bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Structural and Functional Characterization of G Protein-Coupled Receptors with Deep Mutational Scanning

Authors: Eric M. Jones[1†‡], Nathan B. Lubock[1†‡], AJ Venkatakrishnan[2,3], Jeffrey Wang[1], Alex M. Tseng[3], Joseph M. Paggi[3], Naomi R. Latorraca[3], Daniel Cancilla[1], Megan Satyadi[1], Jessica E. Davis[1], M. Madan Babu[2], Ron O. Dror[3*], Sriram Kosuri[1*]

Affiliations:
[1]Department of Chemistry and Biochemistry, UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, Quantitative and Computational Biology Institute, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, and Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA 90095, USA
[2]MRC Laboratory of Molecular Biology, Francis Crick Ave, Cambridge CB2 0QH, UK
[3]Department of Computer Science, Stanford University, Stanford, CA 94305, Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA 94305, Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA
Current Address:
‡Octant, Inc. 570 Westwood Plaza, Los Angeles, CA 90095

*Correspondence should be addressed to S.K. (sri@ucla.edu) and R.O.D. (ron.dror@stanford.edu)
†Equal Contribution

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

In humans, the 813 G protein-coupled receptors (GPCRs) are responsible for transducing diverse chemical stimuli to alter cell state, and are the largest class of drug targets. Their myriad structural conformations and various modes of signaling make it challenging to understand their structure and function. Here we developed a platform to characterize large libraries of GPCR variants in human cell lines with a barcoded transcriptional reporter of G-protein signal transduction. We tested 7,800 of 7,828 possible single amino acid substitutions to the beta-2 adrenergic receptor (β_2AR) at four concentrations of the agonist isoproterenol. We identified residues specifically important for β_2AR signaling, mutations in the human population that are potentially loss of function, and residues that modulate basal activity. Using unsupervised learning, we resolve residues critical for signaling, including all major structural motifs and molecular interfaces. We also find a previously uncharacterized structural latch spanning the first two extracellular loops that is highly conserved across Class A GPCRs and is conformationally rigid in both the inactive and active states of the receptor. More broadly, by linking deep mutational scanning with engineered transcriptional reporters, we establish a generalizable method for exploring pharmacogenomics, structure and function across broad classes of drug receptors.

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Main Text

Introduction

G protein-coupled receptors (GPCRs) are central mediators of mammalian cells' ability to sense and respond to their environment. The 813 human GPCRs respond to a wide range of chemical stimuli such as hormones, odors, natural products, and drugs by modulating a small set of defined pathways that affect cellular physiology (*1*). Their central role in altering relevant cell states makes them ideal targets for therapeutic intervention, with ~34% of all U.S. Food and Drug Administration (FDA)-approved drugs targeting the GPCR superfamily (*2*).

Understanding GPCR signal transduction is non-trivial for several reasons. First, GPCRs exist in a complex conformational landscape, making traditional biochemical and biophysical characterization difficult (*3*, *4*). Consequently, most experimentally-determined GPCR structures are truncated, non-native, or artificially stabilized (*5*). Even when structures exist, the majority are of inactive states - GPCR conformations that cannot couple with a G protein and cause it to stimulate intracellular signaling. Second, the function of a GPCR depends on its ability to change shape. Static structures from both X-ray crystallography and cryo electron microscopy do not directly probe structural dynamics (*6*). Tools such as double electron-electron resonance (DEER) spectroscopy, nuclear magnetic resonance (NMR) spectroscopy, and computational simulation have aided our understanding of GPCR dynamics, but interpreting how structural dynamics relate to function is still difficult (*7*, *8*).

Structure- and dynamics-based analyses generate sets of candidate residues that are potentially critical for function and warrant further characterization. These approaches are complemented by methods that directly perturb protein function such as mutagenesis followed by functional screening. Several reporter gene and protein complementation assays measure GPCR signal transduction by activation of a transcriptional reporter, and are often used to identify and validate important structural residues (*9*–*11*). Such transcriptional reporter assays exist for most major drug receptor classes, including the major GPCR pathways: $G_{\alpha s}$, $G_{\alpha q}$, $G_{\alpha i/o}$, and arrestin signaling (*12*–*14*).

Recent advances in DNA synthesis, genome editing, and next-generation sequencing have enabled deep mutational scanning (DMS) approaches that functionally assay all possible missense mutants of a given protein (*15*, *16*). Several new methods allow for the generation and screening of DMS libraries in human cell lines (*17*–*20*). Function is usually assessed by next-generation sequencing using screens that are bespoke to each gene's function, or by more general approaches that allow characterization of expression levels rather than function (*21*). For GPCRs, the DMS of the CXCR4, CCR5, and T1R2 GPCRs used binding to external epitopes to test expression and ligand binding (*22*, *23*). Unfortunately, such assays tell us little

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

about the signaling capacity of these mutants, which is the primary function of GPCRs and many other drug receptors.

Here we develop an experimental approach to simultaneously profile variant libraries with barcoded transcriptional reporters in human cell lines using RNA-seq. This method is widely applicable to GPCRs and across the druggable genome where transcriptional reporters exist. As a proof-of-principle, we perform DMS on a prototypical GPCR, the β_2-adrenergic receptor (β_2AR) and measure the consequences of these mutations through the cyclic AMP (cAMP) dependent pathway.

Results

Multiplexed screening platform for G_s-coupled GPCR signaling

We developed a system to build, stably express, and assay individual variants of the β_2AR in human cell lines. The β_2AR primarily signals through the heterotrimeric G_s protein, activating adenylyl cyclase upon agonist binding. In our platform, cAMP production stimulates transcription of a barcoded reporter gene, controlled by multimerized cAMP response elements (CRE), which can be quantified by RNA-seq (Fig. 1A). Initially, we generated a HEK293T-derived cell line for stable integration of the GPCR-reporter construct (Fig. 1B, S1A and S1B). We also modified a previously developed Bxb1-landing pad system to allow for stable, once-only integration at the transcriptionally-silent H11 safe-harbor locus to avoid placing the reporter within transcribed genes (*24–26*). To prevent endogenous signaling, we knocked out the gene encoding for β_2AR, *ADRB2,* and verified loss of reporter gene activity in response to the β_2AR agonist, isoproterenol (Fig. S1C). Our donor vector configuration ensures the receptor and resistance marker are only activated upon successful integration into the landing pad. This vector architecture ameliorates toxicity related to receptor overexpression during library integration (Fig. 1B). Lastly, we included several sequence elements in the donor vector to improve signal-to-noise of the assay: an insulator upstream of the CRE reporter and an N-terminal affinity tag to the receptor (Fig. S1D and S1E). As a result, upon integration of a donor vector expressing wild-type (WT) β_2AR, isoproterenol induces reporter gene expression in a dose-dependent manner (Fig. 1B).

We designed and synthesized the receptor's 7,828 possible missense variants in eight segments on oligonucleotide microarrays (Fig. 1C). We amplified the mutant oligos, attaching a random 15 nucleotide barcode sequence, and cloned them into one of eight background vectors encoding the upstream, wild-type portion of the gene. In this configuration, we mapped barcode-variant pairs with next-generation sequencing and subsequently utilized Type IIS restriction enzymes to insert the remaining sequence elements between the receptor and barcode. In the resulting mature donor vector, the barcode is located in the 3' untranslated region (UTR) of the reporter gene. We integrated the library into our engineered cell line, and developed protocols to ensure proper quantification of library members, most notably vastly

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

increasing the numbers of cells we assayed and RNA processed for the RNA-seq (Fig. S1F and S2A).

Measurement of mutant activities and comparison to evolutionary metrics

We screened the mutant library at four concentrations of the β_2AR full-agonist isoproterenol: vehicle control, an empirically determined half-maximal activity (EC_{50}), full activity (EC_{100}), and beyond saturation of the WT receptor (E_{max}). We obtained measurements for 99.6% (7,800/7,828) of possible missense variants (412 residues * 19 amino acids = 7,828 possible missense variants) with two biological replicates at each condition (Fig. 1C). We normalized these measurements against forskolin treatment, which induces cAMP signaling independent of the β_2AR. Forskolin treatment maximally induces the reporter gene, therefore the relative barcode expression is proportional to the physical composition of the library. Each variant was represented by 10 barcodes (median), with biological replicates displaying Pearson's correlations of 0.868 to 0.897 at the barcode level and 0.655 to 0.750 when summarized by individual variants.

The heatmap representation of the variant-activity landscape reveals global and regional trends in response to specific subtypes of mutations (Fig. 2A). For example, the transmembrane domain and intracellular helix 8 are more sensitive to substitution than the termini or loops, and this effect becomes more pronounced at higher agonist concentrations (all $p < 0.001$; Mann-Whitney U). In general, the transmembrane domain and intracellular helix 8 are especially sensitive to helix-disrupting proline substitutions (Figure S2B). Microarray-derived DNA often contains single-base deletions that will introduce frameshift mutations into our library (*27*). As expected, frameshifts consistently display lower activity than missense mutations (Fig. 2B; $p << 0.0001$; Mann-Whitney U). Furthermore, frameshifts occurring in the C-terminus have a significantly increased activity relative to those in the rest of the protein (Figure 2C; $p << 0.0001$; Mann-Whitney U). We also built and integrated six previously characterized mutants (*28–30*) into our system individually and measured activity with a luciferase reporter gene at the same induction conditions, which largely agreed with the results of our multiplexed assay (Fig. 2C).

Metrics for sequence conservation and covariation are often used to predict the effects a mutation will have on protein function (*31–33*). Mutational tolerance, the mean activity of all amino acid substitutions per residue at each agonist concentration, is highly correlated to conservation, both across species for the β_2AR (Spearman's $\rho = -0.743$), and across all Class A GPCRs (Spearman's $\rho = -0.676$; Fig. 3A, S3A, and S3B) (*31*, *33*, *34*). Correlation between our data and both predictors increases with agonist concentration up to EC_{100} (Fig. S4). Next, we examined residues more intolerant to mutation than expected given their conservation across Class A GPCRs. We found a subset of residues in extracellular loop 2 (ECL2), including C184 and C190 that form an intraloop disulfide bridge, suggesting a fairly specific functional role for this motif in the β_2AR (Fig. 3A, S3A and S3B). On an individual variant level, mutational responses correlate (Spearman's $\rho = 0.521$) with EVmutation, a predictor of mutational effects from sequence covariation (Fig. 3B and S3C) (*31*, *33*, *34*).

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Population genetics and structural analysis of individual variants

In addition to evolutionary metrics, understanding the functional distribution of *ADRB2* variants found within the human population is important given the extensive variation found among GPCR drug targets (*35*). The Genome Aggregation Database (gnomAD) reports variants found across 141,456 individuals (*36*), and the majority of the 180 *ADRB2* missense variants are of unknown significance. We classified 11 of these variants as potentially loss of function, by comparing their activity to the distribution of frameshift mutations found in our assay (Fig. 3C; see Methods). Given that measurements of individual mutations are noisy (average coefficient of variation = 0.55), this analysis is best suited as a funnel to guide further characterization (see Discussion).

However, our analysis is more robust in aggregate. Therefore, we compiled a list of the 100 most activating mutations at vehicle control and the 100 least active mutations at EC_{100} and mapped their location on the β_2AR structure. As expected, the least active mutations tended to reside within the core of the transmembrane domain (Fig. S3D and S3E). Alternatively, the most activating mutations mapped to TM1, TM5, TM6, and residues that typically face away from the internal core of the receptor (Fig. 3D and 3E). Of note, a group of these mutations in TM5 face TM6, which undergoes a large structural rearrangement during receptor activation (*37*). Activating mutants are also enriched in the termini, ICL3, and Helix 8. Concentration at the termini is unsurprising, as these regions have known involvement in surface expression (*38*). Similarly, the enrichment of activating mutants in ICL3 appears to reflect its role in G protein binding (*39–41*). Lastly, we observe a number of activating mutations in the terminal residue, L413. A recent study of genetic variation in human melanocortin 4 (*MC4R)* GPCR also found a gain-of-function mutation at the terminal residue of the receptor, suggesting a possible conserved role for this position in regulating basal activity of GPCRs (*42*).

Unsupervised learning reveals functionally relevant groupings of residues

Given that our data spans thousands of mutations across several treatment conditions, we used unsupervised learning methods to reveal hidden regularities within groups of residues' response to mutation. In particular, we applied Uniform Manifold Approximation and Projection (UMAP) (*43*) to learn multiple different lower-dimensional representations of our data and clustered the output with density-based hierarchical clustering (HDBSCAN; Fig. S4) (*44*). We found residues consistently separated into 6 clusters that exhibit distinct responses to mutation (Fig. 4A and 4B). Clusters 1 and 2 are globally intolerant to all substitutions, whereas Cluster 3 is vulnerable to proline and charged substitutions. Cluster 4 is particularly inhibited by negatively charged substitutions and Cluster 5 by proline substitutions, while Cluster 6 is unaffected by any mutation. Mapping these clusters onto a 2D snake plot representation shows Clusters 1–5 primarily comprise the transmembrane domain, while Cluster 6 resides in the loops and termini (Fig. 4C). These flexible regions are often truncated before crystal structure determination to minimize conformational variability (*45*). Surprisingly, a number of residues from Cluster 5 also map there, suggesting potential structured regions. However, Cluster 5 assignment is largely

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

based on the response of a single proline mutation, and thus is more susceptible to noise than the other clusters (see Discussion).

Next, we projected the clusters onto the hydroxybenzyl isoproterenol-bound structure (Fig. S5A; PDB: 4LDL). The globally intolerant Clusters 1 and 2 segregate to the core of the protein, while the charge-sensitive Cluster 3 is enriched in the lipid-facing portion. This suggests that differential patterns of response to hydrophobic and charged substitutions could correlate with side chain orientation within the transmembrane domain. Indeed, residues that are uniquely charge sensitive are significantly more lipid-facing than those that are sensitive to both hydrophobic and charged mutations (Fig. 5A, S5B-S5D, $p = 0.000036$; Mann-Whitney U) (*46*).

Mutational tolerance stratifies the functional relevance of structural features

Decades of research have revealed how ligand binding is coupled to G-protein activation through a series of conserved motifs (*37*). This comprehensive, unbiased screen enables us to systematically evaluate and rank the functional importance of every implicated residue. The globally intolerant UMAP clusters (1 and 2) highlight many residues from these motifs and suggest novel residues for investigation (Fig. 5B). We can further resolve the significance of individual residues within these motifs by ranking the mutational tolerance of positions in these clusters at EC_{100} (Fig. 5C). In fact, 11 of the 15 most mutationally intolerant positions belong to the PIF, CWxP, and NPxxY motif, orthosteric site, a water-mediated bond network, an extracellular disulfide bond, and a cholesterol binding site. Interestingly, the second most intolerant residue is the uncharacterized $G315^{7x41}$ (GPCRdb numbering in superscript (*5*)). In the active state, G315's alpha carbon points directly at $W286^{6x48}$ of the CWxP motif, the fourth most intolerant residue, and any substitution at $G315^{7.x41}$ will likely clash with $W286^{6x48}$ (Fig. 5D). We confirmed G315's intolerance with a luciferase reporter gene assay, where mutants G315T and G315L resulted in complete loss of function (Fig. S6A).

Recent simulations suggest water-mediated hydrogen bond networks play a critical role in GPCR function (*47*, *48*). The third most intolerant residue in our assay, $Y326^{7x53}$ of the NPxxY motif, is especially important as it switches between two of these networks during receptor activation. In the inactive state, $Y326^{7x53}$ contacts $N51^{1x50}$ and $D79^{2x50}$, two of the top 15 most intolerant positions (Fig. 5E). N51L and N51Y also result in complete loss of function when assayed individually (Fig. S6A). The movement of $Y326^{7x53}$ is also part of a broader rearrangement of residue contacts that are conserved across Class A GPCRs, with the majority of these residues being intolerant to mutation (Fig. S6B) (*49*). Aside from $G315^{7x41}$, the other uncharacterized residues in the top 15 include $W99^{23x50}$, $S319^{7x46}$, and $G83^{2x54}$. Given the correlation between mutational tolerance and functional relevance, further investigation of these residues will likely reveal insights into GPCR biology.

Next, we hypothesized residues in the orthosteric site that directly contact isoproterenol would respond uniquely to mutation, however no crystal structure of β_2AR bound to isoproterenol exists. Using the crystal structure of the β_2AR bound to the analog, hydroxybenzyl isoproterenol (PDB: 4LDL), we find that residues responsible for binding the derivatized hydroxybenzyl tail

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

have significantly higher mutational tolerance than residues that contact the catecholamine head common to both isoproterenol and hydroxybenzyl isoproterenol at EC_{100} ($p = 0.0162$; Fig. 5F, Fig. S6C). Given this discrimination, we believe DMS can be a powerful tool for mapping functional ligand-receptor contacts in GPCRs.

GPCR signaling is dependent on a series of intermolecular interactions, and the numerous β_2AR crystal structures enable us to comprehensively evaluate residues mediating such interactions. For example, cholesterol is an important modulator of GPCR function (*50*), and the timolol-bound inactive-state β_2AR structure elucidated the location of a cholesterol binding site (PDB: 3D4S) (*51*). Of residues in this pocket, W158^{4x50} is predicted to be most important for cholesterol binding, and in agreement, W158^{4x50} is the most mutationally intolerant (Fig. S6D). Similarly, a number of studies have mutagenized residues at the $G_{\alpha s}$-β_2AR interface (*11*, *52–58*), but a complete understanding of the relative contribution of each residue to maintaining the interface is unknown. Most residues are more mutationally tolerant than residues in the intolerant clusters 1 and 2, but the four most intolerant positions are I135^{3x54}, V222^{5x61}, A271^{6x33}, and Q229^{5x68} respectively (Fig. S6E). Q229^{5x68} appears to coordinate polar interactions between D381 and R385 of the α5 helix of $G_{\alpha s}$, whereas V222^{5x61} and I135^{3x54} form a hydrophobic pocket on the receptor surface (Fig. S6F).

A structural latch is conserved across Class A GPCRs

Analysis of the mutational tolerance data has highlighted the functional importance of a previously uncharacterized residues. In particular, W99^{23x50} of extracellular loop 1 (ECL1) is the 13th most intolerant residue, which is unusual as mutationally intolerant residues are rare in the flexible loops. Furthermore, W99^{23x50} is proximal to the disulfide bond C106^{3x25}-C191^{45x50}, an important motif for stabilization of the the receptor's active state (*59*, *60*). While, aromatic residues are known to facilitate disulfide bond formation, only tryptophan is tolerated at this position (*61*). We hypothesize W99's indole group hydrogen bonds with the backbone carbonyl of the neighboring uncharacterized and mutationally intolerant G102^{3x21}, positioning W99^{23x50} towards the disulfide bond. Other aromatic residues are unable to form this hydrogen bond and are less likely to be positioned properly. G102^{3x21} also hydrogen bonds with the backbone amide of C106^{3x25}, further stabilizing this region. To verify this claim, we individually confirmed the mutational intolerance of both W99^{23x50} and G102^{3x21} (Fig. S7A).

Interestingly, W99^{23x50}, G102^{3x21}, and C106^{3x25} are almost universally conserved across Class A GPCRs (Fig. 6A and S7B). Comparison of over 25 high-resolution structures of class A GPCRs from five functionally different sub-families and six different species revealed that these residues consistently contact each other (Fig. 6B and 6C). Based on the evolutionary and structural conservation across Class A GPCRs, we find W99^{23x50}, G102^{3x21}, and the C106^{3x25}-C191^{45x50} disulfide bond represent a conserved WxxGxxxC motif, forming an extracellular "structural latch" that is maintained consistently throughout GPCRs spanning diverse molecular functions and phylogenetic origins. While a minority of Class A GPCRs lack the Trp/Gly combination of residues in the ECL1 region, these receptors have varying structures in ECL1: an alpha helix

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(sphingosine S1P receptor), beta strand (adenosine receptor), or even intrinsically disordered (viral chemokine receptor US28) (Fig. S7C).

To better understand the dynamics of the structural latch, we compared the active and inactive state crystal structures of four representative GPCRs. While the overall RMSD between the inactive and active states for the β2AR, M2 muscarinic receptor, μ opioid receptor and κ opioid receptor are 1 Å,1.5 Å, 1.7 Å and 1.5 Å respectively, the conformation of the latch in the active and inactive states is nearly identical in each receptor (Fig. 6D). This suggests that the extracellular structural latch is part of a larger rigid plug present at the interface of the transmembrane and extracellular regions, which could be important for the structural integrity of the receptor and possibly guide ligand entry.

In Class A receptors lacking components of the WxxGxxxC motif, introducing the Trp-Gly interaction could increase the stability of the receptor for structural studies. In fact, in the BLT1 receptor structure, a Gly mutation at 3x21 was found to be thermostabilizing (*62*). Other candidate receptors lacking a Gly at 3x21 include the alpha2B receptor and the neuropeptide FF2 receptor, where the R81G and D112G mutations respectively have potential to increase receptor stability. More broadly, these conserved ECL1/TM3 positions could serve as candidate sites for introducing thermostabilizing mutations.

Discussion

Our findings showcase a new, generalizable approach for DMS of human protein targets with transcriptional reporters. Such reporters enable precise measurements of gene-specific function that can be widely applied across the druggable genome. We show comprehensive mutagenesis can illuminate the structural organization of the protein and the local environment of individual residues. These results suggest DMS can work in concert with other techniques (e.g. X-ray crystallography, Cryo-EM, and molecular dynamics) to augment our understanding of GPCR structure-function relationships. Moreover, we identify key residues for β2AR function including uncharacterized positions that inform about receptor stability and activation. Importantly, these approaches can be undertaken when direct structural information is unavailable but reporters exist, which is true for most GPCRs.

There are still a number of limitations to our current approach that we expect will improve as we develop the method. Importantly, we did not quantify cell-surface expression directly in our high-throughput functional assay, and thus we cannot distinguish between mutations that substantially affect G protein signaling and those that affect cell-surface expression. In particular, mutations that lead to increased signal in our assays could in fact work by reducing GPCR internalization and not by increasing the intrinsic activity of the receptor. However, we express our variant library in a genomic context at a controlled copy number, dampening the effects of expression-related artifacts typically associated with assays that involve transiently transfected receptor. In addition, expression level alterations can affect the dynamics of signaling and thus may be physiologically relevant. For example, the GPCR *MC4R* is

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

haploinsufficient, and rare heterozygous mutations that eliminate or reduce receptor expression are associated with obesity (*42*, *63*, *64*). Combining our assay with new generalized, multiplexed assays of protein expression levels in human cells can help tease apart mechanistic reasons for differences in signaling (*21*). Secondly, the current signal-to-noise ratio of this approach at single-variant resolution restricted our analyses to mutations with extreme effects on receptor function. This made interpreting single mutations challenging. For example, several mutations within the C-terminus exhibited a sensitivity to proline substitution. This was surprising because the C-terminus is thought to be a flexible, disordered region (*65*, *66*). We individually synthesized and tested three of these mutations (E369P, R253P, and T360P), and found them to maintain similar activity to the WT receptor, suggesting these substitutions do not disrupt function (Fig. S6A). Thus, individual variant data should be confirmed by more traditional assays until the signal-to-noise ratio is improved. However, our measurements are robust in aggregate, and pointed to new receptor biology, providing structural and functional insights. Further improvements to the signal to noise will facilitate the exploration of more subtle aspects of individual mutations.

Looking forward, our method is well-poised to investigate many outstanding questions in GPCR and drug receptor biology. First, individual GPCRs signal through multiple pathways, including pathways mediated by various G proteins and arrestins (*67*–*70*). By leveraging transcriptional reporters for each of these pathways, we can understand the mechanisms that underpin signal transduction and biased signaling (*71*). Second, GPCRs are often targeted by synthetic molecules with either unknown or predicted binding sites, and often have no known structures. We find ligands imprint a mutational signature on their receptor contacts which could potentially reveal the binding site for allosteric ligands. We also found several regions on the external surface of the receptor where activating mutants are clustered. Since perturbations at these sites appear to increase receptor activity, they could potentially be targeted by positive allosteric modulators or allosteric agonists (*50*). Third, the identification of mutations that can stabilize specific conformations or increase receptor expression can aid in GPCR structure determination (*72*, *73*). Fourth, the development of stable cell libraries expressing human medicinally-related GPCR variants can be combined with large-scale profiling against small molecule libraries to build very large-scale empirical maps for how small molecules modulate this broad class of receptors (*74*–*76*). Finally, our approach is generalizable to many classes of drug receptors where transcriptional reporters exist or can be developed (*77*), enabling the functional profiling, structural characterization, and pharmacogenomic analysis for most major drug target classes.

Acknowledgements:

We thank the Kosuri Lab for helpful discussions, the UCLA Broad Stem Cell Research Center Sequencing and Flow Cytometry Core, and the Technology Center for Genomics and Bioinformatics for providing next-generation sequencing. We thank Deborah Marks and Jung-Eun Shin for advice implementing EVmutation. We thank Robert J. Lefkowitz and Laura Wingler for suggesting control mutations to test and technical guidance. We thank Hiroaki Matsunami, Joshua S. Bloom, Rishi Jajoo and Rocky O. Cheung for expert technical

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

assistance. Funding: National Science Foundation, Brain Initiative (1556207 to S.K), Ruth L. Kirschstein National Research Service Award (GM007185 to N.B.L.), the USPHS National Research Service Award (5T32GM008496 to E.M.J.), the NIH (DP2GM114829 to S.K.), the Medical Research Council (MC_U105185859 to M.M.B. and A.J.V.), and UCLA. Data and materials availability: Processed data and analysis scripts are available on https://github.com/KosuriLab/b2-dms. Raw data is available with accession number XXX. Plasmids and cell lines are available upon request.

Author Contributions:

E.M.J. and S.K. conceptualized the experiments. E.M.J., D.C., M.S., J.D. and J.W. performed the experiments. N.B.L., E.M.J., A.J.V., A.M.T., J.M.P., N.R.L., M.M.B., R.O.D., and S.K. analyzed the results. E.M.J., N.B.L., A.J.V., A.M.T., J.M.P., N.R.L., M.M.B., R.O.D., and S.K. wrote and edited the manuscript.

Declaration of Interests:

E.M.J. and N.B.L. are employed by and hold equity and S.K. consults for and holds equity in Octant Inc. to which patent rights based on this work have been licensed. All other authors declare that they have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figures

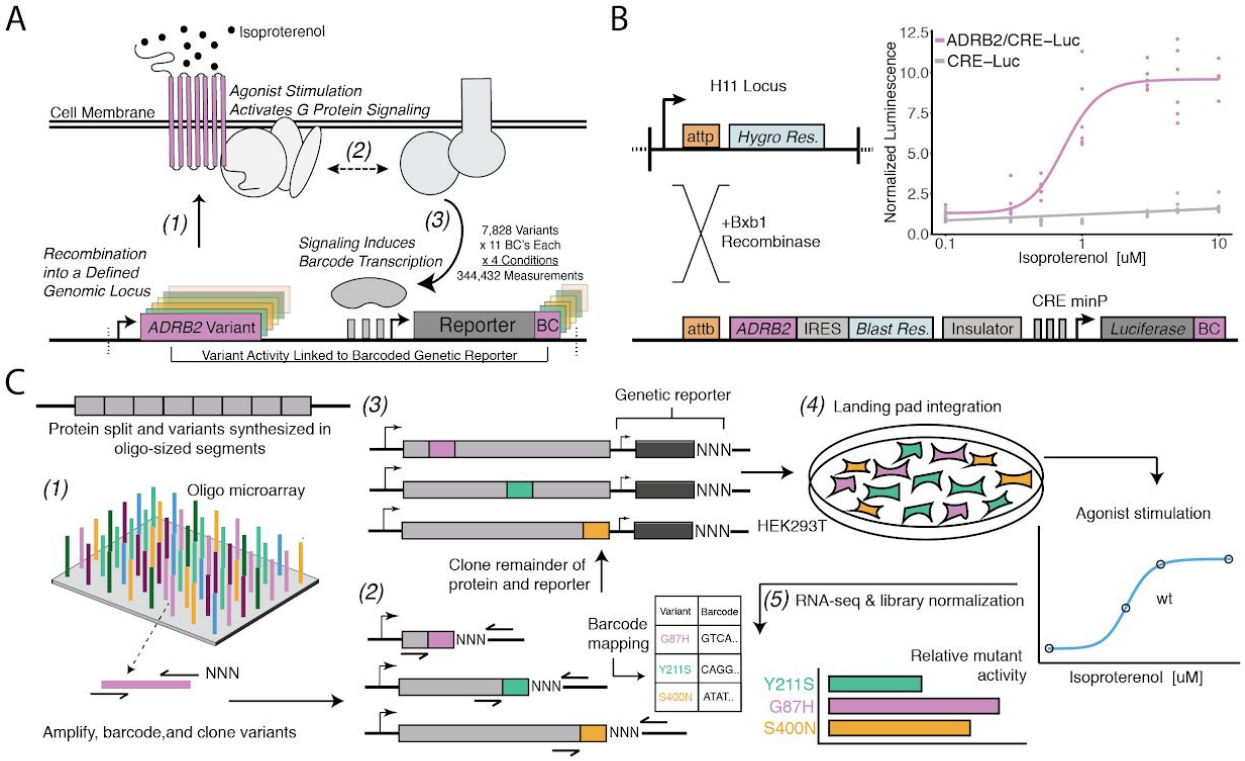

Fig. 1. A Platform for Deep Mutational Scanning of GPCRs. A. Overview of the Multiplexed GPCR Activity Assay. Plasmids encoding *ADRB2* variants, a transcriptional reporter of signaling activity, and 15 nucleotide barcode sequences that identify the variant are integrated into a defined genomic locus such that one variant is present per cell. Upon stimulation by isoproterenol, G protein signaling induces transcription of the cAMP-responsive genetic reporter and the barcode. Thus, the activity of a given variant is proportional to the amount of barcode mRNA which can be read out in multiplex by RNA-seq. **B.** Schematic detailing the recombination of the reporter-receptor expression plasmid into the landing pad locus. Top Right: Activation of the genetic reporter integrated with (purple) or without (grey) exogenous *ADRB2* into the landing pad when stimulated with isoproterenol in Δ*ADRB2* cells via a luciferase reporter gene assay. **C.** Overview of Library Generation and Functional Assay. Missense variants are synthesized on an oligonucleotide microarray, the oligos are amplified with random DNA barcode sequences appended, and the variants are cloned into wild-type background vectors. Barcode-variant pairs are mapped with next-generation sequencing and the remaining wild-type receptor and genetic reporter sequences are cloned into the vector. Next, the variant library is integrated *en masse* into the serine recombinase (Bxb1) landing pad engineered at the H11 locus of Δ*ADRB2* HEK293T cells. This integration strategy ensures a single pair of receptor variant and barcoded genetic reporter pair is integrated per cell and avoids crosstalk. After selection, the library is stimulated with various concentrations of the β_2AR agonist, isoproterenol. Finally, mutant activity is determined by measuring the relative abundance of each variant's barcoded reporter transcript with RNA-seq.

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

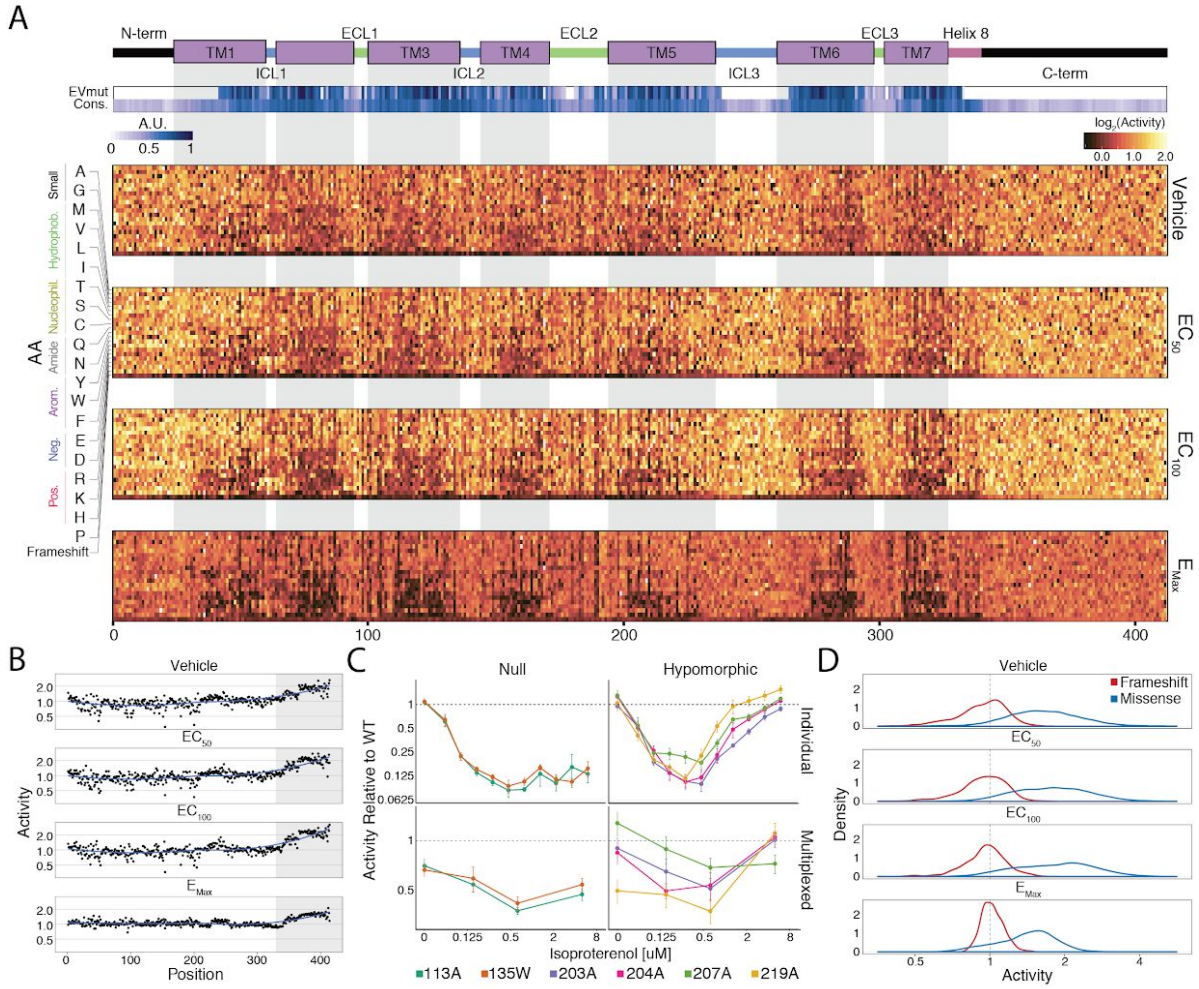

Fig. 2. Variant-Activity Landscape for 7,800 Missense Variants of the β2AR and Multiplexed Assay Validation. A. Top: Secondary structure diagram of the β_2AR: the N and C termini are black, the transmembrane helices are purple blocks, and the intra- and extracellular domains are colored blue and green respectively. The EVmutation track (EVmut.) displays the mean effect of mutation at each position as predicted from sequence covariation (*33*). Conservation track (Cons.) displays the sequence conservation of each residue across 55 β_2AR orthologs (*31*). A.U. stands for arbitrary units and the scale is different for the EVmutation and sequence conservation tracks. The shaded guides represent positions in the transmembrane domain. Bottom: The heatmap representation of mutant activity at each agonist condition. Variants are colored by their activity relative to the mean frameshift mutation. **B.** Relative effect of the mean frameshift mutant activity per position is markedly decreased in the unstructured C-terminus of the protein (shaded region), and is consistent across agonist concentration. Blue line represents the LOESS fit. **C.** Mutant activity measured individually with a luciferase reporter gene compared to the multiplexed assay. Both null and hypomorphic mutations are comparable in both approaches. **D.** The distribution of frameshift mutant activity (red) is significantly different than the distribution of designed missense mutations (blue) across increasing isoproterenol concentrations ($p << 0.001$). Mean frameshift activity marked with a dashed line.

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

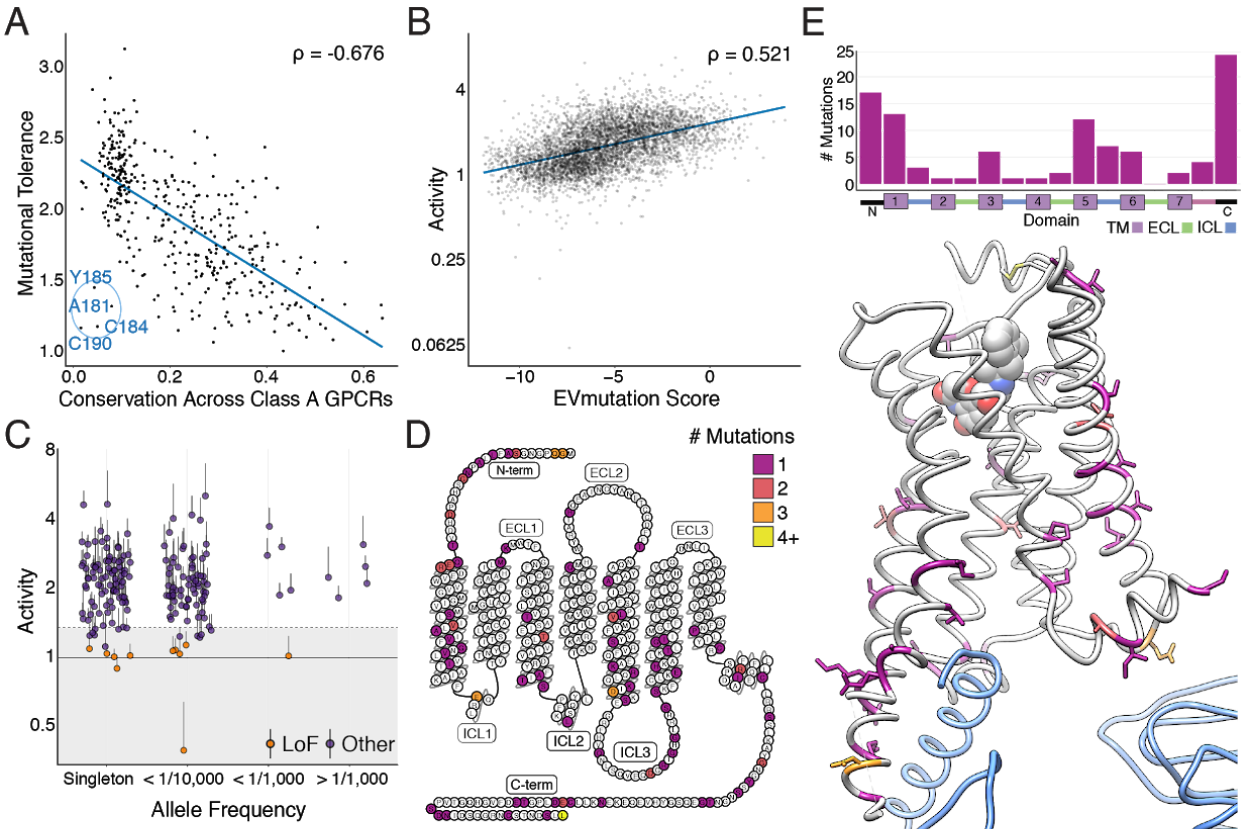

Fig. 3. Individual Mutations and Residues Reveal Evolutionary and Structural Insights into β2AR Function. A. Positional conservation across Class A GPCRs correlates with mutational tolerance (Spearman's ρ = -0.676), the mean activity of all amino acid substitutions per residue at each agonist concentration, at EC_{100}. However, four of the least conserved positions (C190, C184, A181, Y185) are highly sensitive to mutation and are located in ECL2, suggesting this region is uniquely important to the β2AR. The blue line is a simple linear regression. **B.** Individual mutant activity correlates with EVmutation (Spearman's ρ = 0.521) at EC_{100}. The blue line is a simple linear regression. **C.** Activity of individual mutants present in the human population from the gnomAD database stratified by allele frequency. Mutations are classified as potential loss of function (LoF) mutations (orange) are classified as such (shaded region) if the mean activity at EC_{100} plus the standard error of the mean (upper error bar) is more than two standard deviations below mean frameshift mutant activity (dashed line). **D.** The distribution of the 100 most basally activating mutations across the β2AR snake plot reveals a clustering of mutants in the termini, TM1, TM5 and TM6. **E.** Top: Distribution of the 100 most basally activating mutations stratified by domain. Bottom: The distribution of the 100 most basally activating mutations across the β2AR 3D structure (PDB: 3SN6). These positions (colored as in D) are concentrated on the surface of the β2AR ($G_{\alpha s}$ shown in blue).

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

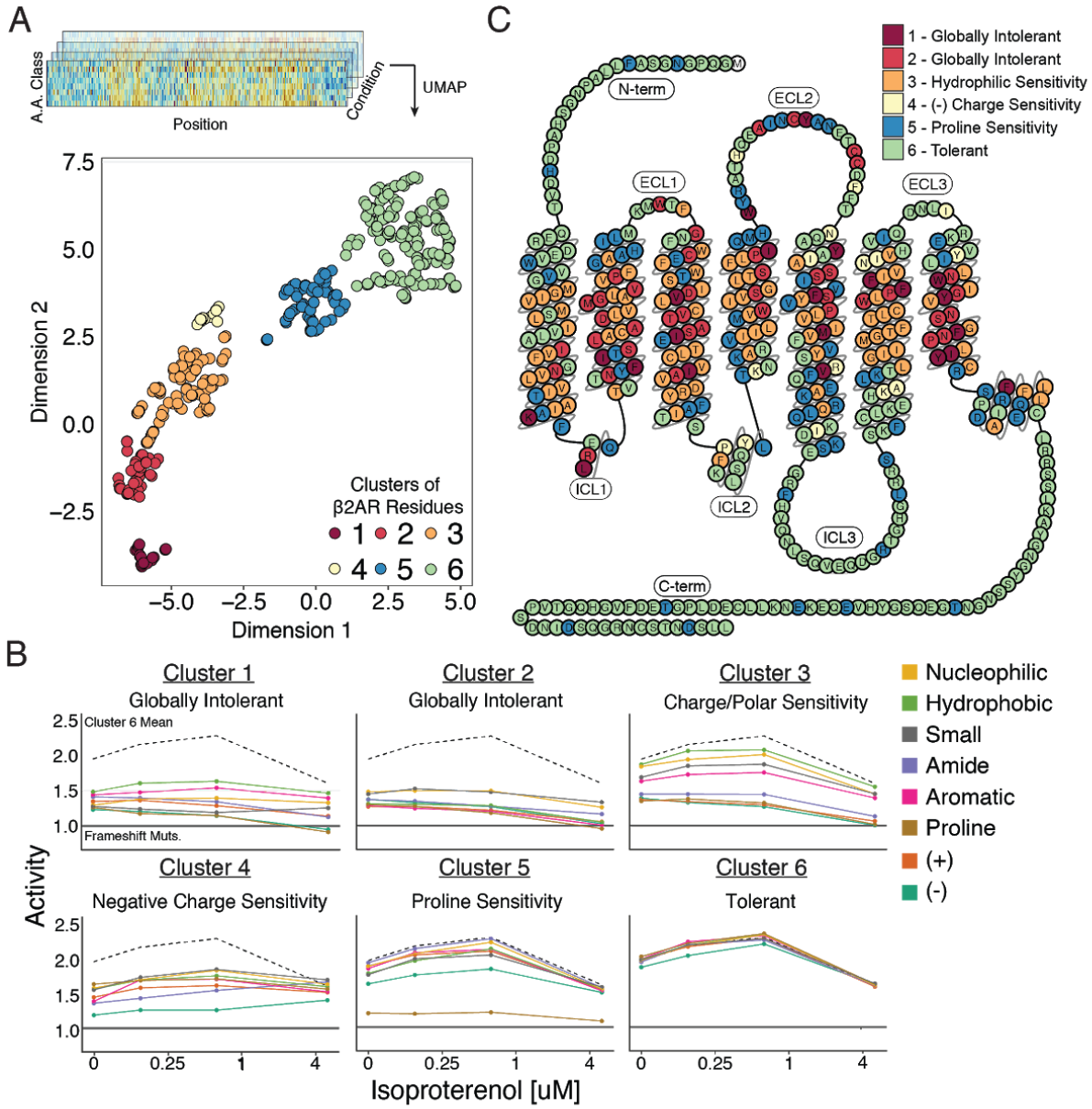

Fig. 4. Unsupervised Learning Segregates Residues into Clusters with Distinct Responses to Mutation. A. Amino acids were segregated into classes based on their physicochemical properties and mean activity scores were reported by class for each residue. With Uniform Manifold Approximation and Projection (UMAP) a 2D representation of every residue's response to each mutation class across agonist conditions was learned. Each residue is assigned into one of six clusters using HDBSCAN (see Fig. S5). **B.** Class averages for each of these cluster reveal distinct responses to mutation. The upper dashed line represents the mean activity of Cluster 6 and the lower solid line represents the mean activity of frameshift mutations. **C.** A 2D snake plot representation of β_2AR secondary structure with each residue colored by cluster identity.

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

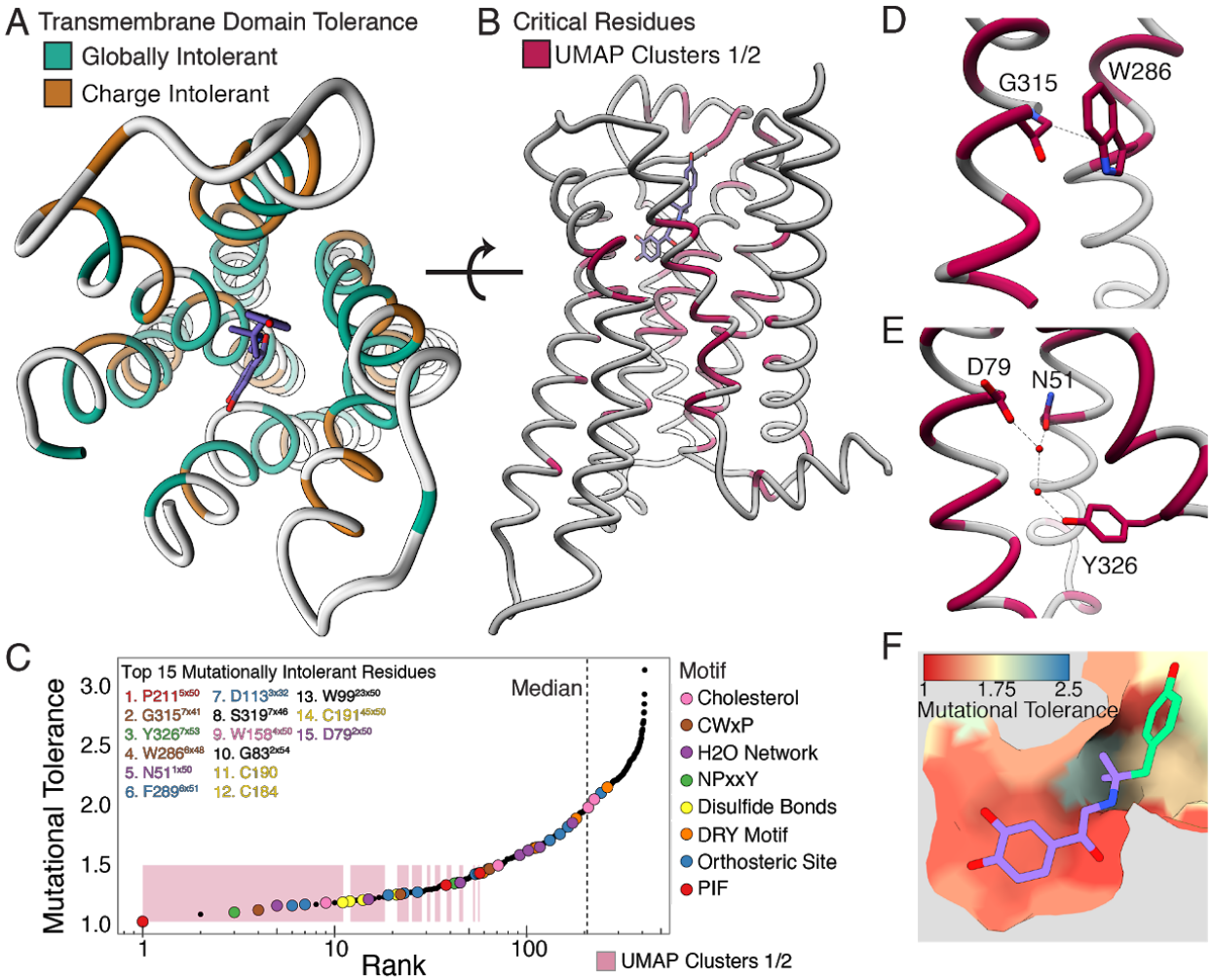

Fig. 5. Mutational Tolerance Elucidates Broad Structural Features and Critical Residues of the β_2AR. A. Residues within the transmembrane domain colored by their tolerance to particular classes of amino acid substitution. Teal residues are intolerant to both hydrophobic and charged amino acids (globally intolerant), and brown residues are tolerant to hydrophobic amino acids but intolerant to charged amino acids (charge intolerant). The charge sensitive positions' side chains are enriched pointing into the membrane, while the globally intolerant positions' side chains face into the core of the protein (see Fig. S5). **B.** The crystal structure of the hydroxybenzyl isoproterenol-activated state of the β_2AR (PDB: 4LDL) with residues from the mutationally intolerant Clusters 1 and 2 highlighted in maroon. **C.** 412 β_2AR residues rank ordered by mutational tolerance at the 0.625 uM isoproterenol condition. Residues in known structural motifs (colored points) are significantly more sensitive to mutation than other positions on the protein ($p << 0.001$). Dashed line demarcates the median of the ranking. The top 15 mutationally intolerant residues are listed and colored by motif association. **D-F.** Selected vignettes of residues from the mutationally intolerant UMAP clusters and ranking. **D.** W286^{6x48} of the CWxP motif and the neighboring G315^{7x41} are positioned in close proximity. Substitutions at G315^{7x41} are likely to cause a steric clash with W286^{6x48} (PDB: 4LDL). **E.** An inactive-state water-mediated hydrogen bond network (red) associates N51^{1x50} and Y326^{7x53} (PDB: 2RH1). Disruption of this network may destabilize the receptor. **F.** The ligand-bound orthosteric site surface colored by mutational tolerance. Receptor-ligand contacts with the catecholamine head

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(present in agonist used in assay) are more intolerant to mutation than those in the hydroxybenzyl tail (not present in agonist used in assay) of the isoproterenol analog depicted in this crystal structure (PDB: 4LDL).

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

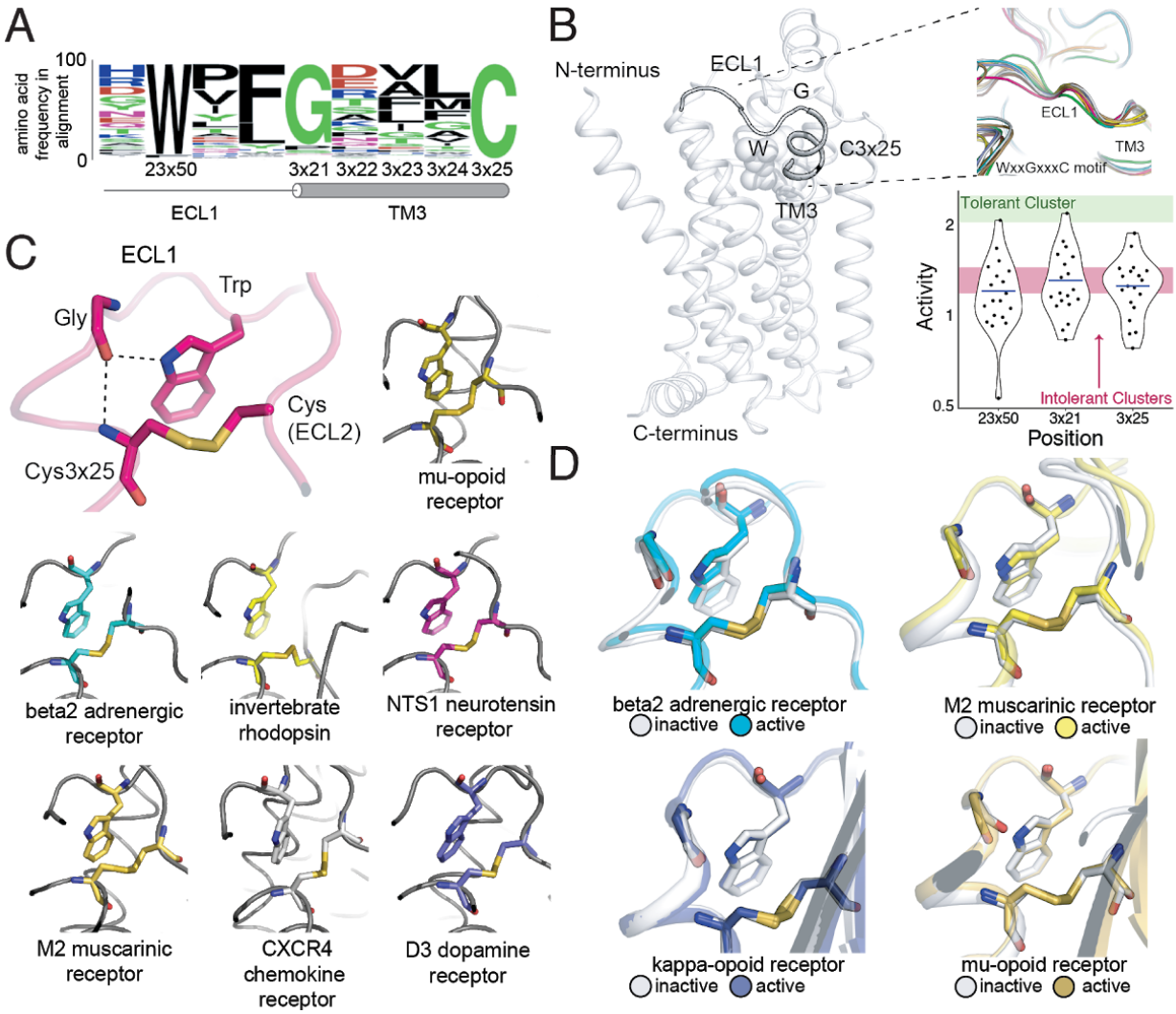

Fig. 6 A Conserved Extracellular Tryptophan-Disulfide "Structural Latch" in Class A GPCRs is Mutationally Intolerant and Conformation-Independent. A. Sequence conservation of extracellular loop 1 (ECL1) and the extracellular interface of TM3 (202 GPCRs). **B.** Left: Depiction of the interaction of W99^{23x50}, G102^{3x21}, and C106^{3x25} in ECL1 of the β_2AR. Top Right: Conservation of the structure of the ECL1 region across functionally different class A GPCRs. Bottom Right: Activity of all 19 missense variants assayed for each of the three conserved residues, with the mean activity (mutational tolerance) shown as a blue bar. The shaded bars represent the mean mutational tolerance ± 1 SD of residues in the tolerant cluster 6 (green) and the intolerant clusters 1 and 2 (red). **C.** A hydrogen bond network between mutationally intolerant positions W99^{23x50}, G102^{3x21}, and C106^{3x25}. Representative examples of the structural latch are shown. **D.** This structural latch is maintained in both the inactive and active state structures for the β_2AR (inactive: 2RH1, active: 3P0G), the M2 muscarinic receptor (inactive: 3UON, active: 4MQS), the kappa-opioid receptor (inactive: 4DJH, active: 6B73), and the mu-opioid receptor (inactive: 4DKL, active: 5C1M)

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

References

1. R. Fredriksson, M. C. Lagerström, L.-G. Lundin, H. B. Schiöth, The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).

2. A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth, D. E. Gloriam, Trends in GPCR drug discovery: new agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).

3. B. K. Kobilka, X. Deupi, Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **28**, 397–406 (2007).

4. X. Deupi, B. K. Kobilka, Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. *Physiology* . **25**, 293–303 (2010).

5. V. Isberg *et al.*, GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.* **44**, D356–64 (2016).

6. S. Granier, B. Kobilka, A new era of GPCR structural and chemical biology. *Nat. Chem. Biol.* **8**, 670–673 (2012).

7. N. R. Latorraca, A. J. Venkatakrishnan, R. O. Dror, GPCR Dynamics: Structures in Motion. *Chem. Rev.* **117**, 139–155 (2017).

8. A. Manglik, B. Kobilka, The role of protein dynamics in GPCR function: insights from the β2AR and rhodopsin. *Curr. Opin. Cell Biol.* **27**, 136–143 (2014).

9. A.-M. Schönegge *et al.*, Evolutionary action and structural basis of the allosteric switch controlling β 2 AR functional selectivity. *Nat. Commun.* **8**, 2169 (2017).

10. G. Pei *et al.*, A constitutively active mutant beta 2-adrenergic receptor is constitutively desensitized and phosphorylated. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 2699–2702 (1994).

11. L. Valentin-Hansen *et al.*, The arginine of the DRY motif in transmembrane segment III functions as a balancing micro-switch in the activation of the β2-adrenergic receptor. *J. Biol. Chem.* **287**, 31973–31982 (2012).

12. P. Azimzadeh, J. A. Olson Jr, N. Balenga, Reporter gene assays for investigating GPCR signaling. *Methods Cell Biol.* **142**, 89–99 (2017).

13. Z. Cheng *et al.*, Luciferase Reporter Assay System for Deciphering GPCR Pathways. *Curr. Chem. Genomics*. **4**, 84–91 (2010).

14. W. K. Kroeze *et al.*, PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat. Struct. Mol. Biol.* **22**, 362–369 (2015).

15. D. M. Fowler, S. Fields, Deep mutational scanning: a new style of protein science. *Nat.*

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Methods. **11**, 801–807 (2014).

16. L. M. Starita *et al.*, Variant Interpretation: Functional Assays to the Rescue. *Am. J. Hum. Genet.* **101**, 315–325 (2017).

17. E. Kotler *et al.*, A Systematic p53 Mutation Library Links Differential Functional Impact to Cancer Mutation Pattern and Evolutionary Conservation. *Mol. Cell*. **71**, 178–190.e8 (2018).

18. A. R. Majithia *et al.*, Prospective functional classification of all possible missense variants in PPARG. *Nat. Genet.* (2016), doi:10.1038/ng.3700.

19. L. M. Starita *et al.*, A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. *Am. J. Hum. Genet.* **103**, 498–508 (2018).

20. J. M. Lee *et al.*, Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8276–E8285 (2018).

21. K. A. Matreyek *et al.*, Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat. Genet.* **50**, 874–882 (2018).

22. J. D. Heredia *et al.*, Mapping Interaction Sites on Human Chemokine Receptors by Deep Mutational Scanning. *J. Immunol.* **200**, 3825–3839 (2018).

23. J. Park *et al.*, Structural architecture of a dimeric class C GPCR based on co-trafficking of sweet taste receptor subunits. *J. Biol. Chem.* **294**, 4759–4774 (2019).

24. R. Cheung *et al.*, A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions. *Mol. Cell*. **73**, 183–194.e8 (2019).

25. K. A. Matreyek, J. J. Stephany, D. M. Fowler, A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res.* **45**, e102 (2017).

26. X. Duportet *et al.*, A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.* **42**, 13440–13451 (2014).

27. E. M. LeProust *et al.*, Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.* **38**, 2522–2540 (2010).

28. T. Sato, H. Kobayashi, T. Nagao, H. Kurose, Ser203 as well as Ser204 and Ser207 in fifth transmembrane domain of the human beta2-adrenoceptor contributes to agonist binding and receptor activation. *Br. J. Pharmacol.* **128**, 272–274 (1999).

29. C. E. Elling, K. Thirstrup, B. Holst, T. W. Schwartz, Conversion of agonist site to metal-ion chelator site in the beta(2)-adrenergic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12322–12327 (1999).

30. S. K. Shenoy *et al.*, beta-arrestin-dependent, G protein-independent ERK1/2 activation by

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the beta2 adrenergic receptor. *J. Biol. Chem.* **281**, 1261–1273 (2006).

31. J. A. Capra, M. Singh, Predicting functionally important residues from sequence conservation. *Bioinformatics*. **23**, 1875–1882 (2007).

32. I. Adzhubei, D. M. Jordan, S. R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **Chapter 7**, Unit7.20 (2013).

33. T. A. Hopf *et al.*, Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017).

34. A. M. Altenhoff *et al.*, The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. *Nucleic Acids Res.* **46**, D477–D485 (2018).

35. A. S. Hauser, S. Chavali, I. Masuho, L. J. Jahn, M. Madan Babu, Pharmacogenomics of GPCR Drug Targets. *Cell*. **172** (2017), doi:10.1016/j.cell.2017.11.033.

36. K. J. Karczewski *et al.*, Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* (2019), p. 531210.

37. W. I. Weis, B. K. Kobilka, The Molecular Basis of G Protein-Coupled Receptor Activation. *Annu. Rev. Biochem.* **87**, 897–919 (2018).

38. C. Dong, C. M. Filipeanu, M. T. Duvernay, G. Wu, Regulation of G protein-coupled receptor export trafficking. *Biochim. Biophys. Acta*. **1768**, 853–870 (2007).

39. C. Ozgur, P. Doruker, E. D. Akten, Investigation of allosteric coupling in human β2-adrenergic receptor in the presence of intracellular loop 3. *BMC Struct. Biol.* **16**, 9 (2016).

40. O. Ozcan, A. Uyar, P. Doruker, E. D. Akten, Effect of intracellular loop 3 on intrinsic dynamics of human β2-adrenergic receptor. *BMC Struct. Biol.* **13**, 29 (2013).

41. G. M. West *et al.*, Ligand-dependent perturbation of the conformational ensemble for the GPCR β2 adrenergic receptor revealed by HDX. *Structure*. **19**, 1424–1432 (2011).

42. L. A. Lotta *et al.*, Human Gain-of-Function MC4R Variants Show Signaling Bias and Protect against Obesity. *Cell*. **177**, 597–607.e9 (2019).

43. L. McInnes, J. Healy, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv [stat.ML]* (2018), (available at http://arxiv.org/abs/1802.03426).

44. R. J. G. B. Campello, D. Moulavi, J. Sander, in *Advances in Knowledge Discovery and Data Mining* (Springer Berlin Heidelberg, 2013), pp. 160–172.

45. D. M. Rosenbaum *et al.*, GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science*. **318**, 1266–1273 (2007).

46. S. Mitternacht, FreeSASA: An open source C library for solvent accessible surface area

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

calculations. *F1000Res.* **5**, 189 (2016).

47. A. J. Venkatakrishnan *et al.*, Stable networks of water-mediated interactions are conserved in activation of diverse GPCRs. *bioRxiv* (2018), p. 351502.

48. A. J. Venkatakrishnan *et al.*, Diverse GPCRs exhibit conserved water networks for stabilization and activation. *Proc. Natl. Acad. Sci. U. S. A.* (2019), doi:10.1073/pnas.1809251116.

49. A. J. Venkatakrishnan *et al.*, Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* (2016), doi:10.1038/nature19107.

50. D. M. Thal, A. Glukhova, P. M. Sexton, A. Christopoulos, Structural insights into G-protein-coupled receptor allostery. *Nature*. **559**, 45–53 (2018).

51. M. A. Hanson *et al.*, A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor. *Structure*. **16**, 897–905 (2008).

52. S. G. F. Rasmussen *et al.*, Crystal structure of the β2 adrenergic receptor-Gs protein complex. *Nature*. **477**, 549–555 (2011).

53. O. Moro, J. Lameh, P. Högger, W. Sadée, Hydrophobic amino acid in the i2 loop plays a key role in receptor-G protein coupling. *J. Biol. Chem.* **268**, 22273–22276 (1993).

54. S. P. Sheikh *et al.*, Similar Structures and Shared Switch Mechanisms of the β2-Adrenoceptor and the Parathyroid Hormone Receptor: Zn(II) BRIDGES BETWEEN HELICES III AND VI BLOCK ACTIVATION. *J. Biol. Chem.* **274**, 17033–17041 (1999).

55. B. F. O'Dowd *et al.*, Site-directed mutagenesis of the cytoplasmic domains of the human beta 2-adrenergic receptor. Localization of regions involved in G protein-receptor coupling. *J. Biol. Chem.* **263**, 15985–15992 (1988).

56. M. Valiquette, S. Parent, T. P. Loisel, M. Bouvier, Mutation of tyrosine-141 inhibits insulin-promoted tyrosine phosphorylation and increased responsiveness of the human beta 2-adrenergic receptor. *EMBO J.* **14**, 5542–5549 (1995).

57. G. Swaminath, T. W. Lee, B. Kobilka, Identification of an Allosteric Binding Site for Zn2+on the β2 Adrenergic Receptor. *J. Biol. Chem.* **278**, 352–356 (2003).

58. A. D. Jensen *et al.*, Agonist-induced Conformational Changes at the Cytoplasmic Side of Transmembrane Segment 6 in the β2 Adrenergic Receptor Mapped by Site-selective Fluorescent Labeling. *J. Biol. Chem.* **276**, 9279–9290 (2001).

59. K. Noda, Y. Saad, R. M. Graham, S. S. Karnik, The high affinity state of the beta 2-adrenergic receptor requires unique interaction between conserved and non-conserved extracellular loop cysteines. *J. Biol. Chem.* **269**, 6743–6752 (1994).

60. H. G. Dohlman, M. G. Caron, A. DeBlasi, T. Frielle, R. J. Lefkowitz, Role of extracellular disulfide-bonded cysteines in the ligand binding function of the beta 2-adrenergic receptor. *Biochemistry*. **29**, 2335–2342 (1990).

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

61. R. Bhattacharyya, D. Pal, P. Chakrabarti, Disulfide bonds, their stereospecific environment and conservation in protein structures. *Protein Eng. Des. Sel.* **17**, 795–808 (2004).

62. T. Hori *et al.*, Na+-mimicking ligands stabilize the inactive state of leukotriene B4 receptor BLT1. *Nat. Chem. Biol.* **14**, 262–269 (2018).

63. A. V. Khera *et al.*, Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell*. **177**, 587–596.e9 (2019).

64. I. S. Farooqi *et al.*, Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N. Engl. J. Med.* **348**, 1085–1095 (2003).

65. V. Cherezov *et al.*, High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science*. **318**, 1258–1265 (2007).

66. S. G. F. Rasmussen *et al.*, Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature*. **450**, 383–387 (2007).

67. S. Galandrin, G. Oligny-Longpré, M. Bouvier, The evasive nature of drug efficacy: implications for drug discovery. *Trends Pharmacol. Sci.* **28**, 423–430 (2007).

68. D. Hilger, M. Masureel, B. K. Kobilka, Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* **25**, 4–12 (2018).

69. L. M. Luttrell, Reviews in molecular biology and biotechnology: transmembrane signaling by G protein-coupled receptors. *Mol. Biotechnol.* **39**, 239–264 (2008).

70. S. M. DeWire, S. Ahn, R. J. Lefkowitz, S. K. Shenoy, β-Arrestins and Cell Signaling. *Annu. Rev. Physiol.* **69**, 483–510 (2007).

71. E. Reiter, S. Ahn, A. K. Shukla, R. J. Lefkowitz, Molecular mechanism of β-arrestin-biased agonism at seven-transmembrane receptors. *Annu. Rev. Pharmacol. Toxicol.* **52**, 179–197 (2012).

72. C. G. Tate, G. F. X. Schertler, Engineering G protein-coupled receptors to facilitate their structure determination. *Curr. Opin. Struct. Biol.* **19**, 386–395 (2009).

73. M. J. Serrano-Vega, F. Magnani, Y. Shibata, C. G. Tate, Conformational thermostabilization of the beta1-adrenergic receptor in a detergent-resistant form. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 877–882 (2008).

74. E. M. Jones *et al.*, A Scalable, Multiplexed Assay for Decoding GPCR-Ligand Interactions with RNA Sequencing. *Cell Syst*. **8**, 254–260.e6 (2019).

75. A. Botvinik, M. J. Rossner, Linking cellular signalling to gene expression using EXT-encoded reporter libraries. *Methods Mol. Biol.* **786**, 151–166 (2012).

76. S. Galinski, S. P. Wichert, M. J. Rossner, M. C. Wehr, Multiplexed profiling of GPCR activities by combining split TEV assays and EXT-based barcoded readouts. *Sci. Rep.* **8**, 8137 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/623108; this version posted May 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

77. D. J. O'Connell *et al.*, Simultaneous Pathway Activity Inference and Gene Expression Analysis Using RNA Sequencing. *Cell Syst*. **2**, 323–334 (2016).

78. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

79. J. D. Bloom, Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinformatics*. **16**, 168 (2015).

80. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

81. M. A. Lomize, I. D. Pogozheva, H. Joo, H. I. Mosberg, A. L. Lomize, OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–6 (2012).

82. H. M. Berman *et al.*, The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

83. G. Pándy-Szekeres *et al.*, GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.* **46**, D440–D446 (2018).

84. F. Crameri, Geodynamic diagnostics, scientific visualisation and StagLab 3.0. *Geoscientific Model Development*. **11**, 2541–2562 (2018).

85. O. Tange, Others, Gnu parallel-the command-line power tool. *The USENIX Magazine*. **36**, 42–47 (2011).