97 signaling responses to individual mutations. Pooled approaches have been
98 used to characterize surface expression[21–25], ligand binding[23,26], and single or
99 few dose transcriptional responses downstream of GPCRs[27–30]. However, the
100 limited throughput[18–20] and lack of quantification of signaling parameters limit
101 the insights that can be obtained by these approaches[21–30].

102 Many of the most basic questions in GPCR signaling remain unanswered.
103 Challenges include describing and understanding how the sequence of a
104 receptor determines signaling properties and the causal mapping from sequence
105 to biophysics to signaling characteristics. Properties to understand include
106 cellular abundance, maximum response, basal signaling, and potency of
107 response of a receptor to ligand binding. Mutating receptor residues to alanine
108 has provided important insights into this, but testing only 1 of 19 alternative amino
109 acids in each position provides a limited view of the functional consequences of
110 sequence change. An additional challenge is to understand how the sequence
111 of a receptor controls bias in the recruitment of different transducer proteins[31].

112 With respect to drug discovery, a key challenge is to understand how
113 ligands modulate receptor signaling. Are all ligand-contacting sites equally
114 important for a response or just a subset of "hotspot"[32] residues? Are all ligand
115 contacts equally important for different functional responses to a ligand or are
116 some contacts more important for particular responses? Where else in the
117 receptor mediates the response to a ligand? Are residues important for the
118 response to ligand equally important for different functional responses, including
119 signaling bias?

120 A major goal in GPCR drug discovery is to identify allosteric sites on a
121 receptor where perturbations can inhibit, activate, or modulate signaling.
122 GPCRs, like all proteins, have many potential small pockets to therapeutically
123 target[33]. For each receptor it would be extremely useful to know at which of
124 these sites perturbations alter receptor activity and how, including quantifying
125 changes in agonist responses and altered bias.

126 Here we present a general high-throughput method, GPCR-MAPS
127 (GPCR multiplexed assays with pathway specificity), to rapidly generate
128 comprehensive functional maps of GPCRs. As a first application, we apply
129 GPCR-MAPS to the beta-2 adrenergic receptor (β2AR), one of the most
130 extensively characterized receptors. β2AR is expressed in the cardiovascular
131 system where it modulates smooth muscle relaxation and bronchodilation;
132 β2AR agonists are therefore used for treating asthma as well as chronic
133 obstructive pulmonary disease[34].

134 GPCR-MAPS has four components: First, we generate a complete site-
135 saturation library of a receptor, consisting of all possible amino acid substitutions
136 in all positions. Second, we quantify the cell surface expression of the library of
137 variants to quantify defects in fold stability and trafficking. Third, we use pooled
138 experiments to quantify the recruitment of conformation-specific interaction
139 partners to every receptor mutant in response to ligand treatment. Using ligand
140 titrations allows the complete dose-response activation functions of thousands of
141 receptor variants to be quantified. Measuring the recruitment of different
142 interaction partners quantifies changes in receptor bias. Fourth, we integrate the
143 resulting data to generate rich multi-modal functional maps of how the

144 quantitative function of a receptor is altered by diverse perturbations throughout
145 its entire sequence.
146 Applied to β2AR, GPCR-MAPS quantifies the activation functions and
147 multi-modal characteristics of >7,500 different protein variants in pooled
148 experiments. The resulting multi-modal maps reveal a modular receptor
149 architecture, with a core activation network surrounded by residues controlling
150 quantitative parameters and bias. The multi-modal maps also identify multiple
151 allosteric surface pockets, including pockets bound by serendipitously
152 discovered β2AR allosteric ligands and a novel pocket with no known ligands.
153 We believe application of the GPCR-MAPS approach across the GPCR
154 superfamily will advance our understanding of the largest family of human
155 membrane receptors and enable rational, precision drug design.
156
157 **Results**
158 **GPCR-MAPS: multiplexed assay with pathway specificity**
159 To allow quantification of multi-modal receptor phenotypes for thousands
160 of receptor variants in parallel, including full dose-response curves, we
161 developed a modular platform that we call GPCR-MAPS. Our goal is to read out
162 receptor states as directly as possible. In contrast to other approaches[27–30], we
163 therefore focused on quantifying direct binding interactions with receptors.
164 Quantifying direct binding has proven powerful for understanding signaling by
165 soluble proteins[35] and GPCRs[36] and avoids the confounding influence of non-
166 linearities, feedback, and differential signal amplification when quantifying
167 activity downstream of endogenous signaling networks.

168 GPCR-MAPS is compatible with cassette-based[37,38] or SUNi
169 mutagenesis[39] to generate a single pooled library of receptor variants
170 containing all possible single amino acid substitutions. The library is barcoded
171 and rapidly integrated into a single locus in engineered HEK293 “landing pad”
172 cells[40] so that each cell expresses a single copy of a single variant.
173 In a first selection experiment, the cell surface expression of each variant
174 is measured using FACS followed by sequencing[21,22,25] to precisely quantify
175 changes in receptor fold stability and trafficking (Figure 1a, Materials and
176 Methods).
177 Additional selections then quantify direct binding of receptor interaction
178 partners. To quantify recruitment of receptor interaction partners in response to
179 ligand treatment we developed a modular protein interaction assay based on a
180 previously described platform[41] but extensively re-engineered to enable
181 massively multiplexed experiments (see Materials and Methods). We also
182 improved the signal to noise of the platform and extended it to allow modular read
183 out of the recruitment of diverse GPCR protein interaction partners including the
184 active-state-specific nanobody Nb80[42], β-arrestin and miniG_s[43] (Supplementary
185 Figure 1b).
186 The assay uses three genetic constructs to report on GPCR activation: 1)
187 a GPCR of interest with a fusion of the exogenous transcription factor Gal4,
188 linked by sequence encoding an evolved light, oxygen, and voltage sensitive
189 (eLOV) domain as well as tobacco etch virus protease cleavage site (TEVcs), 2)
190 a GPCR binding partner (e.g. β-arrestin) fused to NanoLuciferase (NLuc) and
191 tobacco etch virus protease (TEV), and 3) the fluorescent protein Citrine under
192 the control of the UAS promoter (Figure 1b and Supplementary Figure 1e).

193 When co-expressed in cells, agonist-mediated activation of the GPCR causes
194 recruitment of the binding partner to the receptor. In the absence of light, the
195 eLOV domain cages the TEVcs and prevents cleavage; however, co-addition of
196 the NLuc substrate furimazine results in bioluminescence, and when the
197 transducer brings NLuc into close proximity with the eLOV domain, the steric
198 hindrance is relieved and TEVp can cleave at TEVcs. This cleavage results in
199 liberation of Gal4, which can move to the nucleus and drive transcription of the
200 UAS-Citrine reporter construct, which is read out by short-read RNA sequencing
201 (Figure 1b).

202 The recruitment assay allows the binding of multiple GCPR protein
203 interaction partners to be quantified in a single condition (for example ligand
204 concentration). However, because of its highly multiplexed RNA-seq based
205 readout, it also allows the quantification of protein recruitment across many
206 different conditions, for example ligand combinations, changes in culture
207 conditions or, as we demonstrate here, ligand titrations allowing the construction
208 of full dose-response curves for thousands of different receptors in a single
209 pooled experiment.

210 **Application to the beta-2 adrenergic receptor (β2AR)**

211 We generated a saturation mutagenesis library of β2AR in a vector also
212 containing: an N-terminal HA epitope tag and C-terminal fusions of eLOV,
213 TEVcs, and Gal4, and additionally the UAS-Citrine module (attB-HA-b2AR-
214 eLOV-TEVcs-Gal4-UAS-Citrine, Supplementary Figure 1e) including all
215 missense and nonsense substitutions, synonymous mutations where possible
216 (i.e. the wild-type amino acid can be encoded by alternate codons), and single
217 residue deletions. We then introduced short degenerate nucleotide barcodes

218 into the 3' UTR of Citrine and used long read sequencing to make 212,840
219 β2AR genotype-barcode links (Supplementary Figure 1c, Materials and
220 Methods). These represented 8,970 out of an intended 9,046 unique genotypes
221 (99.2%). We introduced this library into HEK293T landing pad cells[40], ensuring
222 a single variant is expressed per cell.

223 **β2AR surface expression**

224 The first selection experiment of GPCR-MAPS is a FACS-based
225 multiplexed surface expression experiment. Cells were immunostained with
226 anti-HA antibody, then sorted into four bins based on fluorescence intensity
227 (Supplementary Figure 1d, Materials and Methods). DNA from the cells sorted
228 into each bin was purified and converted into sequencing libraries, and the
229 barcodes were sequenced with short-read sequencing (Figure 1a, Materials
230 and Methods).

231 All counts from barcodes representing the same variant were tallied and
232 combined with the average fluorescence value of each FACS bin to arrive at
233 surface expression estimates. Variants with low frequency in the library were
234 filtered to yield 8,740 high confidence surface expression measurements that
235 were highly reproducible across three replicates (Pearson r=0.94-0.96, Figure
236 1c, Materials and Methods, Supplementary Table 1). Scores were normalized
237 between the median value of premature stop codons, set to 0, and the wild-type
238 surface expression value, set to 1. A heatmap representation of the data shows
239 nonsense mutations are uniformly damaging, except those occurring within
240 intracellular loop 1 (ICL1) and those in the C-terminus, suggesting the receptor
241 can be stably expressed as either a single-pass transmembrane protein, or as a
242 C-terminal tail truncation. The transmembrane helices are more mutation

243 sensitive than extra- or intra-cellular loops, consistent with previous work[21–25,28]
244 (Figure 1d). Synonymous wild-type and nonsense variants have well separated
245 expression distributions, with missense variants spanning the full range
246 between those two classes (Figure 1e). Our surface expression estimates are in
247 line with previously estimated surface expression levels of alanine
248 substitutions[18] (Spearman's ρ = 0.43, Supplementary Figure 1i).

249 **β-arrestin and Nb80 recruitment**

250 The second selection experiments of GPCR-MAPS are protein
251 interaction recruitment assays. For β2AR we quantified recruitment of two
252 conformation-specific binders: Nb80 (as a surrogate for G-protein, as Nb80
253 returns higher signal to noise than miniG_s (Supplementary Figure 1b)) as well
254 as β-arrestin. Namely, we transiently transfected one of the binder constructs
255 (pUC19-NLuc-[Nb80/βArr]-uTEV1Δ, Supplementary Figure 1e) into cells
256 bearing the β2AR receptor construct in the landing pad. We also included a
257 condition where Gal4 was transfected alone (pUC19-CMV-Gal4, Supplementary
258 Figure 1e); barcode frequencies in this control condition should be related to
259 relative abundance in the library combined with any barcode-specific effects on
260 RNA stability. The next day, saturating concentration of isoprenaline (10 μM)
261 was added to the experimental samples to induce activation, and 24 hours later,
262 those cells were collected. The Gal4 maximum induction samples were
263 collected 30 hours after transfection. RNA was collected from all samples,
264 reverse transcribed, converted to short-read sequencing libraries and
265 sequenced (Supplementary Figure 1f-g, Materials and Methods).

266 Recruitment scores were calculated using the variant frequency in the
267 experimental samples compared to variant frequency in the Gal4 condition and

268 re-scaled with 0 representing the median of nonsense variants recruitment and
269 1 representing wild-type recruitment. Low frequency variants were filtered to
270 yield 7,850 high confidence Nb80 and β-arrestin recruitment scores (Materials
271 and Methods, Supplementary Table 1). Replicates were well correlated
272 (Pearson's r=0.9-0.91 for Nb80 replicates, r=0.88-0.89 for β-arrestin replicates,
273 Figure 1c). Heatmap representations of the data reveal similar overall trends
274 between Nb80 and β-arrestin (as well as with the surface expression data)
275 defined by transmembrane regions being most sensitive to mutation (Figure 1f-
276 h). Synonymous wild-type variants are again well separated from nonsense
277 variants, and missense variants again span the whole distribution from
278 nonsense to wild-type-like (Figure 1g, i). The Nb80 and β-arrestin recruitment
279 scores are also correlated with each other (Spearman's ρ = 0.87), as well as
280 with previous multiplexed measurements of β2AR variants signaling through the
281 G_s pathway[27] (Spearman's ρ = 0.44 and 0.47 for Nb80 and β-arrestin
282 recruitment, respectively, Supplementary Figure 1j-k).

283 **Deconvolving recruitment and expression**

284 Plotting Nb80 and β-arrestin recruitment scores against receptor
285 expression reveals that changes in expression are an important cause of
286 changes in recruitment (ρ=0.55, 0.6, for Nb80 and β-arrestin respectively, Figure
287 2a-b, Supplementary Figure 1h). However, some mutations have much larger
288 effects on recruitment than can be explained by changes in expression. For
289 example, missense mutations at the nine positions making up the isoprenaline
290 binding site[44] have minimal effect on expression (Figure 1j), while these same
291 mutations are strongly deleterious for recruitment of both binding partners (Figure
292 1k-l).

293 **Residues contributing to recruitment beyond expression changes**

294 The relationship between recruitment and expression is non-linear (Figure
295 2a-b). We therefore fit a LOWESS (locally weighted scatterplot smoothing)
296 regression to the data, which acts as a null expectation of recruitment produced
297 by a given surface expression level. This null expectation explains a significant
298 amount of variance (R^2 = 0.37 and 0.42 for Nb80 and β-arrestin, respectively).

299 We next quantified changes in recruitment not explained by expression
300 changes by calculating recruitment residuals and subsequently z-scores of all
301 variants to the global model (i.e. residual divided by standard error of the mean
302 for each variant, Figure 2c-e). Comparing the group of z-scores at each position
303 with the z-scores from all other positions with a two-sided Mann-Whitney U test
304 allows us to identify residues where mutations have greater or lesser recruitment
305 than would be expected based on their surface expression level (Supplementary
306 Table 2).

307 **Complete functional maps of Nb80 and β-arrestin recruitment**

308 For both Nb80 and β-arrestin, more residues confer loss-of-function
309 (LoF) than gain-of-function (GoF) when mutated, and the magnitude of loss is
310 greater than the magnitude of gain (Figure 2f-g). In total, 49 and 73 residues
311 are LoF for Nb80 and β-arrestin, respectively (false discovery rate, FDR = 0.1,
312 Supplementary Table 2). Of these, 47 are shared (Figure 2h, OR=304.6,
313 $p=2.2x10^{-40}$, two-sided Fisher exact test). An example of a strong LoF residue is
314 $D113^{3x32}$ (Figure 2i, GPCRdb numbering[45] in superscript), which participates
315 directly in isoprenaline binding. There are 31 and 53 GoF residues for Nb80 and
316 β-arrestin (FDR=0.1), respectively, of which 20 are shared (OR=19.2, $p=1.3x10^{-}$
317 ^{12}, two-sided Fisher exact test). $E268^{6x30}$ is the most significant GoF (Figure 2j),

318 consistent with previous work demonstrating constitutively activating mutations
319 at this site[46]. Finally, we extended this analysis to highlight residues where
320 mutations have highly variable effects on recruitment. Namely, after excluding
321 the LoF and GoF residues, we compared the interquartile range (IQR) of z-
322 scores at all positions (Supplementary Figure 2a-b). We categorized those
323 residues with IQR above the 90th percentile as change-of-function (CoF): 34
324 and 31 residues for Nb80 and β-arrestin, respectively, with 8 shared (Figure 2h,
325 OR=2.9, p=0.037). An example of a strong CoF residue is L75^{2x46} (Figure 2k),
326 which is near the allosteric sodium binding site and neighbors LoF residue
327 A76^{2x47}. At this position, mutations to small, uncharged residues (G, S, T, A, C)
328 decrease surface expression, but increase Nb80 and β-arrestin recruitment
329 relative to the modeled expectation. On the other hand, positively charged
330 residues (H, R, K) are well expressed but strongly decrease recruitment of both
331 proteins.

332 Average position z-scores and position IQRs are well correlated (r = 0.73
333 and 0.56, respectively, Supplementary Figure 2c-d), indicating that many
334 residues contribute similarly to recruitment of both proteins. Plotting the
335 residues identified for both proteins onto the structure of isoprenaline-stimulated
336 β2AR (PDB ID: 7DHR) reveals that LoF residues are predominantly within the
337 core of the receptor (Figure 2l), whereas GoF and CoF residues are dispersed
338 on the periphery (Figure 2m-n).

339 **Loss-of-function residues define a core activation network**

340 The majority of the shared LoF residues (39 out of 47) form an
341 interconnected path linking the ligand binding site all the way to the transducer
342 binding site in the active state receptor structure (Figure 3a, hereafter referred

343 to as the “core activation network”). All 9 residues of the isoprenaline binding
344 pocket are identified (D113^{3x32}, V117^{3x36}, F193^{45x52}, S203^{5x43}, S204^{5x44},
345 S207^{5x461}, F290^{6x52}, N293^{6x55}, and N312^{7x38}) along with many residues in the
346 second shell of the ligand binding pocket (i.e. residues that contact residues of
347 the isoprenaline binding site, Figure 3b). These residues include W286^{6x48} of
348 the CWxP motif and I121^{3x40} of the PIF motif[47,48]. The core activation network
349 then connects to N322^{7x49} of the critical NPxxY motif[47] via two pathways: first,
350 W286^{6x48} contacts N318^{7x45} which then connects to N322^{7x49}. Second, a
351 pathway from A119^{3x38} goes through A78^{2x49} and D79^{2x50} of the allosteric
352 sodium binding pocket which then contacts N322^{7x49}. Of the NPxxY motif,
353 N322^{7x49}, I325^{7x52}, Y326^{7x53} are identified as LoF. Finally, in the active state
354 structure, Y326^{7x53} of the NPxxY motif then contacts R131^{3x50} (of the DRY motif)
355 and I127^{3x46} of helix 3. Y326^{7x53} is also thought to make an important water-
356 mediated hydrogen bond with Y219$^{5.58}$,[49] which is also identified as LoF and
357 makes a direct contact with R131^{3x50}.

358 Of the eight LoF residues not connected to the core activation network,
359 three likely directly disrupt transducer binding[42,50] (P138^{34x50}, F139^{34x51}, and
360 Y141^{34x53}), while three more take part in critical disulfide bonds between ECL2
361 and TM3 (C106^{3x25}, C190^{ECL2}, C191^{45x50}). F193^{45x52} is a disconnected LoF
362 residue that makes direct contact with bound isoprenaline. Finally, P88^{2x58} in
363 TM2 is the last disconnected residue.

364 Previous work has used structural analysis to define an activation
365 network for class A GPCRs[51] and alanine scanning mutagenesis to define an
366 activation network for β2AR[18]. About half of the residues identified in our
367 network are present in at least one of the others (19/39, Figure 3c). We identify

368 more residues near the ligand binding site compared to the other networks
369 (Supplementary Figure 2e-g). Interestingly, of the six residues shared by the
370 other two networks that are not part of our core activation network, we identify
371 four of these as either CoF (S120^{3x39} and F332^{8x50}) or GoF (L124^{3x43} and
372 D130^{3x43}), emphasizing the value of empirically assaying all 19 substitutions per
373 position. At S120^{3x39}, which neighbors I121^{3x40} of the PIF motif, positively
374 charged substitutions (R, K, H) yield stable receptors with significantly reduced
375 recruitment, while the E and Q also yield stable receptors but significantly
376 increase recruitment (Figure 3d). At L124^{3x43}, most variants decrease surface
377 expression but increase relative recruitment, with the exception of the aromatic
378 amino acids (W, F, and Y) and proline which recruit Nb80 less than expected
379 given their expression level (Figure 3e).

380 **Gain-of-function sites are located at the ionic lock, near the NpxxY motif,**
381 **and in the extracellular vestibule**

382 We next examined the location of the GoF residues (Supplementary
383 Figure 3a, f). Two of the GoF residues, D130^{3x49} and E268^{6x30}, participate in the
384 highly conserved ionic lock (with R131^{3x50} of the DRY motif[52]). While R131^{3x50} of
385 the DRY motif is a LoF residue that contacts Nb80[42] and is predicted to contact
386 β-arrestin,[50] some mutations at D130^{3x49} of the DRY motif have been shown to
387 constitutively activate the receptor[53]. The other GoF residues are predominantly
388 in two clusters: one consisting of TMs 3, 5, and 6, near the NPxxY and DRY
389 motifs (Supplementary Figure 3b). L124 sits in the center of the receptor, near
390 the I of PIF motif, DRY motif, and NPxxY. A few mutations at L124^{3x43} have
391 been shown to cause constitutive activity at this site[54] but we show that most
392 variants activate more than expected (discussed above). The other cluster is

393 near the ligand binding site and in the extracellular vestibule (Supplementary
394 Figure 3c). E187 and T189 are in ECL2, where they may affect ligand binding
395 or activation[55]. W313^{7x39} and V87^{2x57} form two van der Waals interactions in the
396 inactive (PDB:2RH1), but only one in the active state (PDB:3SN6,
397 GetContacts); our data suggest that disrupting this interaction leads to
398 increased recruitment. Finally, there are 7 GoF residues in the C-terminal tail
399 (Supplementary Figure 3g).

400 **Change-of-function sites are adjacent to the core network**

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

410 **Differential contribution of residues in the isoprenaline binding pocket**

411 Next, we examined the effects of substitutions at residues making up the
412 isoprenaline binding site. While all residues are LoF, the magnitude of effects
413 differs among the constituents. D113^{3x32}, which forms two hydrogen bonds with
414 isoprenaline, has strongly damaging effects (median expression-adjusted Nb80
415 and β-arrestin recruitment = -0.94 and -0.89, Figure 3f-g and Supplementary
416 Figure 3h-i). N293^{6x55} also forms a hydrogen bond with isoprenaline, but
417 substitutions here are the least damaging of the binding site (median

418 expression-adjusted Nb80 and β-arrestin recruitment = -0.24 and -0.4). This
419 tolerance is surprising as N293^{6x55} also forms a hydrogen bond with S204^{5x44},
420 which doesn't directly contact the ligand but has strongly deleterious effects
421 (median expression-adjusted Nb80 and β-arrestin recruitment = -0.96 and -
422 0.81). This suggests the role of S204^{5x44} may be more related to receptor
423 activation than ligand binding.

424 **Genetic control of receptor recruitment bias**

425 Next, we directly compared mutational effects on Nb80 and β-arrestin
426 recruitment (Figure 4a). Missense mutational effects on recruitment of Nb80
427 and β-arrestin are well correlated (Pearson's r=0.89, Figure 4b). However,
428 single mutants previously shown to bias the receptor toward G protein signaling
429 (F193A[56], Y219A[57], and Y199A[58]) are indeed biased towards Nb80 in our
430 experiments (Figure 4c). To find residues with stronger effects on one pathway
431 compared to the other, we again fitted a LOWESS regression curve to the data,
432 and calculated residuals and subsequently z-scores for all variants in the
433 dataset. Then, we compared z-scores of mutations at each position with z-
434 scores of mutations at all other positions. This identifies 19 residues where
435 mutations have significantly greater effects on β-arrestin recruitment and 7
436 positions where mutations have significantly greater effects on Nb80
437 recruitment (FDR=0.1, Figure 4d, Supplementary Table 3). As before, we then
438 compared the IQR of all remaining residues; here, to more closely match the
439 number of biased residues, we prioritized the top 5% of residues as recruitment
440 bias switches (Supplementary Figure 4a).

441 Examples of strongly biased residues are shown (Figure 4e-f) as well as
442 a strong bias switching residue (Figure 4g). One expectation is that mutations at

443 residues making direct contacts with either Nb80 or β-arrestin could result in
444 biased recruitment. Nb80 and β-arrestin have overlapping but distinct binding
445 interactions with the intracellular portion of β2AR. Nb80 forms direct contacts
446 with 16 residues in TMs 3, 5, 6, and 7[42]. Since there is no experimentally
447 determined structure of β2AR bound by β-arrestin, we defined this interface
448 from a simulation study[50]. The β-arrestin binding interface is predicted to be
449 larger, encompassing 28 residues across ICLs 1, 2, 3, and TMs 6 and 7, with 10
450 residues overlapping the Nb80 interface.

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

462 **The extracellular vestibule allosterically controls recruitment bias**

463 In contrast, there are many residues biased toward Nb80 recruitment
464 located away from the orthosteric site in the extracellular vestibule: W99^{23x50},
465 F101^{23x52}, G102^{3x21}, W173^{ECL2}, A181^{ECL2}, C184 ^{ECL2}, Y185 ^{ECL2}, N196^{5x36}, and
466 E306^{7x32} (Figure 4h, k). The most strongly Nb80 biased residue, W99^{23x50}, was
467 previously identified as particularly mutationally intolerant[27]; however only by

468 measuring recruitment of both proteins could we uncover its strongly biased
469 effects (Figure 4f). Considered together, the extracellular loops (ECLs) are
470 enriched in Nb80 biased residues compared to the whole receptor (OR=6.4,
471 $p=1.0x10^{-3}$, one-sided Fisher exact test), and in particular, ECL2 is enriched for
472 Nb80 biased residues (OR=5.6, p=0.014, one-sided Fisher exact test).
473 The importance of the extracellular vestibule for G-protein biased
474 activation revealed in our mutational data is also consistent with structural data.
475 Salmeterol is a G-protein biased β2AR agonist that binds the orthosteric site in
476 a similar way as the endogenous ligand (epinephrine) except that it has a long
477 aryloxyalkyl tail that extends into the extracellular vestibule, where it interacts
478 extensively with ECL2[59]. Also, comparative structural analyses of other GPCRs
479 have suggested that the extracellular vestibule adopts a more constricted
480 conformation in the β-arrestin-bound state compared to the G-protein-bound
481 state.[60] Taken together, these data point to the extracellular vestibule, and in
482 particular ECL2, as hotspots for G-protein-biased agonism at the β2AR.
483 Among the bias switching residues, the second strongest is the vestibule
484 residue, $Y308^{7x34}$ (Supplementary Figure 4d). This residue has been proposed
485 to be necessary for G-protein signaling[61], as well as forming a hydrophobic gate
486 with $F193^{45x52}$ that interacts with and dehydrates ligand molecules as they
487 traverse the vestibule and approach the orthosteric binding site[62]. In the active
488 state, these two residues form a “lid” over the ligand, with $F193^{45x52}$ making a
489 direct contact with isoprenaline (Figure 4i). Based on the alanine mutation,
490 $F193^{45x52}$ was postulated to coordinate β-arrestin signaling[56]. Interestingly, while
491 this single variant does have a stronger effect on β-arrestin recruitment, other
492 variants at the same site have the opposite effect (Supplementary Figure 4e).

493 We examined the effects of mutations at these two sites and found that
494 recruitment bias is correlated with mutant hydrophobicity for $Y308^{7x34}$ ($\rho = 0.67$,
495 $p = 4.3x10^{-3}$, Figure 4j), but not for $F193^{45x52}$ ($\rho = -0.36$, $p = 0.16$,
496 Supplementary Figure 4f). This suggests that the role of $Y308^{7x34}$ in determining
497 bias is directly tied to its hydrophobicity, while the relationship of $F193^{45x52}$ to
498 bias is more nuanced.

499 **Recruitment bias in TM7**

500 Besides the vestibule, there are multiple bias-functional residues (Nb80-
501 or β-arrestin-biased or bias switches) in or near the NPxxY motif of TM7(Figure
502 4l). In fact, TM7 is enriched for bias-functional residues compared to the whole
503 receptor (OR=3.7, p=0.011, one-sided Fisher exact test, Supplementary Figure
504 4g). $P323^{7x50}$ is not predicted to contact directly β-arrestin[50], suggesting that the
505 Nb80 bias of the residue cannot be explained simply by disrupting a direct
506 contact. $N318^{7x45}$, $F332^{8x50}$, and $R328^{7x55}$ are all bias switches in close proximity
507 to the NPxxY motif (Figure 4k). A role for TM7 in determining bias is consistent
508 with prior nuclear magnetic resonance experiments demonstrating changes in
509 TM7 conformation upon binding of biased agonists[63].

510 **Activation functions for >7,500 β2AR variants**

511 The data presented thus far quantifies recruitment to β2AR with
512 saturating levels of isoprenaline, which convolutes the contribution of different
513 pharmacological parameters, such as potency (the amount of agonist needed to
514 elicit a certain response) and efficacy (the maximum recruitment response
515 elicited by saturating concentrations of agonist). However, because of its pooled
516 design and RNA-seq based output, GPCR-MAPS can also be used to quantify

517 the full dose-response curves (activation functions) of all receptor variants in a
518 library.

519 To demonstrate this capability, we performed the Nb80 GPCR-MAPS
520 recruitment assay with 12 serially diluted concentrations of isoprenaline, ranging
521 from 10 pM to 10 uM (Figure 5a). In order to compare the magnitude of
522 recruitment across experimental conditions, following RNA isolation, we
523 performed quantitative PCR (qPCR) to estimate the total amount of reporter
524 transcription (Figure 5b, Supplementary Figure 5a, Materials and Methods).
525 Then, by deep sequencing the barcodes and normalizing the frequencies by the
526 estimated total reporter RNA, we arrived at normalized scores across the 12
527 concentrations (Figure 5c and Supplementary Figure 5b, Materials and
528 Methods). This allowed us to quantify dose-response relationships for 7,680
529 different receptor variants.

530 Heatmap and correlation representations of the data reveals induction of
531 Nb80 recruitment across the ~92,000 measurements (Figure 5d and
532 Supplementary Figure 5c-d). Clustering the induction dataset revealed three
533 predominant clusters- these can be assigned as wild-type-like, moderately
534 damaging, and strong loss-of-function (Figure 5e).

535 Next, we fit a four-parameter Hill curve to the wild-type genotype,
536 normalizing E_{min} and E_{max} to 0 and 1, respectively, and deriving a half maximal
537 response (EC_{50}) estimate of 26.2 nM and slope (Hill coefficient) of 0.84 (Figure
538 5f). Then we fit curves to all variants; after quality control (Supplementary
539 Figure 5e-f, Materials and Methods), this resulted in 5,588 high quality dose-
540 response curves with well-estimated parameters (Figure 5g-i). We note that
541 variants with reduced activation level are more poorly fit and therefore likelier to

542 be filtered (Supplementary Figure 5e). Our EC_{50} measurements correlate well
543 with the results of individually assayed substitutions[18] (Pearson r=0.74, n=317
544 variants, Figure 6b).

545 **Surface expression is a major determinant of receptor potency**

546 Changes in potency are often measured by comparing EC_{50} values
547 (Figure 6a). To investigate determinants of potency, we first quantified the
548 correlations between the four Hill function parameters, as well as their
549 dependence upon surface expression. Expression level has been predicted to
550 be an important influence on signaling parameters by theoretical work[64–66], and
551 by experimental work on a plant hormone receptor[35]. The strongest
552 relationships we found were between slope and E_{max} (ρ = -0.46) and between
553 surface expression and log EC_{50} (ρ = -0.41, Figure 6c-d). Surface expression
554 level is thus a major influence on potency in the β2AR. To mechanistically
555 understand changes in potency it is therefore necessary to again quantitatively
556 deconvolve direct mutational effects from those mediated by changes in
557 receptor expression.

558 **Potency tuning**

559 To isolate mutational effects on EC_{50} beyond those caused by changes
560 in expression we again used LOWESS regression to model the relationship
561 between surface expression and EC_{50} (Supplementary Figure 6a). We
562 quantified residuals to that model to identify positions where mutations have
563 consistently higher or lower EC_{50} than expected from their impact on
564 expression. 29 residues were identified as GoF (i.e. they had EC_{50} lower than
565 expected) and 59 were identified as LoF (i.e. higher EC_{50} than expected,
566 FDR=0.1, Figure 6e-f). There is overlap between these sets and the sets

567 identified from the saturating concentration experiment (23/59 LoF shared, 4/29
568 GoF shared). Finally, the interquartile range was used again to identify 29 EC_{50}
569 CoF residues (one shared with saturating concentration experiment,
570 Supplementary Figure 6b, Supplementary Table 4).

571 In principle, mutations could affect potency beyond changes in
572 expression by changing the affinity of isoprenaline binding, or by affecting the
573 probability of isoprenaline-bound receptor to adopt the active conformation[67].
574 While mutations at residues making direct contact with isoprenaline would be
575 expected to reduce affinity, mutations at these sites mostly totally abolish
576 activation; dose-response curves could only be fit to 32 out of 132 missense
577 variants at these sites. However, these 32 variants do have over an order of
578 magnitude reduced potency (median EC_{50}=477 nM, compared to wild-type 26.2
579 nM).

580 Interestingly, EC_{50} LoF residues are enriched in the second shell (15/59
581 LoF sites (25.4%, OR=4.12, $p=1.9x10^{-4}$, Fisher exact test). While not
582 statistically enriched, a higher fraction of EC_{50} CoF sites (5/24, 20.8%) are in the
583 second shell compared to only 1 out of 29 GoF residues (3.4%) in the second
584 shell (Figure 6i). This shows that whereas mutations in residues directly
585 contacting the ligand have very strong effects on affinity and so completely
586 inactivate recruitment, mutations in second shell residues frequently can have
587 more moderate effects and modulate the response. These effects are, by
588 definition, allosteric, either indirectly altering ligand binding affinity or altering the
589 consequences of ligand engagement. This also suggests that mutations in the
590 second shell are likelier to reduce affinity than increase it.

591 Beyond changes in ligand affinity mutations can also affect recruitment
592 by altering the likelihood of ligand-bound receptor adopting the active
593 conformation[67]. One way to disrupt the activation pathway could be for
594 mutations to interfere with the core activation network identified in the maximum
595 recruitment experiment. We therefore quantified the distance of EC_{50}-affecting
596 residues to this network and found that LoF and CoF residues are closer in 3D
597 space than GoF residues ($p=7.4x10^{-4}$, $1.3x10^{-3}$, respectively, two-sided Mann
598 Whitney U test, Figure 6j). Taken together, as with the maximum recruitment
599 analysis, these data suggest that mutations that reduce potency do so by
600 perturbing the core activation pathway whereas the mechanism-of-action of
601 GoF mutations is likely distinct.
602 Three of the four most significant EC_{50} GoF residues are $E268^{6x30}$
603 (Figure 6g), $D130^{3x49}$, and $L272^{6x34}$. $E268^{6x30}$ and $D130^{3x49}$ were already
604 identified in the maximum recruitment analysis and make up the conserved
605 ionic lock while $L272^{6x34}$ is in close proximity to the ionic lock, where mutations
606 are also thought to destabilize the inactive state[67]. Nearby $E268^{6x30}$ in
607 intracellular loop 3 (ICL3), there are nine EC_{50} GoF residues, which is enriched
608 compared to the whole receptor (OR=3.38, $p=7.3x10^{-3}$, one-sided Fisher exact
609 test). This loop connects TM5 and TM6, which make dramatic movements upon
610 receptor activation (Figure 6k). These GoF residues could act in a similar way
611 as $D130^{3x49}$, $E268^{6x30}$, and $L272^{6x34}$, weakening the interaction between TM3
612 and TM6 or other helices. Alternatively, previous studies suggest that ICL3 may
613 have an autoinhibitory role that sterically occludes transducer binding in the
614 inactive state, and that some mutations can abrogate this autoinhibitory effect[68].
615 Further experiments would be necessary to determine the exact mechanism of

616 the EC_{50} GoF phenotypes in ICL3. Additionally, there are two GoF residues in
617 the N-terminus that coincide with the two N-glycosylation sites[69] (N5 and N15,
618 Supplementary Figure 6d) suggesting that loss of these glycosylation
619 modifications increases receptor potency.

620 F282^{6x44} of the PIF motif is the strongest EC_{50} CoF residue (Figure 6h)
621 consistent with previous data showing multiple constitutive activating mutations
622 and at least one attenuating mutation[70]. Likewise, S120^{3x39} and L212^{5x51} are
623 identified as CoF and are directly adjacent to the other two members of the PIF
624 motif (I121^{3x40} and P211^{5x50}) reinforcing that some perturbations to this key
625 switch motif can result in gains or loss of recruitment. There are also several
626 EC_{50} CoF residues in TMs 5 and 6 (4 in each, Supplementary Figure 6c),
627 reflecting the key role of these helices in activation.

628 **Annotation of surface pockets**

629 Finally, we used our β2AR data to illustrate how GPCR-MAPS can be
630 used to functionally annotate and prioritize potential pockets in receptors.
631 AS408 is a negative allosteric modulator (NAM) that binds away from the
632 orthosteric and transducer binding sites, between TMs 3 and 5[71] (Figure 7a).
633 Across the three analyses (maximum recruitment, biased recruitment, and
634 EC_{50}), the 7 residues making up this pocket have 7 total annotations, compared
635 to 96 annotations for all 220 structurally resolved surface exposed residues
636 (>10% relative solvent accessible surface area in 2RH1), indicating a
637 functionally important pocket. In particular, V129^{3x48} is LoF for maximum
638 recruitment as well as EC_{50}, while E122^{3x41} is CoF for maximum recruitment as
639 well as EC_{50}. In addition, C125^{3x44} is CoF for EC_{50} and I214^{5x53} is LoF for EC_{50}.

640 Mutations at P211^{5x50} are biased toward Nb80 recruitment, though AS408 is not
641 thought to be a biased NAM (Supplementary Figure 7a).
642 While the AS408 pocket is replete with functional residues, Cmpd-6, a
643 positive allosteric modulator[72] (PAM), binds in a pocket formed by 9 residues
644 between TMs 2, 3, 4, and ICL2 (Figure 7b) with only 4 annotations. However,
645 one of the residues is D130^{3x49}, where mutations cause strong GoF for both
646 maximum recruitment and EC_{50} (Supplementary Figure 7b), potentially
647 explaining the PAM activity of Cmpd-6. Additionally, while Cpmd-6 mostly forms
648 hydrophobic interactions with β2AR, it does form one hydrogen bond with
649 K149^{4x41}, which is an EC_{50} CoF residue. Taken together, our functional
650 annotations are consistent with the activities of the two known β2AR allosteric
651 modulators.
652 To explore prospective functional annotation, we used Fpocket[73] to
653 predict pockets in the inactive β2AR (PDB ID:2RH1). One additional pocket,
654 between TMs 5 and 6 (Figure 7c), has 8 functional annotations among 10
655 residues, with most being maximum recruitment or EC_{50} LoF or CoF, but also
656 one GoF residue (Supplementary Figure 7c), suggesting a highly active pocket.
657 Indeed, an allosteric agonist was recently found to bind and activate another
658 GPCR, GPR88, at this same site[74]. On the other hand, a nearby and partially
659 overlapping pocket also between TMs 5 and 6 (Figure 7d) has only two
660 annotations among 9 residues (Supplementary Figure 7d), suggesting that
661 perturbation of this pocket is less likely to have functional effects on receptor
662 activation.
663
664 **Discussion**

665 We have presented here GPCR-MAPS, a general, modular, and
666 massively scalable approach that uses pooled experiments to precisely quantify
667 the effects of mutations on the activity and bias of GPCRs. GPCR-MAPS makes
668 multiple important technical advancements over existing technologies for
669 multiplexed receptor assessment[27–30].
670 First, GPCR-MAPS directly quantifies the recruitment of proteins to
671 receptors, so minimizing the confounding effects of non-linear and unequal
672 signal amplification, and feedback by endogenous signaling networks. Second,
673 GPCR-MAPS is highly modular - we demonstrate here the successful use of a
674 nanobody, a mini-G protein, and β-arrestin as signal transducers. The receptor-
675 proximal readout also facilitates comparisons across these signal transducers,
676 which can otherwise be confounded by differential signal amplification and
677 crosstalk[75]. This modular design allows direct study of GPCR bias at the level of
678 recruitment to a receptor. To further demonstrate generalizability, we also show
679 the vasopressin 2 receptor recruitment of miniGs and β-arrestin in response to
680 the endogenous ligand arginine vasopressin (Supplementary Figure 6e). Third,
681 GPCR-MAPS measures and quantitatively deconvolves the strong confounding
682 influence of receptor expression changes on signaling parameters. Fourth, due
683 to its highly multiplexed design and straightforward read out, GPCR-MAPS
684 extends GPCR mutational scanning to the quantification of full dose-response
685 curve activation functions for all possible mutations in a receptor. Comparing
686 our data to previous β2AR mutagenesis studies reveals the importance of all
687 four of these innovations (Supplementary Figure 1i-k, Figure 6b).
688 Applied to β2AR, GPCR-MAPS generates over >150,000 measurements
689 and charts a complete amino acid resolution picture of how quantitative multi-

690 modal receptor phenotypes are encoded and how they can be tuned. The β2AR
691 maps demonstrate the importance of quantitatively controlling for receptor
692 expression changes. After deconvolving mutational effects on expression and
693 recruitment, the residues that disrupt recruitment form a contiguously connected
694 core activation network linking the ligand binding site to the transducer coupling
695 site (Figure 8a). The data also identify the extracellular vestibule, particularly
696 ECL2, as a major allosteric site to control receptor bias, and TM7 near the NPxxY
697 motif as an internal region that allosterically controls receptor activation bias
698 (Figure 8b).

699 Quantifying the full activation functions of >7,500 receptor variants in a
700 single experiment (Supplementary Figure 8) revealed that expression is also a
701 major influence on EC_{50}. This has been previously predicted by theoretical[64–66]
702 models and agrees with recent work on plant hormone receptors[35]. After
703 correcting for the influence of expression changes, EC_{50}-modulating mutations
704 are topologically localized, including potency increasing variants in ICL3 and at
705 both N-glycosylation sites in the unstructured N-terminus (Figure 8c). Mutations
706 that decrease potency are in close proximity to the core activation network and
707 so likely modulate it. In contrast, residues increasing potency are far from the
708 core activation network, likely affecting potency via mechanisms independent of
709 the core activation network. This distinction in the location and likely mechanism
710 of LoF and GoF variants parallels observations in the Src kinase, where
711 mutations that decrease kinase activity are enriched close to the active site but
712 mutations that increase it are not[76].

713 GPCRs are the largest family of membrane receptors and the most
714 successful class of drug targets, with one third of current clinical therapeutics

715 targeting GPCRs[7]. Precision modalities such as biased signaling and allosteric
716 modulation could lead to GPCR drugs with improved efficacy and safety, but
717 these drugs have historically only been found by serendipity. Despite a
718 revolution in GPCR structural biology, we lack sufficient functional
719 understanding of GPCR activation and effector recruitment to design biased or
720 allosteric drugs. We believe that quantifying multi-modal mutational effects will
721 allow the high-resolution functional annotation of many human receptors and
722 accelerate the development of new precision drugs.

723 Residues biased toward Nb80 recruitment are enriched in ECL2, which
724 forms part of the exosite of salmeterol, a β2AR agonist strongly biased toward G
725 protein signaling[59], suggesting mutant behavior is related to drug activity binding
726 at the same sites. Likewise, known allosteric modulators bind in pockets with high
727 impact mutations that match their activities. AS408, a NAM, binds in a pocket with
728 loss- and change-of-function residues, while Cmpd-6FA, a PAM, binds in a
729 pocket with gain- and change-of-function residues. In addition to these known
730 allosteric sites, GPCR-MAPS prioritizes an additional – and, to our knowledge,
731 novel – allosteric pocket to target in β2AR and deprioritizes one where mutations
732 only have limited effects.

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

740 **Materials and Methods**

741 **Plasmids**

742 Plasmids encoding SPARK2 components were gifts from Christina Kim
743 and Alice Ting and were obtained from Addgene (Addgene plasmid # 125228,
744 104839, 104840, 104841). From these plasmids we generated new plasmids
745 using Gibson cloning. attB-HA-b2AR-eLOV-TEVcs-Gal4-UAS-Citrine encodes
746 the receptor fusion followed by the bGH poly(A) signal, then the UAS sequence
747 followed by Citrine fluorescent protein. A 26 nucleotide barcode was inserted in
748 the 3' UTR of the Citrine sequence. To make cloning more efficient, we
749 transferred the fusion coding sequence from the NanoLuc-15aa linker-
750 βarrestin2-HA-GS linker-TEVp plasmid from the pAAV vector (which contains
751 GC-rich inverted terminal repeats) to the pUC19 vector. We then used site
752 directed mutagenesis to introduce the S153N variant to TEV, generating
753 uTEV1Δ). We then made versions of this plasmid with Nb80 and miniG_s
754 sequence instead of β-Arrestin. We also modified the miniG_s vector so that
755 uTEV1Δ was fused N-terminally (NanoLuc-15aa linker-GS linker-TEVp-miniG_s).
756 pCAG-NLS-HA-Bxb1 was a gift from Kenny Matreyek and Doug Fowler.

757 **Development of GPCR-MAPS recruitment assays**

758 We started with the SPARK2 system[41] and extensively re-engineered the
759 system to enable massively multiplexed experiments. First, we combined the
760 GPCR construct and the reporter construct onto a single plasmid and added a
761 DNA barcode (26 nucleotides) in the 3' untranslated region (UTR) of the Citrine
762 reporter so that the GPCR variant on any given plasmid can be identified
763 (Figure 1a and Supplementary Figure 1e). We also added the attB
764 recombination site to enable integration into HEK293T landing pad cells[40],

765 which ensures copy number of exactly one in each cell. We used the active-
766 state-specific nanobody Nb80[42] as the binding partner to optimize the assay.
767 While the original assay used transient transfection to introduce transgenic
768 material and took place over the course of minutes, we found that substantially
769 increased times were necessary to obtain detectable signal, presumably due to
770 the reduction of copy number of the receptor and reporter constructs (24 hours,
771 Supplementary Figure 1a). Further, we tested evolved variants of both TEV and
772 the LOV domain, uTEV1Δ[77] and hLOV1[78]; the combination that yielded the
773 highest signal to noise ratio was uTEV1Δ and eLOV (Supplementary Figure
774 1b). Next, we also validated the use of β-arrestin and miniG_s[79] as the binding
775 partner in this system (Supplementary Figure 1b). Finally, in order to account for
776 differences in barcode abundance in the library, as well as barcode-specific
777 effects on RNA abundance, we also include a condition in which Gal4 is
778 expressed alone (Supplementary Figure 1f); this condition should maximally
779 stimulate all barcodes and provides normalization for the experimental
780 conditions[27].

781 **Saturation mutagenesis library creation**

782 We designed six pools of mutation-bearing oligonucleotides, each pool
783 representing mutated coding sequence for 66-70 amino acids of the full-length
784 (413 amino acids) b2AR with flanking wild-type sequence for amplification from
785 the pool (using primers dialout_tile[1-6]_[F/R], Supplementary Table 5). For
786 each pool we designed primers to amplify the whole plasmid except for the
787 mutated segment (using primers b2AR_satmut_tile[1-6]_[F/R]), then used
788 Gibson assembly to introduce the mutagenic oligos. Oligos were designed to
789 encode all possible single amino acid changes (using the human optimized

790 codon for each alternative amino acid), all premature stop codons, and all
791 synonymous codons when possible (using the most optimal codon, excluding
792 the wild-type codon).

793 **Barcode integration and variant-barcode association**

794 10 µg of the attB-HA-b2AR-eLOV-TEVcs-Gal4-UAS-Citrine saturation
795 mutagenesis plasmid was digested with 40 units PpuMI (New England Biolabs)
796 overnight at 37°, then in the morning 10 units more PpuMI were added and 30
797 minutes more incubation at 37°. Then the cut band was gel purified. The
798 barcode construct (oTLM_257) was made double stranded by PCR amplifying
799 with oTLM_258 and oTLM_259 (Supplementary Table 5). Double stranded
800 barcode was incorporated into the cut plasmid backbone by Gibson assembly
801 and transformed into electrocompetent bacteria. Following recovery in SOC
802 media, we plated 0.1% of the transformation to estimate the number of clones,
803 and inoculated the rest into LB media with ampicillin for overnight growth.
804 Estimated number of clones was 300,000. In the morning the bacteria was
805 pelleted and plasmid was midi-prepped (Qiagen Plasmid Plus Midi Kit).

806 We used PacBio long read sequencing to associate each variant with the
807 barcode. We digested the plasmid pool with BsiWI (New England Biolabs) to
808 release the coding sequence and barcode and SMRT bell adapters were ligated
809 according to manufacturer's instructions. The library was sequenced on a
810 PacBio Sequel IIe. We used alignparse[80] to align reads, quality filter, and call
811 sequences for each barcode (Supplementary Figure 1).

812 **Landing pad integration**

813 HEK 293T LLP-iCasp9-Blast[40] cells were a gift from Kenny Matreyek.
814 Cells were grown in DMEM with GlutaMAX (Gibco), with Tetracyline-free fetal

815 bovine serum. 14 million cells were plated onto a T225 flask. The next day,
816 transfection was done with Lipofectamine 3000 (Invitrogen) with 17.5 µg of
817 pCAG-NLS-HA-Bxb1 and 17.5 µg of the mutated and barcoded attB-HA-b2AR-
818 eLOV-TEVcs-Gal4-UAS-Citrine. The cells were allowed to incubate for two
819 days. Then, the media was removed and replaced with media containing 2
820 µg/mL doxycycline and incubated for one day. Then, the media was changed
821 for media containing 2 µg/mL doxycycline and 10 nM rimiducid
822 (MedChemExpress) and incubated for one day. Then media was changed to
823 include just 2 µg/mL doxycycline and outgrowth of cells was allowed for ~one
824 week, passaging cells when they reached 90% confluency.

825 **GPCR-MAPS recruitment experiment**

826 10 million cells were plated onto each T175 flask and allowed to incubate
827 overnight. The next day, the cells were transfected with 36 µg of transducer
828 plasmids (pUC19-NanoLuc-β-arrestin-uTEVp1 or pUC19-NanoLuc-Nb80-
829 uTEVp1), using Lipofectamine 3000 and allowed to incubate overnight. The
830 next morning, cells were treated with media containing furimazine and either
831 drug or DMSO: [13.05 mL DMEM, 126.56 µL furimazine, 270 µL 1M HEPES,
832 and 13.5 µL isoproterenol or DMSO]. Furimazine was obtained from the Nano-
833 Glo Live Cell Assay System (Promega). Isoprenaline was obtained from Sigma
834 (catalog number 15627). Cells were incubated in the dark (covered with
835 aluminum foil) at 37° for 24 hours. Then, cells were lifted with Trypsin-EDTA,
836 pelleted and stored in -80 until RNA extraction. For the Gal4 maximum
837 stimulation condition, after 10 million cells were plated and grown overnight, 36
838 µg of pUC19-CMV-Gal4 was transfected with Lipofectamine 3000. The cells

839 were allowed to incubate for 30 hours, then cells were pelleted and stored at -
840 80°.

841 **RNA isolation and sequencing library preparation**

842 For the first set of experiments (Nb80 and β-arrestin recruitment with 10
843 μM isoprenaline), cell pellets were thawed and resuspended in buffer RLT with
844 added beta-mercaptoethanol. Then, they were homogenized by centrifuging
845 through QIAshredder columns. RNA was then purified from the homogenate
846 with RNeasy Midi-prep kit (Qiagen) and eluted with 400 μL of water. Between
847 120 and 160 μg of total RNA was used as input for reverse transcription (RT)
848 reactions (6-8 μg per reaction, 20 reactions per sample) using SuperScript III
849 reverse transcriptase (Invitrogen). For the dose response experiment,
850 Dynabeads mRNA direct purification kit (ThermoFisher Scientific) was used
851 according to manufacturers instructions and mRNA was eluted in 80 μL of Tris-
852 HCl. Total mRNA yields were between 11.5 μg and 16 μg. Between 5.75 and 8
853 μg mRNA was used as input for RT (1.15-1.6 μg per reaction, 10 reactions per
854 sample). For both sets of reactions, RT was carried out with 2 pmol of primer
855 (oTLM_266, Supplementary Table 5) and incubated at 55° for one hour then
856 75° for 15 minutes. Following this, RNase A was added to a final concentration
857 of 100 μg/mL and reactions incubated at 37° for 30 minutes. Then, reactions
858 were cleaned up with Nucleospin PCR cleanup columns (Macherey Nagel) and
859 eluted in 50 μL of elution buffer. 0.5 μL of the cDNA eluate was used in a qPCR
860 reaction (Q5 High-Fidelity polymerase, New England Biolabs, and EvaGreen
861 fluorescent dye, Biotium) to determine the number of cycles required to reach
862 near-plateau phase. Then, 40 μL of cDNA was amplified across 4x 50 μL PCR

863 reactions with the program 98° for 30s, followed by 13-16 cycles of [98° for 15s,
864 64° for 30s, 72° for 30s].
865 Reactions were column purified and 4% of this was used as input for a
866 second round PCR, which added the rest of the Illumina adapter as well as
867 unique index sequences. The second round PCR was 98° for 30s, followed by 5
868 cycles of [98° for 15s, 64° for 30s, 72° for 30s]. Frameshifting primers were
869 used to introduce variability in the first few cycles of the sequencing reaction
870 (oTLM_277-282, Supplementary Table 5). Reactions were then gel purified and
871 analyzed on TapeStation (Agilent Technologies). Sequencing was done with
872 Illumina NovaSeq with 2x50 paired end reads.
873 An additional step was taken for the dose response experiment.
874 Following RT, we performed triplicate qPCR reactions with each cDNA sample,
875 along with a set of serial dilutions of the original library plasmid, in order to
876 estimate the relative quantity of Citrine RNA produced from each experiment.
877 We used primers oTLM_021 and oTLM_024, which amplify the 5' UTR of
878 Citrine. We chose to amplify this region instead of the barcode (as was done for
879 the library prep) to avoid idiosyncratic effects of amplification efficiency caused
880 by diverse and potentially different barcode identities across different samples.
881 Following qPCR we fit a standard curve (Supplementary Fig. 5b) then used this
882 to estimate RNA abundance for the dose response samples (Figure 5b).
883 **Surface expression experiment cell culture and FACS**
884 Cells were cultured in 245x245mm cell culture flasks until ~80%
885 confluent. Then, cells were lifted with Trypsin-EDTA, pelleted, supernatant
886 removed, and resuspended in blocking buffer (1% bovine serum albumin in
887 phosphate buffered saline). Cells were counted and 50 M cells were transferred

888 to a 5 mL conical tube with 3.5 mL of blocking buffer, and incubated on ice,
889 covered from light, with rocking for 30 minutes. Then, anti-HA antibody (Cell
890 Signaling HA-Tag 6E2 Alexa Fluor 647 Conjugate) was added at dilution of
891 1:100 and incubated on ice, covered from light, with rocking for 60 more
892 minutes. Then, cells were pelleted, supernatant removed, and resuspended in
893 blocking buffer. Sorting was done on the BD FACSAria instrument. Bins were
894 drawn so that each represented ~25% of cells. In total, ~20M cells were sorted
895 for replicate 1, and ~30M cells were sorted for replicates 2 and 3. 30-45M 2x50
896 basepair reads were collected on an Illumina NovaSeq instrument for each
897 replicate.

898 **Surface expression sequencing library preparation**

899 Sorted cells were pelleted and DNA extracted using Qiagen DNeasy
900 Blood & Tissue kit, using 2 columns for each sample. 200 µL elution buffer was
901 used to elute from each column, then eluates for each sample were combined.
902 Then, 320 µL of each sample was was used as input for 16x 50 µL PCRs with
903 primers adjacent to the barcode that include partial Illumina adapters and also
904 varying lengths of degenerate bases to increase nucleotide complexity in the
905 first few cycles of the sequencing run (oTLM277-oTLM282, Supplementary
906 Table 5). Cycling was 98° for 30s, followed by 25 cycles of [98° for 15s, 64° for
907 30s, 72° for 30s]. Reactions were column purified and 4% of this was used as
908 input for a second round PCR, which added the rest of the Illumina adapter as
909 well as unique index sequences. The second round PCR was 98° for 30s,
910 followed by 5 cycles of [98° for 15s, 64° for 30s, 72° for 30s]. The products of
911 the second PCR were gel purified and analyzed on TapeStation, then
912 sequenced on Illumina NovaSeq with 2x50 paired end reads.

913 **Sequencing data processing**

914 Paired-end sequencing reads were first merged with vsearch[81] and then
915 adapters were trimmed with cutadapt[82]. This leaves each read representing the
916 barcode sequence, which was then matched with barcodes identified in the
917 variant-barcode association with a custom script.

918 **Calculation of surface expression scores**

919 Surface expression scores were calculated as previously described[25].
920 Briefly, the frequency of each variant in each bin is enumerated by combining
921 counts of all barcodes associated with that variant. The frequency, along with
922 the total number of cells sorted into each bin, is used to estimate the number of
923 cells in each bin. Then, the log10 transformed geometric mean fluorescence
924 value of all cells sorted into each bin is combined with the estimated cell counts
925 to arrive at the raw surface expression estimate. These estimates were then
926 normalized between 0 (represented by the median of presumed loss-of-function
927 variants: e.g. premature stop codons before the 300th amino acid) and 1, the
928 wild-type score. Surface expression scores were considered high confidence if
929 the estimated number of cells sorted was >50.

930 **Calculation of recruitment scores**

931 First, barcode read counts for all barcodes representing the same variant
932 were combined and then the frequency of that variant was calculated by
933 dividing that number by the total number of barcode counts in the sample.
934 Then, to arrive at a raw score, we divided the variant frequency in the
935 experimental sample by the variant frequency in the Gal4 sample, and take the
936 $\log_2$ of this. The Gal4 sample represents a maximum stimulation condition in
937 which the barcode frequency in the library, as well as any idiosyncratic effects

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

942 **LOWESS modeling and identifying outlier positions**

943 The python package statsmodels.nonparametric.smoothers_lowess[83]
944 was used with the following parameter values: frac=0.5, it=3, delta=0.0.
945 Following curve fitting, residuals were calculated as the distance in the y-axis
946 between the data point and the line, and z-scores were calculated as the
947 residual / standard error of the mean. In the case of Nb80 versus β-arrestin, the
948 average standard error of the mean between the two measurements was used.

949 **Structural analysis**

950 Residue contacts were defined by GetContacts
951 (https://getcontacts.github.io/index.html). We used the definition of the
952 isoprenaline binding site as presented in the publication solving and describing
953 the structure[44]. Specific interactions between residues and isoprenaline atoms
954 were taken from the publication as well as GetContacts. The binding interface of
955 β2AR with Nb80 was defined by GetContacts analysis of structure. The
956 predicted inactive and active structures in Figure 6k were taken from GPCRdb
957 and were produced with AlphaFold-Multistate[45]. These structures were used to
958 enable visualization of the entire ICL3, which is not resolved in experimentally
959 solved structures.

960 **Dose response curve fitting**

961 In the package neutcurve[84], we used the neutcurve.hillcurve.HillCurve
962 function to fit curves with the following options:

963 infectivity_or_neutralized='neutralized', fix_slope_first=True, fs_stderr=None,
964 fixbottom=False, fixtop=False, fixslope=False, fitlogc=False,
965 use_stderr_for_fit=False, no_curve_fit_first=False. Following the fitting
966 procedure, we then filtered curves based on heuristics: $R^2 > 0.9$, EC50 within
967 the measured range (i.e. between 10pM and 10uM), and $E_{max} < 1.5x$ the
968 maximum measured recruitment value. See Supplementary Figure 5e for the
969 number of variants filtered.

970 **Fpocket pocket prediction**

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

975

976 **Acknowledgments**

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

981

982 **Data availability**

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

986

987 **Code availability**

988 Custom code to reproduce analyses can be found at github
989 (https://github.com/lehner-lab/GPCR-MAPS).
990
991 **Competing interests**
992 CRG and the Genome Research Ltd have filed a patent application for the
993 use of the methods described in this study for mapping allostery in GPCRs; T.M.
994 and B.L. are listed as inventors. B.L. is a founder and shareholder of ALLOX.
995

996 **Figure 1. GPCR-MAPs**

997 **a** Schematic of the experimental framework. A mutagenized GPCR (in this case
998 β2AR) is encoded in a plasmid that also includes the attB recombination site, an
999 N-terminal HA epitope tag, C-terminal fusions of a light, oxygen, and voltage
1000 sensitive (LOV) domain, the cleavage site of TEV protease (TEVcs), and Gal4
1001 transcription factor. Downstream, on the same plasmid, is the fluorescent
1002 protein Citrine, driven by Gal4-responsive UAS promoter, with a degenerate
1003 DNA barcode included in the 3' untranslated region. This construct is
1004 recombined into HEK 293T LLP-iCasp9-Blast landing pad cells. Then, two
1005 assays are performed to measure the surface expression (FACS-based) and
1006 recruitment activity (transcriptional reporter) of variants *en masse***.** The outputs
1007 of these experiments (DNA for FACS and RNA for transcriptional reporter) are
1008 sequenced with short read technology. Estimated surface expression and
1009 recruitment scores are then modeled to understand the contribution of each of
1010 these to the variant effects. **b** Detailed schematic of the recruitment assay. The
1011 receptor construct is expressed from landing pad cells while the transducer
1012 construct is transiently transfected. Addition of agonist (isoprenaline in this
1013 case) and furimazine results in recruitment of transducer to the receptor fusion
1014 and subsequent cleavage of the TEVcs, liberating Gal4 which moves to the
1015 nucleus and drives transcription of the UAS-Citrine-barcode construct. **c**
1016 Replicate Pearson correlations for the surface expression, Nb80 recruitment,
1017 and βArr recruitment. $p<10^{-100}$ in each case. **d** Heatmap representation of the
1018 surface expression data. β2AR positions on the x-axis, mutation on the y-axis.
1019 Gray cells indicate expression similar to wild-type while blue and red cells
1020 indicate reduced and increased expression, respectively. **e** Histogram of

1021 surface expression variant scores; nonsense variants in red, synonymous wild-
1022 type in green, and missense in grey. **f** Heatmap representation of the Nb80
1023 recruitment data. Axes and color scale as in (d). **g** Histogram of Nb80
1024 recruitment scores; colors as in (e). **h** Heatmap representation of β-Arrestin
1025 recruitment scores. Axes and color scale as in (d). **i** Histogram of β-Arrestin
1026 recruitment scores; colors as in (e). **j** Surface expression scores for residues in
1027 the isoprenaline binding site in the solved structure (PDB ID: 7DHR). **k** Nb80
1028 recruitment scores for residues in the isoprenaline binding site. **l** β-Arrestin
1029 recruitment scores for residues in the isoprenaline binding site.
1030
1031 **Figure 2. Deconvolving recruitment and expression**
1032 **a** Surface expression versus Nb80 recruitment scores for missense variants.
1033 $p<10^{-100}$. **b** Surface expression versus β-Arrestin recruitment scores for
1034 missense variants. $p<10^{-100}$. **c** Schematic of the method for identifying residues
1035 contributing to recruitment beyond what can be explained by surface
1036 expression. First, the recruitment data is related to surface expression with a
1037 LOWESS regression. Then, for each variant, residuals to this fit are calculated.
1038 Then, residues with significantly biased residuals are identified as loss-of-
1039 function or gain-of-function via Mann Whitney U test. Finally, positions where
1040 variants have a wide range of effects, quantified using the interquartile range of
1041 variant z-scores, are identified as change-of-function. **d** LOWESS fit of surface
1042 expression versus Nb80 recruitment, with variant points colored by residual to
1043 the fit. **d** Same as (e) but for β-Arrestin recruitment. **f** Volcano plot showing the
1044 position-average z-score and the minus log10 transformed p-value of all
1045 residues for Nb80 recruitment. Residues with significantly biased residuals are

1046 colored according to the average z-score. **g** Same as in (f) but for β-Arrestin
1047 recruitment. **h** Venn diagrams showing the number of residues called loss-,
1048 gain-, and chage-of-function for Nb80 and β-Arrestin recruitment. **i** Surface
1049 expression and recruitment effects of variants at D113^{3x32}, plotted over the
1050 population of variant effects, for Nb80 and β-Arrestin recruitment versus surface
1051 expression. Error bars represent standard error of the mean. **j** Same as (i) but
1052 with variants at position E268^{6x30}. **k** Same as (i) and (j) but with variants at
1053 position L75^{2x46}. **l** Loss-of-function residues highlighted in purple on the
1054 isoprenaline-bound β2AR structure (PDB ID: 7DHR, isoprenaline in yellow). **m**
1055 Gain-of-function residues highlighted in green on β2AR. **n** Change-of-function
1056 residues highlighted in red on β2AR
1057
1058 **Figure 3. A core activation network links ligands to transducers**
1059 **a** Inactive and active state structures of β2AR with LoF residues colored purple.
1060 Those LoF residues that make a direct contact with other LoF residues are
1061 visualized as pseudobonds connecting them, while disconnected LoF residues
1062 are shown with spheres. **b** Enlarged view of the core activation network with
1063 residues labeled and their participation in different structural motifs. **c** Venn
1064 diagram showing the overlap between network residues identified in this work
1065 and previous work. **d** Variant surface expression and Nb80 recruitment at
1066 S120^{3x39}, a CoF residue, are plotted against the population of all variants. **e**
1067 Same as (d) but with variants at L124^{3x43}, a GoF residue. **f** The isoprenaline
1068 binding pocket, with isoprenaline highlighted in yellow, and the residues making
1069 up the binding site colored according to their median expression adjusted Nb80
1070 recruitment level. **g** Violin plot of expression adjusted Nb80 recruitment at the

1071 isoprenaline binding site residues. Violins are colored according to their median
1072 value, with the same color scale as in (f).
1073
1074 **Figure 4. Genetic control of receptor recruitment bias**
1075 **a** Schematic showing β2AR activation through β-Arrestin and Nb80 in response
1076 to isoprenaline. **b** Nb80 versus β-Arrestin recruitment for missense variants.
1077 $p<10^{-100}$, n=6,791. **c** Nb80 versus β-Arrestin recruitment with a LOWESS fit and
1078 variant scores colored by residual to the fit line. Individual variant scores, with
1079 error bars representing standard error of the mean, shown for three variants
1080 previously reported to cause biased signaling. **d** Volcano plot showing the
1081 position-average z-score and the minus log10 transformed p-value of all
1082 residues for Nb80 versus β-Arrestin recruitment. Residues with significantly
1083 biased residuals are colored according to the average z-score. **e** Nb80 and β-
1084 Arrestin recruitment effects of variants at $W99^{23x50}$, plotted over the population
1085 of variant effects, error bars represent standard error of the mean. **f** Same as in
1086 (e) but with variants at $P138^{34x50}$. **g** Same as in (e) but with variants at $F332^{8x50}$.
1087 **h** β-Arrestin biased, Nb80 biased, or bias switch residues highlighted on the
1088 solved structure, 7DHR. Isoprenaline shown in yellow. **i** $F193^{45x52}$ and $Y318^{7x34}$
1089 form a "lid" over isoprenaline in the solved structure (7DHR). **j** Hydrophobicity
1090 (Kyte-Doolittle scale) of the mutant amino acid compared with the Nb80 versus
1091 β-Arrestin recruitment bias. **k** Enlarged view of the extracellular vestibule.
1092 Colored residues are significantly Nb80-biased (orange), β-Arrestin-biased
1093 (blue), or bias switches (magenta). **l** Enlarged view of NPxxY motif and TM7,
1094 colored residues same as in (k).
1095

1096 **Figure 5. Quantifying >7,500 dose-response relationships**

1097 **a** Experimental design: serial dilutions of isoprenaline were used to stimulate
1098 landing pad cells expressing the β2AR library. Then, the GPCR-MAPS
1099 recruitment assay was performed and RNA was collected from the stimulated
1100 cells. **b** The RNA abundance from each sample was first quantified with qPCR,
1101 then the variant frequencies were estimated with short-read sequencing. **c** Per-
1102 variant normalized scores were then calculated by normalizing variant
1103 frequencies in each condition by the RNA abundance of each condition. **d**
1104 Heatmap representations of the 8 highest concentration drug treatment
1105 conditions. β2AR positions on the x-axis, mutation identities on the y-axis. One
1106 y-axis is blown up; the rest of the heatmaps follow this same scheme. Grey cells
1107 indicate wild-type-like recruitment, blue and red indicate reduced and increased
1108 recruitment, respectively. One color bar is blown up; the rest of the heatmaps
1109 follow this same scheme. **e** Clustered heatmap of recruitment intensities. Color
1110 bar on top. Three main clusters correspond to wild-type-like, moderately
1111 damaging, and complete loss-of-function curves. **f** The curve fit to the wild-type
1112 data, with E_{max}, E_{min}**,** EC_{50}, and slope indicated. **g** Fit curves from the wild-type-
1113 like cluster in black, wild-type curve in red. **h** Fit curves from the moderately
1114 damaging cluster in black, wild-type curve in red. **i** Fit curves from the loss-of-
1115 function cluster in black, wild-type curve in red.

1116

1117 **Figure 6. Genetic determinants of potency**

1118 **a** Schematic showing example curves for increased and decreased potency. **b**
1119 Comparison of variant EC_{50} estimates from this study with previously published
1120 estimates[18]. Red dashed line is x=y line. **c** Spearman correlations of curve

1121 fitting parameter estimates and surface expression scores. **d** Surface
1122 expression scores compared with log-transformed EC_{50} estimates. $p<10^{-100}$. **e**
1123 Volcano plot showing the position-average z-score and the minus log10
1124 transformed p-value of all residues for log EC_{50} versus surface expression. **f**
1125 Log-transformed EC_{50} versus surface expression for variants at $R328^{7x55}$,
1126 plotted over the population of variant effects, error bars represent standard error
1127 of the mean. **g** Same as in (f) but with variants at $E268^{6x30}$. **h** Same as in (f) but
1128 with variants at $F282^{6x44}$. **i** EC_{50} loss-of-function, gain-of-function, or change-of-
1129 function residues highlighted on the solved structure, 7DHR. Isoprenaline show
1130 in yellow. **j** Violin plots showing minimum distance between LoF, GoF, or CoF
1131 residues from the core activation network in AlphaFold-Multistate active state
1132 model. Distances capped at 20Å. **k** Portions of TMs 3, 5, and 6, as predicted by
1133 AlphaFold-Multistate and downloaded from GPCRdb[45]. LoF residues colored
1134 aquamarine, GoF residues colored brown, and CoF residues colored salmon.
1135

1136 **Figure 7. Rationalization and prioritization of allosteric pockets**

1137 **a** The binding pocket of known negative allosteric modulator, AS408, shown in
1138 yellow. PDB ID: 6OBA. Residues making direct contact with the NAM are
1139 colored beige. **b** Expression normalized maximum recruitment through Nb80,
1140 recruitment bias, and expression normalized potency of the mutations at sites
1141 making up the binding pocket of AS408. Shown below the heatmaps are the
1142 categorical annotations of significant residues: purple, green, and red indicate
1143 LoF, GoF, and CoF residues for maximum recruitment. Orange, blue, and
1144 magenta indicate Nb80 biased, β-Arrestin, biased, or bias switches.
1145 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 recruitment and potency analyses, the color from maximum recruitment 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 recruitment analysis. Core activation network shown as purple lines, residues or topological domains colored according to annotation. **b** Selected findings from biased recruitment 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.

1160 **References**

1161 1. Bockaert, J. & Philippe Pin, J. Molecular tinkering of G protein-coupled receptors: an
1162 evolutionary success. *EMBO J.* **18**, 1723–1729 (1999).
1163 2. Farrens, D. L., Altenbach, C., Yang, K., Hubbell, W. L. & Khorana, H. G. Requirement
1164 of Rigid-Body Motion of Transmembrane Helices for Light Activation of Rhodopsin.
1165 *Science* **274**, 768–770 (1996).
1166 3. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of
1167 G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
1168 4. Jean-Charles, P.-Y., Kaur, S. & Shenoy, S. K. GPCR signaling via β-arrestin-dependent
1169 mechanisms. *J. Cardiovasc. Pharmacol.* **70**, 142–158 (2017).
1170 5. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there?
1171 *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
1172 6. Sriram, K. & Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs:
1173 How Many Targets and How Many Drugs? *Mol. Pharmacol.* **93**, 251–258 (2018).
1174 7. Wold, E. A., Chen, J., Cunningham, K. A. & Zhou, J. Allosteric Modulation of Class A
1175 GPCRs: Targets, Agents, and Emerging Concepts. *J. Med. Chem.* **62**, 88–127 (2019).
1176 8. Wootten, D., Christopoulos, A., Marti-Solano, M., Babu, M. M. & Sexton, P. M.
1177 Mechanisms of signalling and biased agonism in G protein-coupled receptors. *Nat.*
1178 *Rev. Mol. Cell Biol.* **19**, 638–653 (2018).
1179 9. May, L. T., Leach, K., Sexton, P. M. & Christopoulos, A. Allosteric modulation of G
1180 protein-coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* **47**, 1–51 (2007).
1181 10. Kenakin, T. Biased Receptor Signaling in Drug Discovery. *Pharmacol. Rev.* **71**,
1182 267–315 (2019).

1183 11. Wootten, D., Christopoulos, A. & Sexton, P. M. Emerging paradigms in GPCR
1184 allostery: implications for drug discovery. *Nat. Rev. Drug Discov.* **12**, 630–644
1185 (2013).
1186 12. Lorente, J. S. *et al.* GPCR drug discovery: new agents, targets and indications.
1187 *Nat. Rev. Drug Discov.* 1–22 (2025) doi:10.1038/s41573-025-01139-y.
1188 13. Bermudez, M., Nguyen, T. N., Omieczynski, C. & Wolber, G. Strategies for the
1189 discovery of biased GPCR ligands. *Drug Discov. Today* **24**, 1031–1037 (2019).
1190 14. Allosteric Modulator Discovery: From Serendipity to Structure-Based Design. *J.*
1191 *Med. Chem.* **62**, 6405–6421 (2019).
1192 15. Venkatakrishnan, A. J. *et al.* Diverse activation pathways in class A GPCRs
1193 converge near the G-protein-coupling region. *Nature* **536**, 484–487 (2016).
1194 16. Zhou, Q. *et al.* Common activation mechanism of class A GPCRs. *eLife* **8**, e50279
1195 (2019).
1196 17. Ward, W. H. J., Timms, D. & Fersht, A. R. Protein engineering and the study of
1197 structure—function relationships in receptors. *Trends Pharmacol. Sci.* **11**, 280–284
1198 (1990).
1199 18. Heydenreich, F. M. *et al.* Molecular determinants of ligand efficacy and potency
1200 in GPCR signaling. *Science* **382**, eadh1859 (2023).
1201 19. Morales-Pastor, A. *et al.* Multiple triggers converge to preferential effector
1202 coupling in the CB2R through a complex allosteric communication network.
1203 2024.02.27.582279 Preprint at https://doi.org/10.1101/2024.02.27.582279 (2024).
1204 20. Cao, Y. *et al.* Unraveling allostery within the angiotensin II type 1 receptor for
1205 Gαq and β-arrestin coupling. *Sci. Signal.* **16**, eadf2173 (2023).

1206 21. Penn, W. D. *et al.* Probing biophysical sequence constraints within the
1207 transmembrane domains of rhodopsin by deep mutational scanning. *Sci. Adv.* **6**,
1208 eaay7505 (2020).
1209 22. McKee, A. G. *et al.* Systematic profiling of temperature- and retinal-sensitive
1210 rhodopsin variants by deep mutational scanning. *J. Biol. Chem.* **297**, 101359 (2021).
1211 23. Heredia, J. D. *et al.* Mapping Interaction Sites on Human Chemokine Receptors
1212 by Deep Mutational Scanning. *J. Immunol.* **200**, 3825–3839 (2018).
1213 24. Park, J. *et al.* Structural architecture of a dimeric class C GPCR based on co-
1214 trafficking of sweet taste receptor subunits. *J. Biol. Chem.* **294**, 4759–4774 (2019).
1215 25. Mighell, T. L. & Lehner, B. A small molecule stabilizer rescues the surface
1216 expression of nearly all missense variants in a GPCR. *Nat. Struct. Mol. Biol.* 1–12
1217 (2025) doi:10.1038/s41594-025-01659-6.
1218 26. Yoo, J. I., Daugherty, P. S. & O'Malley, M. A. Bridging non-overlapping reads
1219 illuminates high-order epistasis between distal protein sites in a GPCR. *Nat.*
1220 *Commun.* **11**, 690 (2020).
1221 27. Jones, E. M. *et al.* Structural and functional characterization of G protein–
1222 coupled receptors with deep mutational scanning. *eLife* **9**, e54895 (2020).
1223 28. Howard, M. K. *et al.* Molecular basis of proton sensing by G protein-coupled
1224 receptors. *Cell* **188**, 671-687.e20 (2025).
1225 29. Howard, C. J. *et al.* High resolution deep mutational scanning of the
1226 melanocortin-4 receptor enables target characterization for drug discovery. *eLife*
1227 **13**, (2024).

1228 30. LaFlam, T. N. *et al.* Phenotypic pleiotropy of missense variants in human B cell
1229 confinement receptor P2RY8. *Cell Genomics* 100981 (2025)
1230 doi:10.1016/j.xgen.2025.100981.
1231 31. Fallon, B. S., Rondem, K. E., Mumby, E. J. & English, J. G. Biased Signaling in G
1232 Protein-Coupled Receptors: Understanding the Biological Relevance and Tools for
1233 Probing Functionally Selective Ligands. *Biochemistry* **64**, 1425–1436 (2025).
1234 32. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces1. *J. Mol.*
1235 *Biol.* **280**, 1–9 (1998).
1236 33. Hedderich, J. B. *et al.* The pocketome of G-protein-coupled receptors reveals
1237 previously untargeted allosteric sites. *Nat. Commun.* **13**, 2567 (2022).
1238 34. Ippolito, M. & Benovic, J. L. Biased agonism at β-adrenergic receptors. *Cell.*
1239 *Signal.* **80**, 109905 (2021).
1240 35. Maximilian R Stammnitz & Ben Lehner. The genetic architecture of an allosteric
1241 hormone receptor. *bioRxiv* (2025).
1242 36. Kroeze, W. K. *et al.* PRESTO-Tango as an open-source resource for interrogation
1243 of the druggable human GPCRome. *Nat. Struct. Mol. Biol.* **22**, 362–369 (2015).
1244 37. Macdonald, C. B. *et al.* DIMPLE: deep insertion, deletion, and missense
1245 mutation libraries for exploring protein variation in evolution, disease, and biology.
1246 *Genome Biol.* **24**, 36 (2023).
1247 38. Weile, J. *et al.* A framework for exhaustively mapping functional missense
1248 variants. *Mol. Syst. Biol.* **13**, 957 (2017).
1249 39. Mighell, T. L., Toledano, I. & Lehner, B. SUNi mutagenesis: Scalable and uniform
1250 nicking for efficient generation of variant libraries. *PLOS ONE* **18**, e0288158 (2023).

1251 40. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An
1252 improved platform for functional assessment of large protein libraries in
1253 mammalian cells. *Nucleic Acids Res.* **48**, e1 (2020).
1254 41. Kim, C. K., Cho, K. F., Kim, M. W. & Ting, A. Y. Luciferase-LOV BRET enables
1255 versatile and specific transcriptional readout of cellular protein-protein interactions.
1256 *eLife* **8**, e43826 (2019).
1257 42. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the
1258 β2 adrenoceptor. *Nature* **469**, 175–180 (2011).
1259 43. Nehmé, R. *et al.* Mini-G proteins: Novel tools for studying GPCRs in their active
1260 conformation. *PLoS ONE* **12**, e0175642 (2017).
1261 44. Yang, F. *et al.* Different conformational responses of the β2-adrenergic
1262 receptor-Gs complex upon binding of the partial agonist salbutamol or the full
1263 agonist isoprenaline. *Natl. Sci. Rev.* **8**, nwaa284 (2021).
1264 45. Pándy-Szekeres, G. *et al.* GPCRdb in 2023: state-specific structure models using
1265 AlphaFold2 and new ligand resources. *Nucleic Acids Res.* **51**, D395–D402 (2023).
1266 46. Valentin-Hansen, L. *et al.* The Arginine of the DRY Motif in Transmembrane
1267 Segment III Functions as a Balancing Micro-switch in the Activation of the β2-
1268 Adrenergic Receptor. *J. Biol. Chem.* **287**, 31973–31982 (2012).
1269 47. Nygaard, R., Frimurer, T. M., Holst, B., Rosenkilde, M. M. & Schwartz, T. W.
1270 Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol.*
1271 *Sci.* **30**, 249–259 (2009).
1272 48. Yang, D. *et al.* G protein-coupled receptors: structure- and function-based drug
1273 discovery. *Signal Transduct. Target. Ther.* **6**, 1–27 (2021).

1274 49. Ragnarsson, L., Andersson, Å., Thomas, W. G. & Lewis, R. J. Mutations in the
1275 NPxxY motif stabilize pharmacologically distinct conformational states of the α1B-
1276 and β2-adrenoceptors. *Sci. Signal.* **12**, eaas9485 (2019).
1277 50. Zhao, L., He, X., Jiang, H. & Cheng, X. Computational characterization of
1278 transducer recognition of β2 adrenergic receptor. *Biochem. Biophys. Res. Commun.*
1279 **592**, 67–73 (2022).
1280 51. Zhou, Q. *et al.* Common activation mechanism of class A GPCRs. *eLife* **8**, e50279
1281 (2019).
1282 52. Rovati, G. E., Capra, V. & Neubig, R. R. The highly conserved DRY motif of class
1283 A G protein-coupled receptors: beyond the ground state. *Mol. Pharmacol.* **71**, 959–
1284 964 (2007).
1285 53. Rasmussen, S. G. F. *et al.* Mutation of a Highly Conserved Aspartic Acid in the
1286 *β*2 Adrenergic Receptor: Constitutive Activation, Structural Instability, and
1287 Conformational Rearrangement of Transmembrane Segment 6. *Mol. Pharmacol.* **56**,
1288 175–184 (1999).
1289 54. Tao, Y. X., Abell, A. N., Liu, X., Nakamura, K. & Segaloff, D. L. Constitutive
1290 activation of G protein-coupled receptors as a result of selective substitution of a
1291 conserved leucine residue in transmembrane helix III. *Mol. Endocrinol. Baltim. Md*
1292 **14**, 1272–1282 (2000).
1293 55. Wheatley, M. *et al.* Lifting the lid on GPCRs: the role of extracellular loops. *Br. J.*
1294 *Pharmacol.* **165**, 1688–1703 (2012).
1295 56. Ippolito, M., De Pascali, F., Inoue, A. & Benovic, J. L. Phenylalanine 193 in
1296 Extracellular Loop 2 of the β2-Adrenergic Receptor Coordinates β-Arrestin
1297 Interaction. *Mol. Pharmacol.* **101**, 87–94 (2022).

1298 57. Choi, M. *et al.* G protein–coupled receptor kinases (GRKs) orchestrate biased
1299 agonism at the β2-adrenergic receptor. *Sci. Signal.* **11**, eaar7084 (2018).
1300 58. Sanchez-Soto, M. *et al.* A structural basis for how ligand binding site changes
1301 can allosterically regulate GPCR signaling and engender functional selectivity. *Sci.*
1302 *Signal.* **13**, eaaw5885 (2020).
1303 59. Masureel, M. *et al.* Structural insights into binding specificity, efficacy and bias
1304 of a β2AR partial agonist. *Nat. Chem. Biol.* **14**, 1059–1066 (2018).
1305 60. Bock, A. & Bermudez, M. Allosteric coupling and biased agonism in G protein-
1306 coupled receptors. *FEBS J.* **288**, 2513–2528 (2021).
1307 61. Woo, A. Y.-H. *et al.* Tyrosine 308 is necessary for ligand-directed Gs protein-
1308 biased signaling of β2-adrenoceptor. *J. Biol. Chem.* **289**, 19351–19363 (2014).
1309 62. Dror, R. O. *et al.* Pathway and mechanism of drug binding to G-protein-coupled
1310 receptors. *Proc. Natl. Acad. Sci.* **108**, 13118–13123 (2011).
1311 63. Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. & Wüthrich, K. Biased Signaling
1312 Pathways in β2-Adrenergic Receptor Characterized by 19F-NMR. *Science* **335**, 1106–
1313 1110 (2012).
1314 64. Chure, G. *et al.* Predictive shifts in free energy couple mutations to their
1315 phenotypic consequences. *Proc. Natl. Acad. Sci.* **116**, 18275–18284 (2019).
1316 65. Razo-Mejia, M. *et al.* Tuning Transcriptional Regulation through Signaling: A
1317 Predictive Theory of Allosteric Induction. *Cell Syst.* **6**, 456-469.e10 (2018).
1318 66. Steiner, P. J. *et al.* A Closed Form Model for Molecular Ratchet-Type Chemically
1319 Induced Dimerization Modules. *Biochemistry* **62**, 281–291 (2023).
1320 67. Weis, W. I. & Kobilka, B. K. The Molecular Basis of G Protein–Coupled Receptor
1321 Activation. *Annu. Rev. Biochem.* **87**, 897–919 (2018).

1322 68. Sadler, F. *et al.* Autoregulation of GPCR signalling through the third intracellular
1323 loop. *Nature* **615**, 734–741 (2023).
1324 69. Li, X., Zhou, M., Huang, W. & Yang, H. N-glycosylation of the β2 adrenergic
1325 receptor regulates receptor function by modulating dimerization. *FEBS J.* **284**, 2004–
1326 2018 (2017).
1327 70. Chen, S. *et al.* Mutation of a Single TMVI Residue, Phe282, in the β2-Adrenergic
1328 Receptor Results in Structurally Distinct Activated Receptor Conformations.
1329 *Biochemistry* **41**, 6045–6053 (2002).
1330 71. Liu, X. *et al.* An allosteric modulator binds to a conformational hub in the β2
1331 adrenergic receptor. *Nat. Chem. Biol.* **16**, 749–755 (2020).
1332 72. Liu, X. *et al.* Mechanism of β2AR regulation by an intracellular positive allosteric
1333 modulator. *Science* **364**, 1283–1287 (2019).
1334 73. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An open source platform
1335 for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).
1336 74. Chen, G. *et al.* Activation and allosteric regulation of the orphan GPR88-Gi1
1337 signaling complex. *Nat. Commun.* **13**, 2375 (2022).
1338 75. Onaran, H. O. *et al.* Systematic errors in detecting biased agonism: Analysis of
1339 current methods and development of a new model-free approach. *Sci. Rep.* **7**,
1340 44247 (2017).
1341 76. Beltran, A., Faure, A. J. & Lehner, B. The allosteric landscape of the Src kinase.
1342 2024.04.26.591297 Preprint at https://doi.org/10.1101/2024.04.26.591297 (2024).
1343 77. Sanchez, M. I. & Ting, A. Y. Directed evolution improves the catalytic efficiency
1344 of TEV protease. *Nat. Methods* **17**, 167–174 (2020).

1345 78. Kim, M. W. *et al.* Time-gated detection of protein-protein interactions with
1346 transcriptional readout. *eLife* **6**, e30233 (2017).
1347 79. Carpenter, B. & Tate, C. G. Engineering a minimal G protein to facilitate
1348 crystallisation of G protein-coupled receptors in their active conformation. *Protein*
1349 *Eng. Des. Sel.* **29**, 583–594 (2016).
1350 80. Crawford, K. H. d & Bloom, J. D. alignparse: A Python package for parsing
1351 complex features from high-throughput long-read sequencing. *J. Open Source*
1352 *Softw.* **4**, 1915 (2019).
1353 81. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile
1354 open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).
1355 82. Martin, M. Cutadapt removes adapter sequences from high-throughput
1356 sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
1357 83. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling
1358 with Python. *Proc. 9th Python Sci. Conf.* 92–96 (2010) doi:10.25080/Majora-
1359 92bf1922-011.
1360 84. Loes, A. N. *et al.* High-throughput sequencing-based neutralization assay
1361 reveals how repeated vaccinations impact titers to recent human H1N1 influenza
1362 strains. *J. Virol.* **98**, e00689-24 (2024).
1363

Figure 1

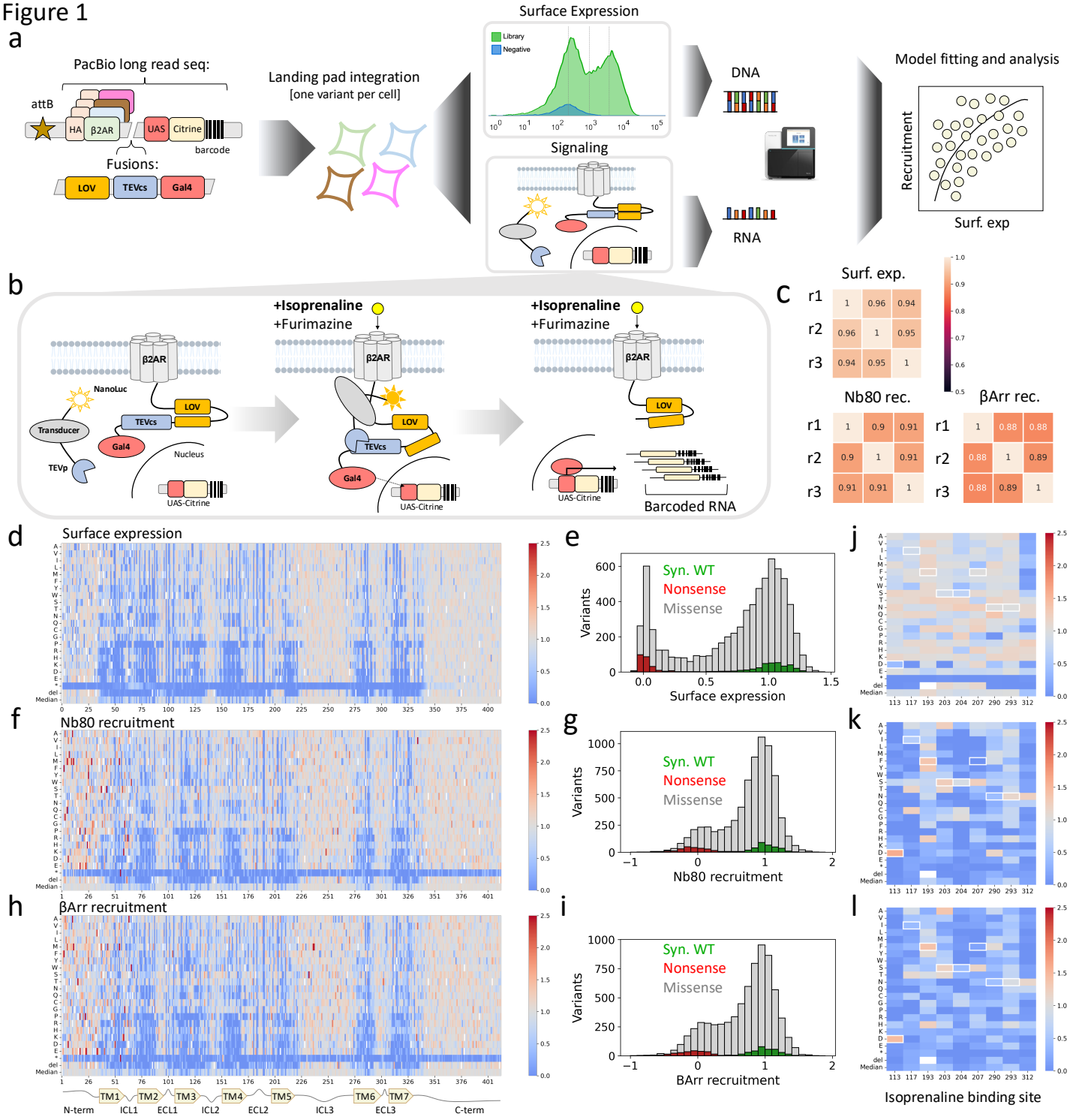

a
PacBio long read seq:
attB
HA
β2AR
UAS
Citrine
barcode
Fusions:
LOV
TEVcs
Gal4
Landing pad integration
[one variant per cell]
Surface Expression
Library
Negative
Signaling
DNA
RNA
Model fitting and analysis
Recruitment
Surf. exp
b
+Isoprenaline
+Furimazine
β2AR
NanoLuc
Transducer
TEVp
LOV
TEVcs
Gal4
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 rec.
r1 1 0.9 0.91
r2 0.9 1 0.91
r3 0.91 0.91 1
βArr rec.
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 recruitment
g
Nb80 recruitment
k
h
βArr recruitment
TM1
TM2
TM3
TM4
TM5
TM6
TM7
N-term
ICL1
ECL1
ICL2
ECL2
ICL3
ECL3
C-term
i
BArr recruitment
l
Isoprenaline binding site

Figure 2

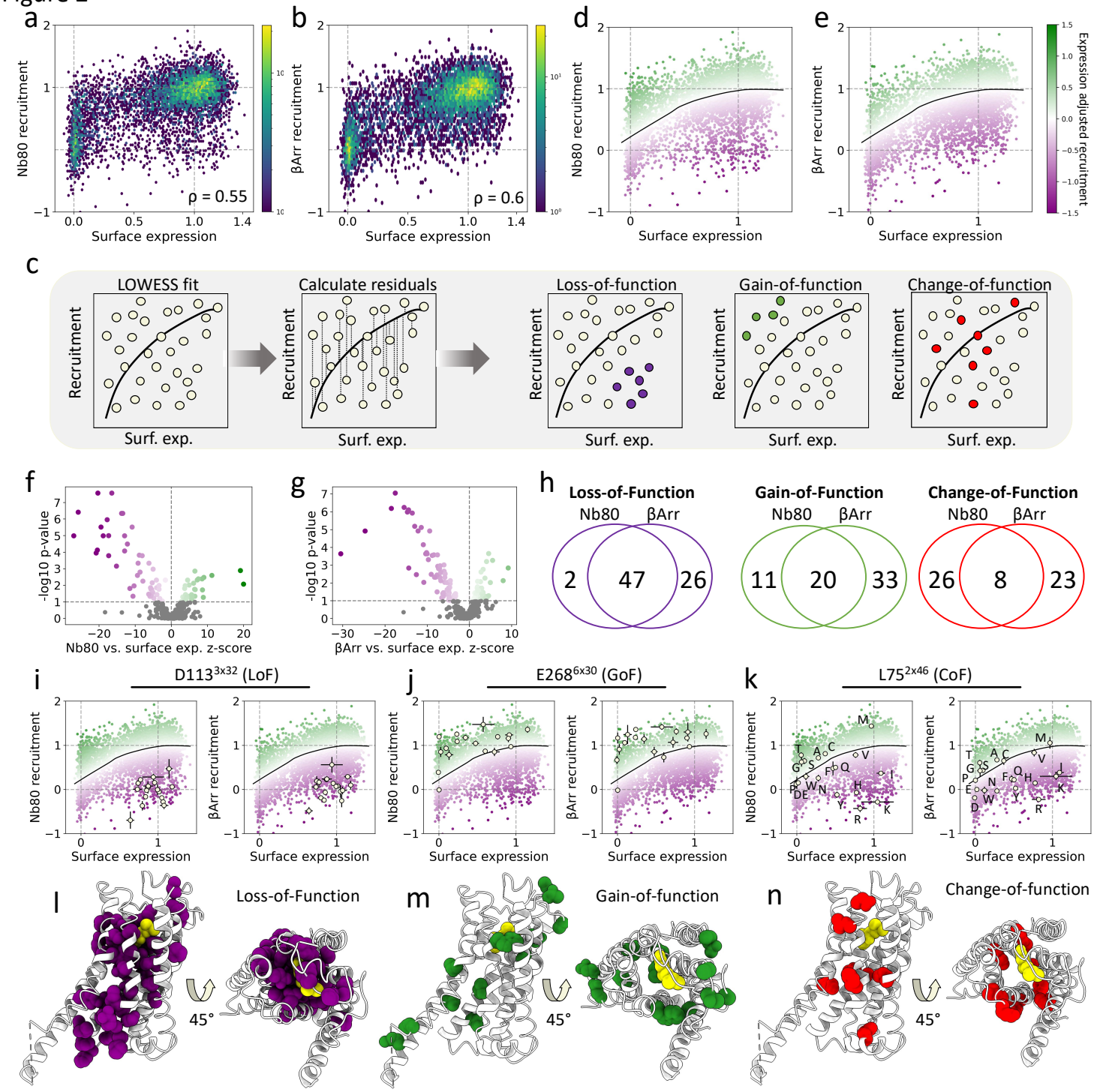

a
Nb80 recruitment
Surface expression
ρ = 0.55
b
βArr recruitment
Surface expression
ρ = 0.6
d
Nb80 recruitment
Surface expression
e
βArr recruitment
Surface expression
Expression adjusted recruitment
c
LOWESS fit
Calculate residuals
Loss-of-function
Gain-of-function
Change-of-function
Recruitment
Surf. exp.
f
-log10 p-value
Nb80 vs. surface exp. z-score
g
βArr vs. surface exp. z-score
h
Loss-of-Function
Gain-of-Function
Change-of-Function
Nb80
βArr
2
47
26
11
20
33
26
8
23
i
D113^3x32 (LoF)
j
E268^6x30 (GoF)
k
L75^2x46 (CoF)
l
Loss-of-Function
45°
m
Gain-of-function
45°
n
Change-of-function
45°

Figure 3

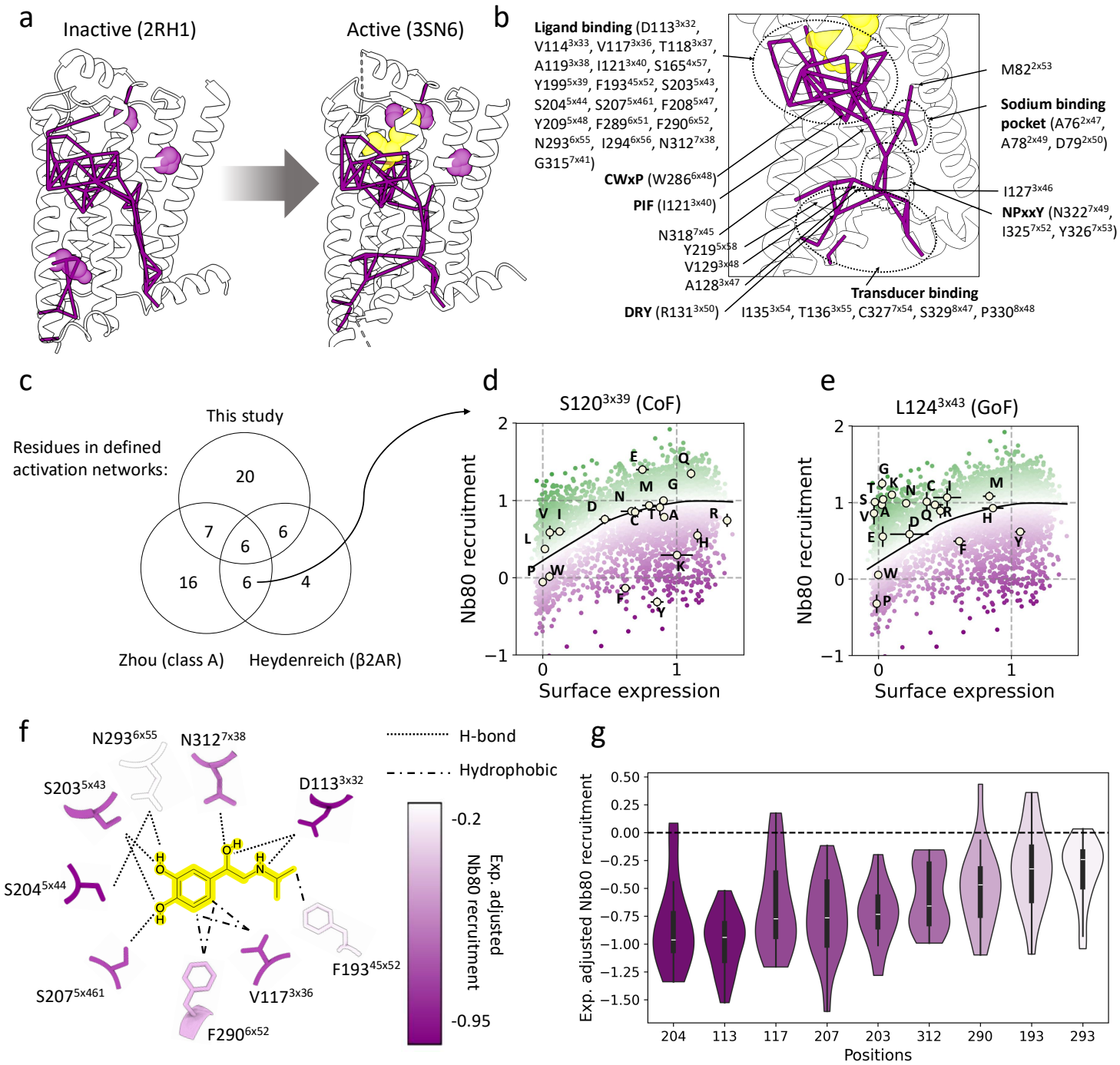

a
Inactive (2RH1)
Active (3SN6)
b
Ligand binding (D113^{3x32}, V114^{3x33}, V117^{3x36}, T118^{3x37}, A119^{3x38}, I121^{3x40}, S165^{4x57}, Y199^{5x39}, F193^{45x52}, S203^{5x43}, S204^{5x44}, S207^{5x461}, F208^{5x47}, Y209^{5x48}, F289^{6x51}, F290^{6x52}, N293^{6x55}, I294^{6x56}, N312^{7x38}, G315^{7x41})
M82^{2x53}
Sodium binding pocket (A76^{2x47}, A78^{2x49}, D79^{2x50})
CWxP (W286^{6x48})
PIF (I121^{3x40})
I127^{3x46}
NPxxY (N322^{7x49}, I325^{7x52}, Y326^{7x53})
N318^{7x45}
Y219^{5x58}
V129^{3x48}
A128^{3x47}
Transducer binding
DRY (R131^{3x50})
I135^{3x54}, T136^{3x55}, C327^{7x54}, S329^{8x47}, P330^{8x48}
c
Residues in defined activation networks:
This study
20
7
6
6
16
6
4
Zhou (class A)
Heydenreich (β2AR)
d
S120^{3x39} (CoF)
Nb80 recruitment
Surface expression
e
L124^{3x43} (GoF)
Nb80 recruitment
Surface expression
f
N293^{6x55}
N312^{7x38}
S203^{5x43}
D113^{3x32}
S204^{5x44}
S207^{5x461}
V117^{3x36}
F193^{45x52}
F290^{6x52}
H-bond
Hydrophobic
-0.2
-0.95
Exp. adjusted Nb80 recruitment
g
Exp. adjusted Nb80 recruitment
Positions
204
113
117
207
203
312
290
193
293

Figure 4

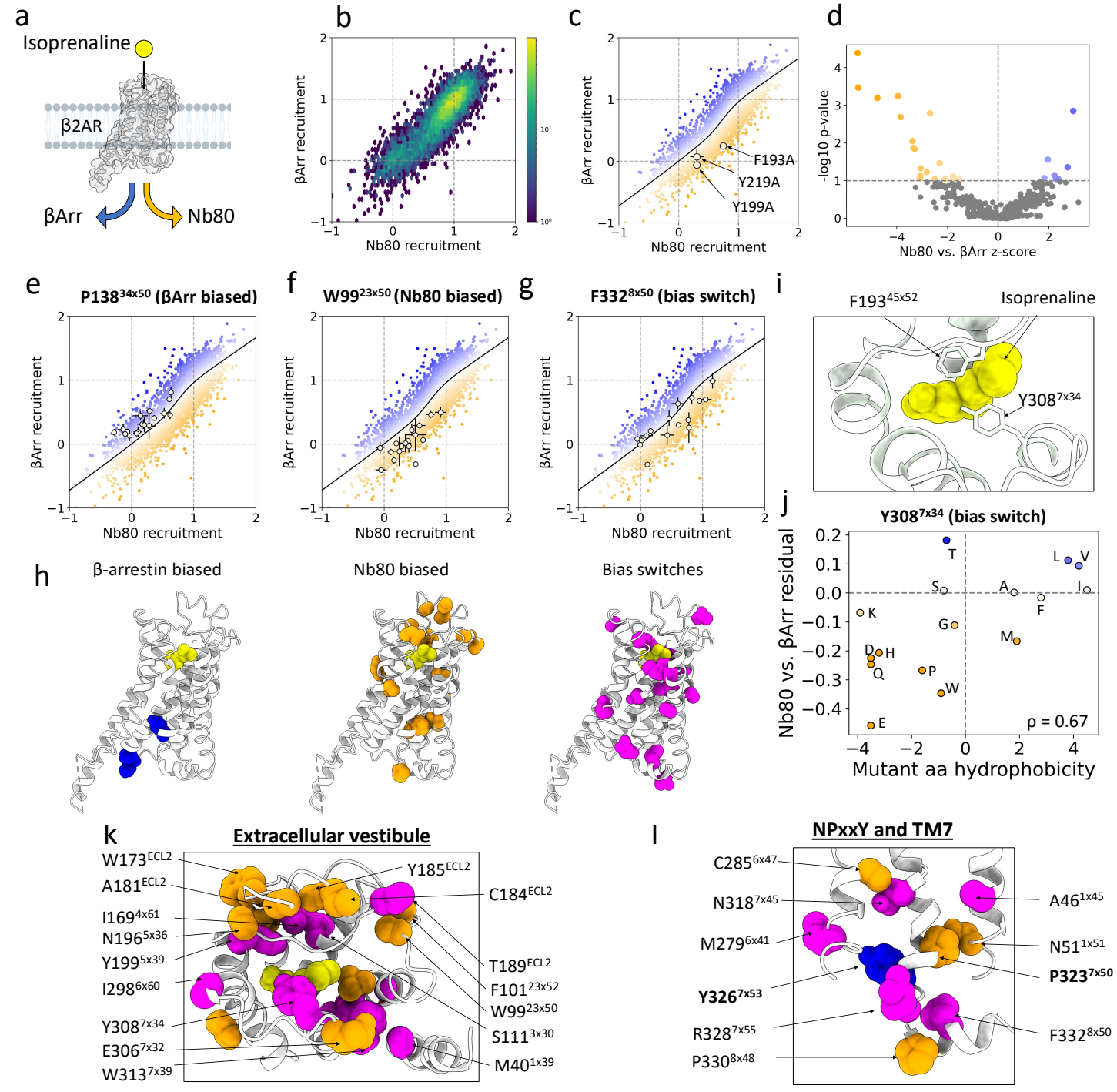

a
Isoprenaline
β2AR
βArr
Nb80
b
βArr recruitment
Nb80 recruitment
c
βArr recruitment
Nb80 recruitment
F193A
Y219A
Y199A
d
-log10 p-value
Nb80 vs. βArr z-score
e
P138^34x50 (βArr biased)
βArr recruitment
Nb80 recruitment
f
W99^23x50 (Nb80 biased)
βArr recruitment
Nb80 recruitment
g
F332^8x50 (bias switch)
βArr recruitment
Nb80 recruitment
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

Figure 5

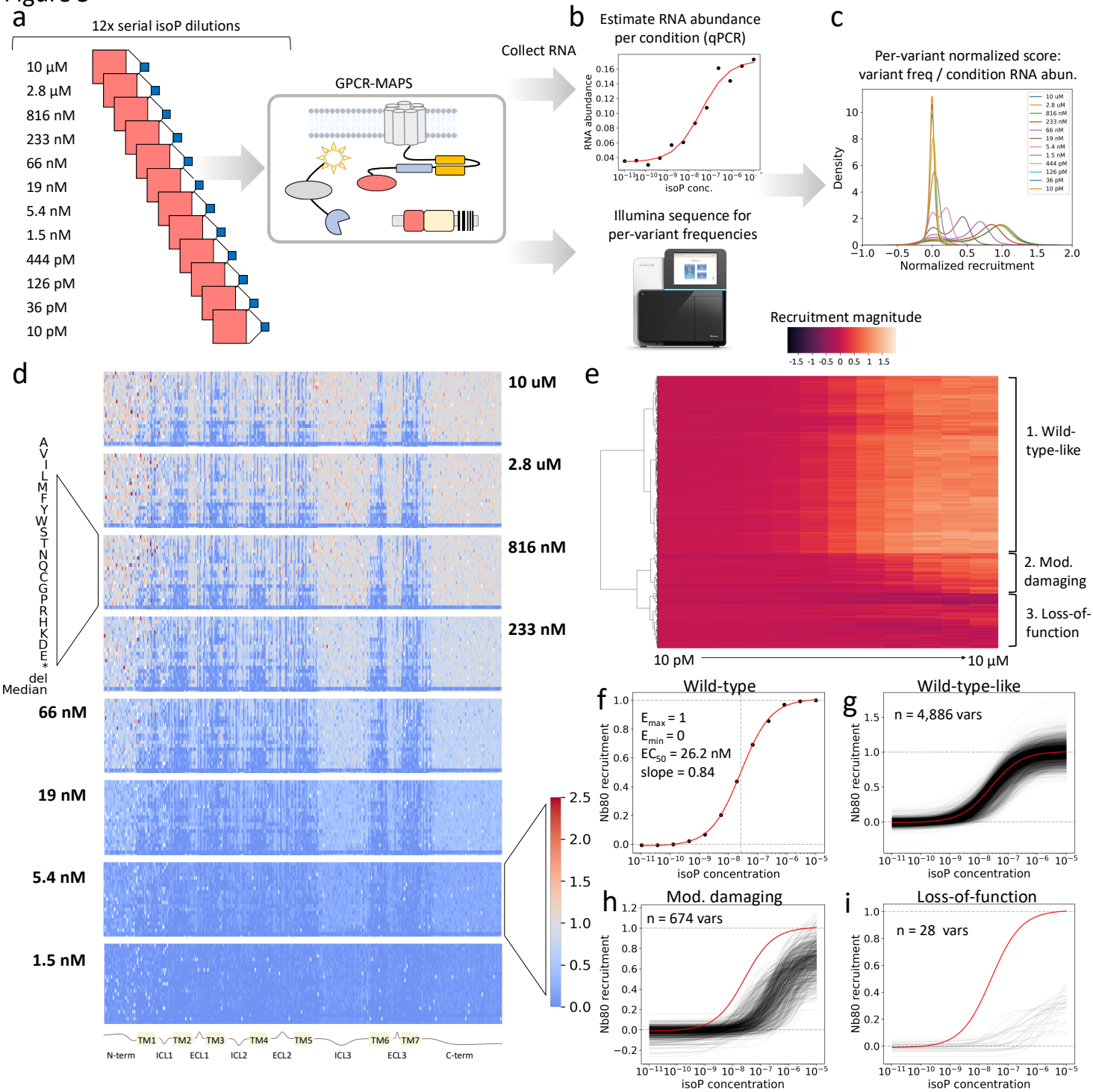

a
12x serial isoP dilutions
10 µM
2.8 µM
816 nM
233 nM
66 nM
19 nM
5.4 nM
1.5 nM
444 pM
126 pM
36 pM
10 pM
GPCR-MAPS
Collect RNA
b
Estimate RNA abundance per condition (qPCR)
RNA abundance
isoP conc.
Illumina sequence for per-variant frequencies
c
Per-variant normalized score: variant freq / condition RNA abun.
Density
Normalized recruitment
Recruitment 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
Nb80 recruitment
isoP concentration
g
Wild-type-like
n = 4,886 vars
h
Mod. damaging
n = 674 vars
i
Loss-of-function
n = 28 vars

Figure 6

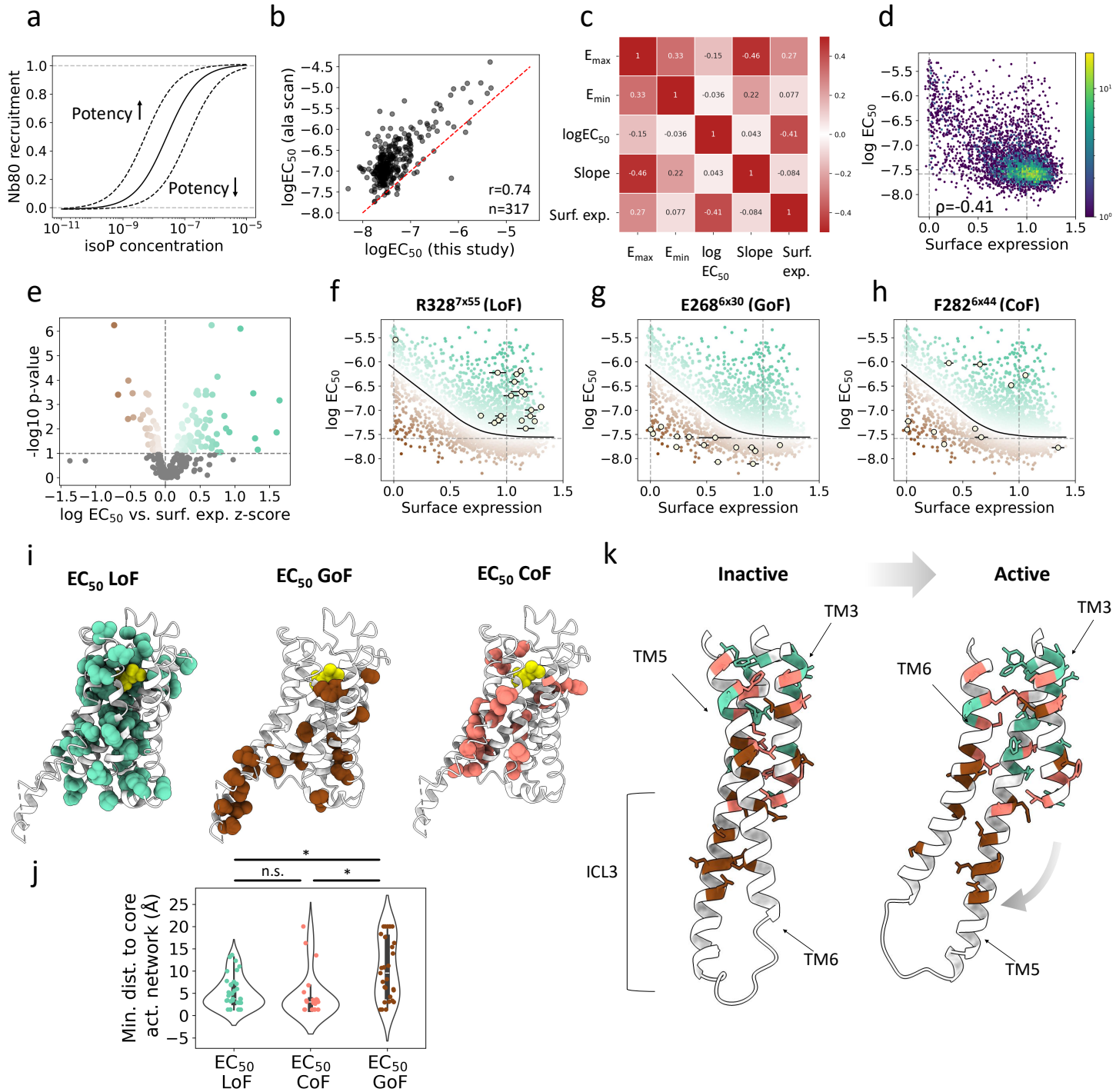

a
Nb80 recruitment
Potency
Potency
isoP concentration
b
logEC50 (ala scan)
r=0.74
n=317
logEC50 (this study)
c
Emax
Emin
logEC50
Slope
Surf. exp.
d
log EC50
ρ=-0.41
Surface expression
e
-log10 p-value
log EC50 vs. surf. exp. z-score
f
R328 7x55 (LoF)
log EC50
Surface expression
g
E268 6x30 (GoF)
log EC50
Surface expression
h
F282 6x44 (CoF)
log EC50
Surface expression
i
EC50 LoF
EC50 GoF
EC50 CoF
j
n.s.
*
*
Min. dist. to core act. network (Å)
EC50 LoF
EC50 CoF
EC50 GoF
k
Inactive
Active
TM3
TM5
TM6
ICL3
TM3
TM6
TM5

Figure 7

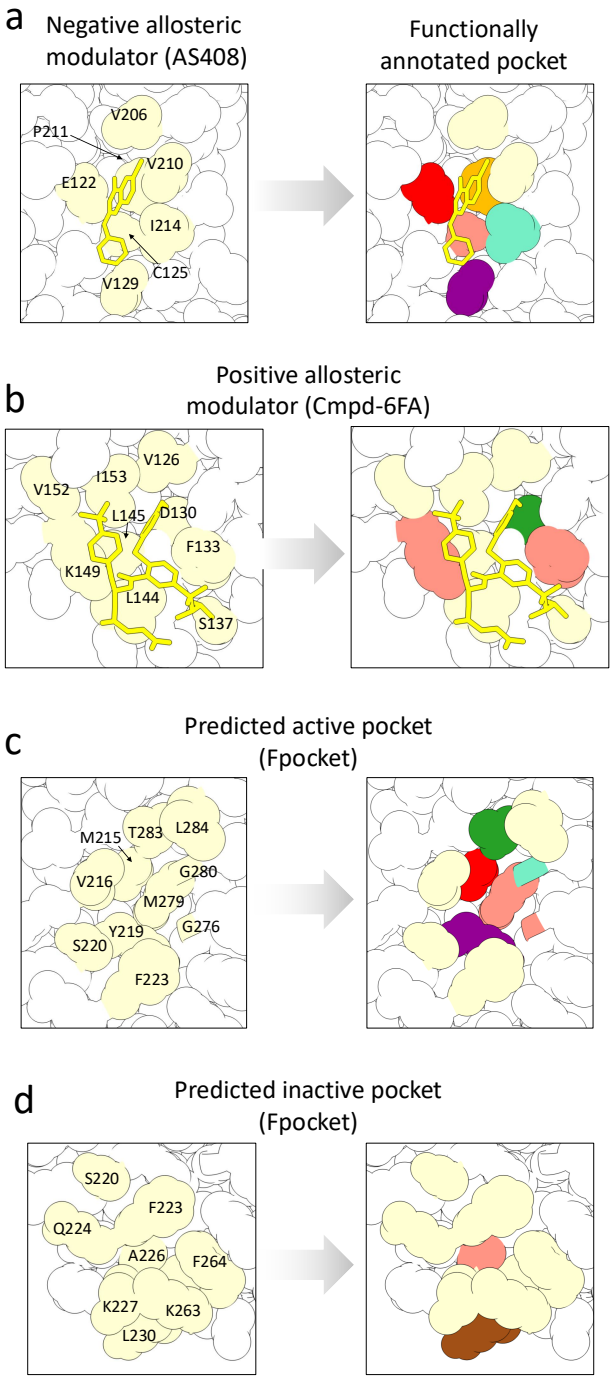
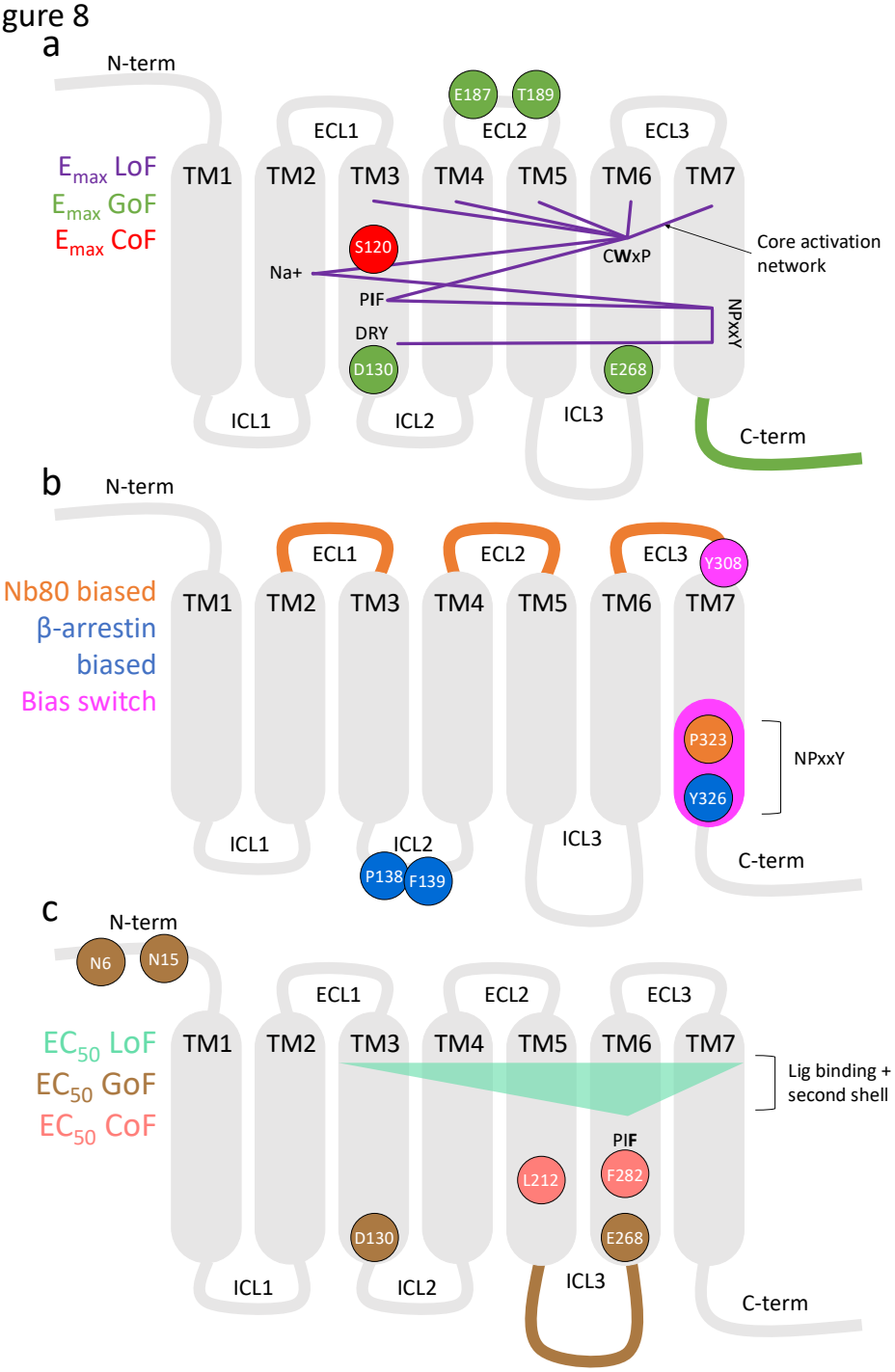

a
Negative allosteric modulator (AS408)
Functionally annotated pocket
P211
V206
V210
E122
I214
C125
V129
b
Positive allosteric modulator (Cmpd-6FA)
V126
I153
V152
L145
D130
F133
K149
L144
S137
c
Predicted active pocket (Fpocket)
M215
T283
L284
V216
G280
M279
G276
Y219
S220
F223
d
Predicted inactive pocket (Fpocket)
S220
F223
Q224
A226
F264
K227
K263
L230

Figure 8


a
N-term
E187
T189
ECL1
ECL2
ECL3
E_{max} LoF
E_{max} GoF
E_{max} 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
EC_{50} LoF
EC_{50} GoF
EC_{50} CoF
TM1
TM2
TM3
TM4
TM5
TM6
TM7
Lig binding + second shell
PIF
L212
F282
D130
E268
ICL1
ICL2
ICL3
C-term