Exploring use of unsupervised clustering to associate signaling profiles of GPCR ligands to clinical response

Besma Benredjem1,2,11, Jonathan Gallion3,11, Dennis Pelletier4, Paul Dallaire1,2, Johanie Charbonneau2, Darren Cawkill4,7, Karim Nagi5, Mark Gosink4, Viktoryla Lukasheva6, Stephen Jenkinson4,8, Yong Ren4,9, Christopher Somps4, Brigitte Murat6, Emma Van Der Westhuizen6,10, Christian Le Gouill6, Olivier Lichtarge3, Anne Schmidt4, Michel Bouvier6 & Graciela Pineyro1,2

Signaling diversity of G protein-coupled (GPCR) ligands provides novel opportunities to develop more effective, better-tolerated therapeutics. Taking advantage of these opportunities requires identifying which effectors should be specifically activated or avoided so as to promote desired clinical responses and avoid side effects. However, identifying signaling profiles that support desired clinical outcomes remains challenging. This study describes signaling diversity of mu opioid receptor (MOR) ligands in terms of logistic and operational parameters for ten different in vitro readouts. It then uses unsupervised clustering of curve parameters to: classify MOR ligands according to similarities in type and magnitude of response, associate resulting ligand categories with frequency of undesired events reported to the pharmacovigilance program of the Food and Drug Administration and associate signals to side effects. The ability of the classification method to associate specific in vitro signaling profiles to clinically relevant responses was corroborated using β2-adrenergic receptor ligands.
protein-coupled receptors (GPCRs) modulate practically
every aspect of human physiology and are the target of
~30% of FDA-approved medicines. When activated these
receptors undergo conformational changes that determine
the type and magnitude of signals triggered within the cell. This
signaling configuration supports ligand-specific activation of
the different pathways, and provides a theoretical opportunity
for directing pharmacological stimulus toward pathways that
underlie desired therapeutic responses and away from those
responsible for undesired side effects. However, in spite of
this promise development of therapeutic biased ligands has
yet to translate into more effective and/or better-tolerated
medicines.

Different challenges have hindered the development of clini-
cally effective biased ligands. Except for limited examples, we
ignore the signals underlying desired and undesired clinical
responses of GPCR ligands. To access this knowledge and apply it
to drug discovery, it is necessary to identify signaling preferences
and to associate distinct signaling profiles to desired/undesired
clinical outcomes. The way in which signaling preferences are
currently identified in drug discovery efforts involves calculation
of “bias factors”, an approach that uses consolidated (Log(τ/KA))
transduction coefficients to measure the extent to which a ligand
preferentially activates one pathway over another. This type
of evaluation compares signals in a pairwise manner, a dichot-
omous approach that provides a fragmented view of a ligand’s
signaling preferences across the multiplicity of pathways. Perhaps
more troubling for the use of “bias factors” as descriptors of
potential clinical responses is the fact that their estimated mag-
nitudes vary with the calculation method used to produce them.
Finally, the same “bias factor” may describe drugs with very
different efficacies at the pathways of interest further ques-
tioning the value of these measures as predictors of desired/undesired
in vivo responses. In an effort to circumvent at least some of these limitations, we sought an alternative way to identify
signaling preferences and classify GPCR ligands.

One of the most studied examples of how biased signaling
may support development of more effective and/or better-
tolerated therapeutic agents is that of opioid analgesics. Pre-
clinical models have indicated that β-arrestin2 (βarr2) knock-
out mitigates constipation and respiratory depression induced
by morphine, pointing to the possibility that mu opioid
receptor (MOR) agonists that preferentially activate G protein
signaling over βarr2 recruitment could induce less of these side
effects. Here, we use this prototypical example to establish
that clustering MOR ligands according to similarities in
pharmacodynamic parameters for multiple responses, captures
their signaling differences and preferences. We show that
ligands with similar G protein/βarr responses cluster together,
and provide evidence that ligands within different categories
display distinct frequencies of gastrointestinal and respiratory
events reported to the FDA pharmacovigilance program.
Moreover, when ligands are clustered according to either G
protein or βarr responses both signals directly associate to side
effects. The practical value of the classification method pro-
posed is further illustrated by the fact that ligand categories
defined by similarity of G protein responses at β2-adrenergic
receptor (β2AR) correlate with sympatholytic CV events and
bronchoconstriction.

Results
Clustering ligands according to pharmacodynamic similarities.

We sought a method to identify and group together ligands with
overall similarities in a multiplicity of signaling pathways while
simultaneously discerning those with overall differences in
features, such as efficacy, potency, and signaling preferences. To
test the ability of the method to accomplish this task independent
of idiosyncrasies in experimental data sets, we generated a set of
320 virtual compounds as variations of 16 prototypical profiles
characterized by a combination of pharmacodynamic features
across six different readouts (see the Methods section). Profiles
are shown in Supplementary Fig. 1. Criteria to classify ligands
according to pharmacodynamic similarities were empirically
established by generating matrices, in which each ligand was
represented by individual logistic (E \(_\text{max}\), pEC50) or operational
(Log(τ), pKA, Log(τ/KA)) parameters, as well as their combi-
nations. Matrices were then subject to nonnegative matrix fac-
torization (NNMF) to identify essential, nonredundant features, and k-means clustering was subsequently used to clas-
sify ligands according to these features. Iterations were used to
incorporate the error associated with each mean parameter value,
ensuring its propagation throughout the clustering procedure
(see the Methods and Supplementary Fig. 2). The result of this
procedure was a ligand × ligand similarity matrix that quantifies
how frequently any two compounds clustered together over the
iterations. Final similarity matrices were then subject to hier-
archical clustering to establish row and column ordering
according to similarity, and visualized as heatmaps. Figure 1
shows heatmaps for the progressive associations of parameters
leading to identification of Log(τ), E \(_\text{max}\) and Log(τ/KA) as a
combination faithfully recreating the 16 profiles initially defined.
Operational efficacy (τ) by itself was not sufficient to fully dis-
tinguish ligands with different profiles (Fig. 1a, d). Introducing
measures of signaling capacity (E \(_\text{max}\)) improved the classification
(Fig. 1e), but discrimination was not optimal unless values for
transduction coefficient Log(τ/KA) were also included (Fig. 1c,
f). Unlike Log(τ) and E \(_\text{max}\) values, transduction coefficients
incorporate potency information and thus provide a differ-
tent dimension on which ligands can be distinguished. In effect,
Log(τ/KA) coefficients were correlated with logistic potency
estimates (pEC50) (Supplementary Fig. 3a), so the classification
afforded by Log(τ)–E \(_\text{max}\)–pEC50 (Supplementary Fig. 3b, c) was
quite similar to the one produced with Log(τ)–E \(_\text{max}\)-Log(τ/KA).
Profiles recreated by classifying ligands according to Log(τ)–
E \(_\text{max}\)-Log(τ/KA), are shown in Supplementary Fig. 4, revealing
a minimal number of displaced compounds.

We then applied the proposed classification strategy on experi-
mental data. Multidimensional signaling profiles for this
analysis were generated using ten different BRET-based biosen-
sors that monitor βarr recruitment and G protein signaling. G
protein signaling was monitored through conformational rearrange-
ments within Gα13, Gβγ, and heterotrimers, at the interface of
Gβγ2/Kir3 channel subunits, or as changes in cAMP levels.
βarr recruitment to the receptor was assessed for βarr1, βarr2, and βarr2 in presence of GRK2, GRK5, or GRK6 to
account for possible impact of expression differences between
the screening system (HEK 293) and target neurons where GRK
levels are higher. Net BRET values obtained in cells co-
expressing human MORs (hMOR) and different biosensors in
presence/absence of the endogenous ligand Met-Enkephalin
(Met-ENK) are shown for reference in Supplementary Fig. 5.
Concentration response curves (CRCs) for 10 known opioids
(Met-ENK) are shown for reference in Supplementary Fig. 5.

Ligands within the same cluster share quality and magnitude of response. An essential pharmacological question is to identify the pathways and pharmacodynamic properties primarily responsible for ligand clustering. The overall resemblance among relative magnitudes of operational and logistic parameters from different functional readouts is shown in Fig. 3b delineating three clusters of parameters. To further characterize differences among ligand categories, we investigated whether the magnitude of parameters describing ligand response in each assay was different across the three clusters of ligands. To do so, we used a Kolmogorov–Smirnoff test to compare the distribution of parameter values in each cluster to that of the overall population (detailed in Supplementary Fig. 7). Only certain parameters in each cluster contributed to ligand discrimination, and they did so to different extents (Fig. 3c). Those in cluster A had the most weight, as 29.9% of comparisons identified at least one distribution of parameters significantly different from that of the whole population. Overall, 14.9% of comparisons in cluster B and 3.5% in cluster C also significantly contributed to ligand discrimination. Alternatively, ordering parameters by type (Fig. 3d) revealed that efficacy-related parameters ($E_{\text{max}}$ and $\tau$) had the most weight in the classification (52.0% and 36.0% of the comparisons, respectively, identified distributions different from the whole population) while $\log(\tau/K_{A})$ played a smaller role separating...
primary compound clusters (11.0%). Finally, Supplementary Table 1 shows assay parameters significantly contributing to ligand clustering. The diversity of signals determining cluster assignment distinguishes this multidimensional classification from dichotomous comparisons underlying bias magnitudes.

Importantly, despite independence from bias magnitudes, the proposed classification strategy still allows to evaluate relative contributions of β-arrestin and G protein signaling to ligand assignment to clusters. To access this information, drugs were re-clustered using subsets of the data corresponding exclusively to G


**Fig. 2** βarr recruitment and G-protein responses generated by opioid ligands. Responses for prescription opioids and known hMOR ligands were monitored using BRET-based biosensors. The results correspond to mean±SEM of at least three independent experiments, normalized to the maximal effect of Met-ENK, which was tested in all experimental runs (n = 16–29). Curves were fit with the operational model and the logistic equation (the results from the logistic fit are shown). a βarr1 recruitment, (b) βarr2 recruitment, (c) βarr2 recruitment in presence of GRK2, (d) βarr2 recruitment in presence of GRK5, (e) βarr2 recruitment in presence of GRK6, (f) cAMP, (g) Ga1 activation, (h) Ga2 activation, (i) Ga2A activation, and (j) Kir3.1/3.2 activation. Net BRET values for Met-ENK are shown in Supplementary Fig. 5 and dose–response curves for all novel compounds appear in Supplementary Fig. 6. Operational and logistic parameters provided in Supplementary Data 1. Source data are provided as a source data file.

protein (Supplementary Fig. 8a) or to βarr (Supplementary Fig. 8b) assays, and the resulting similarity matrix produced with each partial data set was compared with the matrix generated using the complete set of values. Differences between matrices were quantified as described in the Methods section and Supplementary Fig. 9, and expressed as the proportion of changes in ligand distances that were compatible with a switch in clusters between the two compared matrices. Clusters generated with βarr data differed by only 11.5% from clusters produced with the complete data set (Fig. 3e), underscoring the similarity of drug classes defined by βarr signaling patterns and complete signaling profiles. The partial data set for G protein responses differed by 27.2% (Fig. 3e) from the complete ligand similarity matrix. Thus, although clusters generated with βarr or with G protein data sets resembled clusters produced with the complete data set more than did clusters generated with the corresponding randomized values, initial ligand clustering was more faithfully recreated by βarr responses (Fig. 3e), indicating that that this signal was the one primarily driving classification in the complete matrix. Profiles graphically representing $E_{\text{max}}$ and Log(τ/KA) values for G protein and βarr readouts further highlight how the analysis clustered ligands according to type and magnitude of responses elicited (Fig. 4). Ligands in cluster #3 were full, reasonably balanced agonists characterized by maximal effects at βarr and G protein readouts. Ligands in cluster #2 were partial agonists for G protein-mediated responses with measurable βarr recruitment only in presence of overexpressed GRKs, while ligands in cluster #1 displayed minimal or no βarr recruitment and G protein responses were overall smaller than in cluster #2.

While it may be feasible to monitor ten different signaling outcomes for a small group of ligands, it is unlikely that this could be done in high-throughput screening or structure–activity profiling. Hence, once we had identified the signals that contributed the most to drug classification, we determined whether a reduced number of assays could convey similar diversity. To this end, βarr2 + GRK2, βarr2 + GRK6, Ga12, and cAMP were chosen as respective prototypes of βarr- and G protein responses. Similarity matrices generated from these individual signals or from their combinations were compared with the complete similarity matrix. As above, we compared each partial data set to the complete set of hMOR parameters, and then established if the proportion of ligands switching clusters was less than that observed for the corresponding randomized data set. For cAMP, the proportion of changes were larger than the expected random value (Fig. 3f), indicating minimal contribution of this signal to ligand classification. In contrast, for clusters generated with the other data subsets, the proportion of ligands switching clusters was significantly smaller than the random expectation (Fig. 3f), indicating that each of these signals significantly supported ligand discrimination in the complete data set, albeit to different extents.

The combination of βarr2 + GRK2 and βarr2 + GRK6 data was the best at reproducing clustering obtained with the complete hMOR similarity matrix (91.4%; Fig. 3f). In comparison, the Ga12 data set either combined with cAMP or in isolation moderately recreated the clusters of the complete matrix (Ga12 = 74.4%, Ga12 + cAMP = 69.0% differences). Combining all four assays added little extra precision as compared with βarr2 + GRK2 with βarr2 + GRK6 (Fig. 3f). Thus, taken together, these data indicate that it is possible to first identify the signals that primarily contribute to signaling diversity of a group of compounds at a given receptor, and then use these signals as surrogate readouts for screening campaigns over large collections of compounds.

**Ligand clusters are informative of possible side effects.** Preclinical studies suggest that signaling diversity of MOR agonists provides a means of improving tolerability of opioid analgesics12–14. Therefore, it was of interest to determine if the pharmacodynamic clusters just defined could inform us about clinical side effects of ligands in each category. To address this issue, we first used standardized gamma (SD gamma) scores33 to identify adverse events most frequently reported for opioids in the Food and Drug Administration’s pharmacovigilance data base (Adverse Effects Report System (AERS)), and then calculated the scores of these events for each of the prescription opioids used in the study. These measures of side effect prevalence were associated to ligands in the different clusters by using the Euclidian distance between ligands in the similarity matrix. Tramadol was set as the arbitrary origin, and distances separating the rest of prescription opioids from tramadol in the Log(τ)-$E_{\text{max}}$-Log(τ/KA) matrix were consigned as measures of ligand similarity. These measures were then correlated to the SD gamma scores for each ligand’s side effects. A complete list of the 80 events considered along with $r^2$ and p-values for each correlation is provided in Supplementary Data 2. Correlations that were significant (p ≤ 0.05) and/or explained at least 60% of the variance ($r^2$ ≥ 0.60) were considered. Applying these criteria, ligand position in the Log(τ)-$E_{\text{max}}$-Log(τ/KA) matrix was correlated to 6 out of a total of 80 associations considered (7.5%), including gastrointestinal (GI) events, respiratory depression, and somnolence (Table 1), all typically associated to opioid therapy34,35. These correlations confirm that signaling categories defined by unsupervised clustering can be associated to distinct frequency of report of undesired effects of opioid ligands.

Log(τ) and $E_{\text{max}}$ were the main determinants of ligand position in the matrix constituting 89.0% of parameters effectively grouping ligands into clusters (Fig. 3d). Not surprisingly, if ligands were classified exclusively using these efficacy-related parameters, all side effects previously associated with ligand position in the Log(τ)-$E_{\text{max}}$-Log(τ/KA) matrix remained correlated with their positions in this efficacy-only matrix (Table 1). Actually, categories driven by efficacy measures associated with more side effects than clusters established by including Log(τ/KA) as an additional classification criterion (Table 1). Thus, even if functional affinity information within transduction coefficient affords better discrimination of ligands, it also acts as a confounder for cluster association to side effects.

**Associating side effects to specific signals.** Preclinical studies have suggested that MOR agonists that preferentially engage G protein over βarr responses could display less gastrointestinal and
**Fig. 3** Assignment of hMOR ligands into clusters is primarily driven by βarr responses. Ligand (a) and parameter (b) similarity heatmaps for the complete hMOR data set. Yellow and blue, respectively indicate ligands/parameters that never or always cluster together. Distribution of parameters describing ligands within clusters shown in (a) was compared to their distribution in the whole population using a two-sample Kolmogorov-Smirnov test. Resulting p-values were plotted according to clusters shown in (b) (c) or to parameter type (d), mean ± SD are also shown. Red line: p = 0.05. Similarity matrices corresponding to partial data sets for βarr- or G-protein-mediated responses were compared with the complete, reference hMOR data set. Filled bars: proportion of ligands changing clusters when comparing actual βarr and G-protein data sets to the reference; empty bars: proportions observed by comparing simulations of random clustering to the reference data set. **p < 0.001; ***p < 0.001 comparing partial data sets to their randomized controls; z-scores for comparisons between actual and randomized data: βarr2-GRK2: −12.724; βarr2-GRK6: −10.583; βarr2-GRK2/6: −8.835; Gai2: −7.315; cAMP: 6.297; Gai2- cAMP: −2.351; four assays −8.541; βarr2-GRK2/6/Gai2: −7.391, ###p < 0.001; z-score difference: βarr2-GRK2/6 vs. Gai2/ cAMP: −3.308; z-score difference: Gai2 vs βarr2-GRK2/6/Gai2: 3.754 (f). Source data provided in Supplementary Data 1 and as a source data file.
respiratory side effects in the clinic\(^{12-14}\). Hence, we were interested to find out whether AERS reports for opioids would distinctively correlate with ligand categories defined either by G protein or \( \beta \)arr signaling. There has been considerable debate as to whether biased signaling is best identified using \( \log(\tau/KA) \) transduction coefficients\(^{17,24,25}\) or efficacy-related measures\(^{36,37}\). Hence, partial matrices in which drugs were classified according to G protein or \( \beta \)arr responses were generated using either \( \log(\tau) \)-\( E_{\text{max}} \) or \( \log(\tau) \)-\( E_{\text{max}} \)-\( \log(\tau/KA) \) as classification criteria. Supplementary Fig. 10 shows how frequency of faecaloma report correlates to similarity scores in these four partial matrices. Considering the classification based exclusively on \( \log(\tau) \)-\( E_{\text{max}} \) values, categories defined by \( \beta \)arr and G protein responses were both correlated to faecaloma report (Supplementary Fig. 10a), implying no differential association of these signals to the undesired event. In contrast, when \( \log(\tau/KA) \) coefficients were additionally considered, faecaloma report correlated to \( \beta \)arr, but not G protein responses (Supplementary Fig. 10b). However, it is worth considering how inclusion of \( \log(\tau/KA) \) values breaks the correlation previously established with efficacy-based categories. BUP has a high transduction coefficient that cannot be distinguished from those of fentanyl (FEN) or loperamide (LOP), causing the partial agonist to move closer to these efficacious ligands in the matrix. Yet BUP’s transduction coefficient is driven by its high affinity\(^{38,39}\), and regardless of its position among efficacious agonists its side effects profile remains determined by its partial efficacy, disrupting the correlation.

Opioid modulation of acute ileum contractility is G protein-driven by effectors that hyperpolarize myenteric neurons and inhibit neurotransmitter release by vagal terminals\(^{40-42}\). We used this G protein-mediated response to ascertain that failure to correlate faecaloma report to categories partly defined by \( \log(\tau/KA) \) was not due to the method itself. In effect, as shown in Supplementary Fig. 10c–e, frequency of faecaloma report correlated with \( \log(\tau) \) but not \( \log(\tau/KA) \) values describing ligand inhibition of ileum contractions.

**Signaling and structural clusters convey complementary information about side effects.** Structural criteria are used to classify, compare, and infer possible commonalities of in vivo responses for drug candidates\(^{43}\). We therefore compared categories of ligands defined by pharmacodynamic and structural criteria. Ligand structure was described using Tanimoto values\(^{44}\) derived from standard fingerprint representations (Supplementary Data 3–5), and these values were then clustered using the same NNMF/k-means method as previously applied on signaling profiles. The resulting clusters are shown in Supplementary Fig. 11a, and representatives of each structural group are provided in Supplementary Fig. 12. Structural and pharmacodynamic similarity matrices were then compared, indicating that 36.0% of changes in ligand distance were compatible with a switch in cluster when the two different criteria were applied. This value was significantly lower when compared with 43.5% switches observed using randomized structural data (\( z \)-score = –2.803; \( p < 0.01 \)), denoting some degree of statistical similarity between signaling and structural categories (Source data provided). However, the degree of similarity was low as schematically.
Drug effects and even withdrawal symptoms are surprising that structural categories associate with complaints, as well as with reports of withdrawal and events, including pruritus, a typical opioid response and drug levels (Supplementary Data 6). Since structure determines pharmacokinetic properties, it is not surprising that structural categories associate with fluctuations in drug effects and even withdrawal symptoms. On the other hand, and in spite of pharmacodynamic properties also being determined by structure, categories based on structural fingerprint representations failed to identify any of the events that are associated with signaling categories, emphasizing the value of complementing structural information with a signal-based classification.

Table 1 Pharmacodynamic and structural categories associate with frequency of report of undesired events for clinically available hMOR ligands

| Undesired events associated with | Type of side effect | Preferred term | R square | p-value |
|---------------------------------|---------------------|----------------|----------|---------|
| Functional categories defined by Log(τ)/Log(τ/KA) | Functional categories defined by Log(τ)/Log(τ/KA) | Gastrointestinal/nutrition | Cachexia | 0.66 | 0.05 |
| | | | Faecaloma | 0.86 | 0.01 |
| | | | Gastrointestinal motility disorder | 0.81 | 0.01 |
| Respiratory events | | | Respiratory depression | 0.62 | 0.11 |
| Sleep disorder | | | Respiratory rate decreased | 0.60 | 0.13 |
| Structural categories | Functional categories defined by Log(τ)/Log(τ/KA) | Gastrointestinal/nutrition | Cachexia | 0.66 | 0.05 |
| | | | Faecaloma | 0.92 | 0.00 |
| | | | Gastrointestinal motility disorder | 0.62 | 0.06 |
| Respiratory events | | | Hypoventilation | 0.68 | 0.08 |
| Sleep disorder | | | Oxygen saturation decreased | 0.61 | 0.12 |
| Respiratory events | | | Respiratory depression | 0.78 | 0.05 |
| Sleep disorder | | | Respiratory rate decreased | 0.69 | 0.08 |
| Respiratory events | | | Somnolence | 0.75 | 0.06 |
| Neuropsychiatric | | | Hypopnoea | 0.65 | 0.10 |
| Change in drug response | | | Yawning | 0.79 | 0.04 |
| Pain | | | Withdrawing syndrome | 0.74 | 0.06 |
| | | | Analgesic drug level increased | 0.70 | 0.08 |
| | | | Drug effect decreased | 0.64 | 0.10 |
| | | | Drug effect increased | 0.90 | 0.01 |
| | | | Therapeutic response increased | 0.89 | 0.02 |
| | | | Therapeutic response decreased | 0.74 | 0.06 |
| Itching | | | Breakthrough pain | 0.69 | 0.08 |
| Others | | | Complex regional pain syndrome | 0.64 | 0.11 |
| | | | Pruritus generalized | 0.86 | 0.02 |
| | | | Therapy cessation | 0.76 | 0.05 |

*Only significant correlations, and correlations that explained 60% or more of the variance were considered; full information in Supplementary Data 2. Similarity and SD gamma scores for Buprenorphine, Fentanyl, Morphine, Oxycodone, and Tramadol were used to establish correlations. Loperamide was additionally included in correlations for gastrointestinal reports. Source data provided in Supplementary Data 2 and source data files.*

Ligand clusters generated with different GPCRs. We next examined whether clustering analysis could reveal pharmacodynamic similarities and differences among ligand responses generated at different opioid receptor subtypes. To do so, we used the same set of biosensors as for hMORs to monitor ligand activity at rat MORs (rMORs), human delta opioid receptor (hDORs) and rat DORs (rDORs). Corresponding input matrices containing logistic and operational parameters for each receptor (Supplementary Data 7–9) were analyzed as before to yield individual similarity matrices and associated heatmaps (Fig. 5a–c). Differences in clustering across receptor subtypes and species were evaluated by comparing similarity matrices for each receptor and are summarized in Fig. 5d. These comparisons revealed that the pattern of signaling diversity of this group of opioid ligands was reasonably conserved within the same receptor from different species. Indeed, in comparisons between rat and human MORs or rat and human DORs, the proportion of changes in ligand distances that were compatible with a switch in cluster was significantly less for actual as compared with randomized data sets (Fig. 5e), confirming congruent patterns across species.

In contrast, when clusters generated with data sets from MORs and DORs within the same species were compared, their differences were statistically indistinguishable from those obtained by comparing corresponding sets of randomized data (Fig. 5f), indicating that the analysis discerned the distinct pharmacological profiles of the two receptor subtypes. To identify the source of these differences, we compared the relative contribution of βarr and G protein responses in driving ligand clustering this time according to hDOR responses. Clusters generated with each partial data set bore statistical similarity to clustering done using the complete data set, and no statistical difference was revealed between clusters produced with βarr and G protein parameters (Fig. 5g). To more precisely establish the weight of βarr and G protein responses to clustering of ligands according to hDOR signaling, we investigated how every value in the similarity matrix changed when considering hDOR clusters generated with data sets from MORs and DORs within the same species were compared, their differences were statistically indistinguishable from those obtained by comparing corresponding sets of randomized data.
more frequent and larger than those for the corresponding βarr comparison (Fig. 5i). A graphical representation of hDOR clusters is given in Supplementary Fig. 13.

Finally, we assessed whether clustering according to signaling profiles could be extended to GPCRs that couple to effectors different than those activated by opioid receptors. For this purpose, we considered published47 and novel data generated with β2-adrenergic receptor (β2AR) ligands including: (a) G protein-dependent responses (Gas activation, cAMP production, Ca²⁺ mobilization)48, (b) βarr-mediated responses (βarr2 recruitment and receptor internalization), and (c) ERK signaling, a multifaceted response involving both G proteins and βarrs49. Parameters describing concentration response curves for each of the readouts (Supplementary Data 10) were analyzed by NNMF and k-means to reveal four different drug categories (Fig. 6a). Cluster #1, including isoproterenol (ISO) and norepinephrine (NE), was characterized by measurable agonist efficacy at all readouts. Salbutamol (SALB) and salmeterol (SALM) in cluster #2 could be distinguished from the first category because of their minimal responses at βarr-dependent readouts. Carvedilol (CARV) and propranolol (PRO) behaved as agonists only in the ERK pathway (Cluster #3), while ICI118,555 and metoprolol (MET) had no efficacy except for inverse agonism at Gas and cAMP assays (cluster #4). The complete signaling profiles for ligands in different clusters are provided in Supplementary Fig. 14.

As shown in Fig. 6b, partial data sets for Gas/cAMP/Ca²⁺ and for βarr-recruitment/endocytosis recreated original clusters better
to reduced sympathetic tone on cardiovascular function. Hence, their more frequent association with ligands that clustered furthest apart from ISO is entirely consistent with gradual loss of efficacy at β2AR. Moreover, these four events were also negatively correlated with ligand efficacy to induce Gas, cAMP, and Ca2+ signaling (Table 2), an overlap that is consistent both with the well-documented role of these signals in maintaining heart chronotropy, and inotropism51 and with the fact that G protein–dependent signals significantly drove ligand clustering. Gas, cAMP, and Ca2+ signaling categories also showed increasing reports of deleterious effects on respiratory function, some of them such as asthma, asthmatic crisis, status asthmaticus consistent with bronchoconstriction, as distance from ISO and β2AR antagonism increased. Thus, pharmacodynamic categories defined by as few as eight ligands were sufficiently robust to reveal a well-known association between sympatholytic CV events or manifestations of bronchoconstriction and modulation of G protein activity by β2AR ligands. The observation reinforces the notion that unsupervised clustering of multidimensional signaling profiles allows the association of signals generated in simple cellular models to possible clinical effects of GPCR ligands.

**Discussion**

This study introduced a stepwise analysis in which GPCR ligands were organized into pharmacodynamic categories that could be then associated with clinically relevant responses. Pharmacodynamic parameters that best supported classification of ligands ((Log(τ)/-E_max⋅Log(τ/KA)) were empirically chosen by consecutively applying NNMF and k-means clustering to different combination of parameters informative of efficacies and functional affinities. The procedure was successfully applied to classify groups of ligands ranging from 8–320 in number, and representative of a multiplicity of signaling profiles.

Ligand categories generated with (Log(τ)/-E_max⋅Log(τ/KA)) values from hMOR data were primarily driven by ligand diversity in βαr signaling efficacy, but G protein responses also contributed to the classification. Measured similarity among signaling profiles of prescription opioids present in the different categories was correlated with their corresponding frequencies of AERS reports for typical opioid side effects, indicating that the categories established by applying this exploratory method may allow to establish meaningful associations between in vitro signals and clinically relevant drug actions. This notion was further supported by observations that pharmacodynamic categories defined for β2AR ligands were essentially driven by G protein responses and associated to G protein-driven sympatholytic effects. Hence, by unveiling these well-documented associations, we established that clustering analysis of concentration–response parameters allows to associate multidimensional in vitro signaling profiles to clinical responses. Such use of curve parameters should prove beneficial for identifying signals that support specific responses of interest and for which mechanistic information is unavailable. Pharmacodynamic categories defined by efficacy-related parameters (Log(τ)/-E_max) had stronger and more frequent correlations to side effects than those defined by additional inclusion of affinity information provided by transduction coefficients. On the other hand, ligand differentiation was optimal when transduction coefficients were taken into account, calling for a discretionary decision on which parameters to use depending on the goal of the classification.

By classifying opioid ligands according to pathway-specific responses, it was possible to explore whether specific signals were driving typical side effects of opioids. We found that association of faecaloma report to categories defined by G protein responses
In contrast to G proteins, log(τ/KA) values were also considered, suggesting a scenario where G protein signaling would not associate to side effects. This divergence as compared with log(τ/KA) matrices is linked to the fact that despite its partial efficacy and a side effects profile consistent with partial agonism, log(τ/KA) coefficients could not distinguish this low-efficacy–high-affinity ligand from much more efficacious agonists, such as morphine or oxycodone. In contrast to G proteins, βarr categories correlated to faecaloma report independent of the parameters used for classification, as transduction coefficients for βarr responses were consistent with BUP’s low efficacy. Log(τ/KA) transduction coefficients are largely used to identify biased ligands17,25, so a word of caution is warranted for bias measures driven by functional affinity since only efficacy parameters are predictive of the magnitude of in vivo responses24.

It is also of interest that maximal responses for βarr and G proteins decreased in parallel across the different clusters, albeit not to the same extent. Indeed, βarr signals gradually disappeared while those of G proteins grew progressively smaller without completely vanishing. Such systematic imbalance between the two types of signals has been previously reported35, and is akin to system bias25, where βarr responses are less well coupled to the receptor than G protein signals. Within this context, absence of βarr responses may simply indicate partial agonism and not signaling bias. As a matter of fact, all novel biased hMOR ligands presented herein as well as those published to date (i.e.: TRV13012, PZM2113 and Scripps compounds14) are partial agonists at G protein responses. This raises the possibility that currently available biased ligands could simply produce less side effects because they are partially effective at stimulating the receptor, and not necessarily because of greater efficacy for activating the G protein over βarr. A miss-interpretation of bias might be a problem for future clinical applications since a partial agonist may also produce a submaximal analgesic response. In this sense, it is of interest that βarr-G protein signaling profiles of the latest hMOR ligands14 resemble those obtained in this study for BUP, a partially effective analgesic36. It is also worth considering that the clinical profile of TRV130, a partial hMOR agonist which was clinically tested as the first biased agonist for MORs, did not significantly differ from morphine’s profile at doses with equivalent analgesic effects19. Finally, when PZM21 was independently tested after its initial description as biased agonist, it was shown to behave as a partial agonist in βarr and G protein readouts, and to produce respiratory depression commensurate with partial signaling10.

Structural similarity is another means for inferring common in vivo responses of therapeutic drug candidates early in the discovery process43. Here, when clusters established on the bases of signaling and structural resemblance were compared, they displayed nonrandom but marginal similarity. Different reasons could explain the low degree of similarity between categories established with structural and pharmacodynamic criteria, including incomplete representation of structural diversity of opioid ligands within the sample used, or different discriminatory power of signaling profiles and current descriptors of structural properties. Consistent with their low degree of similarity, structural, and pharmacodynamic categories were associated with different types of undesired events. Indeed, structural similarities were more frequently associated with fluctuations in therapeutic response, which are typically associated with pharmacokinetic properties42. On the other hand, signaling categories specifically correlated with on-target side effects, pointing to the complementarity of both approaches when characterizing a limited number of compounds of interest.

In conclusion, we presented an unsupervised classification method that incorporates distinct and complementary data sources to comprehensively describe signaling diversity of GPCR ligands. The procedure identifies signaling imbalance independent of whether bias in the response co-varies with efficacy, it was applied to a large diversity of signaling profiles and distinguishes subtle differences in signaling preferences.

### Table 2 Pharmacodynamic categories associate with frequency of report of undesired cardiovascular and respiratory events for clinically available β2ADR ligands

| Correlations between SD gamma scores for prescription β2ADR ligands* and Functional categories defined by Log(τ)-Emax,Log(τ/KA) | Type of side effect | Preferred term | R square | p-value |
|---|---|---|---|---|---|
| Functional categories defined by Log(τ)-Emax,Log(τ/KA) | CV | Hypotension | 0.60 | 0.04 |
| | | Sinus bradycardia | 0.58 | 0.05 |
| | | Blood pressure decreased | 0.64 | 0.03 |
| | | Atrioventricular block | 0.62 | 0.04 |
| | | Sinus arrest | 0.68 | 0.02 |
| | | Inhalation therapy | 0.60 | 0.04 |
| Functional categories defined by Gs/cAMP/Ca2+ (partial matrix) | Respiratory | Heart rate decreased | 0.67 | 0.05 |
| | | Sinus bradycardia | 0.66 | 0.03 |
| | | Blood pressure decreased | 0.64 | 0.03 |
| | | Sinus arrest | 0.82 | 0.00 |
| | | Ejection fraction decreased | 0.82 | 0.01 |
| Respiratory | Respiratory | Asthma | 0.64 | 0.03 |
| | | Inhalation therapy | 0.59 | 0.05 |
| | | Asthmatic crisis | 0.74 | 0.01 |
| | | Status asthmaticus | 0.89 | 0.00 |
| | | Throat irritation | 0.69 | 0.02 |
| | | Choking | 0.64 | 0.03 |

*Only significant correlations were considered, full information in Supplementary Data 11. Similarity and SD gamma scores for Isoproterenol, norepinephrine, salbutamol, salmeterol, propranolol, carvedilol, and metoprolol were used to establish correlations. Source data provided in Supplementary Data 11 and source data files.
Methods
Materials and reagents. Standard opioid agonists were purchased from Cedarlane (Burliungton, Canada) and Sigma-Aldrich (St. Louis, MO, USA). Fifteen novel compounds were provided by Pfizer Inc. (Worldwide Research and Development). (-)-Isoproterenol hydrochloride, (-)-norepinephrine, DL-propanolol hydrochloride, (±) metoprolol (+) tartrate salt, carvedilol, and salmeterol xinafatoate were purchased from Sigma-Aldrich (St Louis, MO). ICI 118,551 and salbutamol hemisulfate were purchased from Tocris Bioscience (Ellisville, MO). Coelenterazine 400a was purchased from Biotium.

Plasmids and DNA constructs. A cleavable signal sequence of influenza hemagglutinin (MKTIIALSYIFCLVFA) and a Flag tag (MDYKDDDDA) were encoding the Epac-based cAMP sensor (GFP10-Epac-RlucII), RlucII-tagged Gαi2, RlucII-tagged GαoA, and RlucII-tagged Gαs-67-RlucII, 100 ng Gα1. The resulting PCR products were subcloned by Gibson assembly in pLVX-IRES-Puro. Constructs assembled by PCR overlap with the coding sequence of GFP10. The resulting PCR products were purchased from Gen-Synthesis DNA Fragments at Gen-eART (ThermoFisher Scientific). The DNA Strings were subcloned by Gibson assembly (New England Biolabs Canada) in pLVX-IREs-Puro (Clontech Laboratories, Inc). Untagged versions of the receptors were made by an internal NcoI deletion, removing the coding sequence of the Flag tag. Constructs encoding for GFP10-tagged receptors were made by PCR overlap; the coding sequence of each signal-peptide Flag-receptors was PCR-amplified to remove the stop codon and assembled by PCR overlap with the coding sequence of GFP10. The resulting PCR products were subcloned by Gibson assembly in pLVX-IRES-Puro. Constructs encoding the Epac-based cAMP sensor (GFP10-Epac-RlucII), RlucII-tagged Gaαi2, α2, α,a, and Gaα1, Gaα10-Gy1, Gaα10-Gy2, Gaαi2-α1-RlucII, Gaαi2-Gy2, human β2-adrenergic receptor (hβ2AR), and cmyc-hα2AR-GFP10 were previously described (PMID: 15782186, PMID: 22534132, PMID: 23175530, PMID: 24309376, PMID: 19843406, PMID: 26658454, PMID: 16091892, PMID: 15155738). pCDNA3.1 (+) Gpl was bought at Missouri University of Science and Technology (cdn.org). Plasmids encoding for the following proteins were generously provided as follows: GRK6 and GRK2 by Dr Antonio De Blasi (Institute Neurologico Mediterraneo Neuromed, Pozzilli, Italy), GRK5 and GRK2 by Dr Robert Leffkowitz (Duke University, Durham, NC), Kir3.2-GFP10 by Dr Terry Hebert (McGill University, Montréal, Canada), Kir3.1 subunit by Dr. Deborah J. Nelson (University of Chicago, Chicago, IL).

Cell culture and transfection. HEK293 cells were a kind gift of Dr. Laporte, McGill University37. They were cultured in 100 mm Petri dishes (Sarstedt, Germany) at 37 °C and 5% CO₂ in the Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum, 2 mM l-glutamine and 100 unit mL⁻¹ penicillin-streptomycin. Constructs encoding for vectors encoding BRET biosensors in combination with complementary signaling partners were performed in 100 mm Petri dishes (3 × 10⁶ cells) for G protein and Kir3.2 channel activation assays and in 96-wells culture plates coated with polyD-lysine (PerkinElmer, MA, USA) for β-arrestin recruitment, HEK 293 cells were co-transfected with sp-FLAG-DOR-GFP10/β-arrestin2-RlucII, GFP10-Epac-RlucBRET2-cAMP biosensor59 were subsequently distributed into the 96-well-printed plates. Plates were read 5 min to the wells (2.5 nM Coelenterazine 400a was added to a final concentration of 30,000 cells/well in a high glucose medium supplemented with 10% newborn calf serum, and grown on 96-well polystyrene-coated plates for 48 h. Cells were transferred to Tyrode buffer and incubated for 15 min at 37 °C. The cells were washed, incubated with 0.1% BSA and then re-suspended in PBS then incubated in Tyrode buffer and incubated for 15 min at 37 °C.  Coelenterazine 400a was then added to a final concentration of 5 µM. Five min later, forskolin (Bioshop, Canada) was introduced (final concentration: 10 µM for MOR, 15 µM for DOR, and 25 µM for hMOR and hDOR) followed, 3.5 min later, by increasing concentrations of adrenergic compounds were added for 4.5 min. The day of the experiment, cells were washed with PBS then re-suspended in PBS then incubated at a final concentration of 30,000 cells/well in a high glucose medium supplemented with 10% newborn calf serum, and grown on 96-well polystyrene-coated plates for 48 h. Cells were transferred to Tyrode buffer and incubated for 15 min at 37 °C. Coelenterazine 400a was then added to a final concentration of 5 µM. Five min later, forskolin (Bioshop, Canada) was introduced (final concentration: 10 µM for MOR, 15 µM for DOR, and 25 µM for hMOR and hDOR) followed, 3.5 min later, by increasing concentrations of adrenergic ligands. BRET2 readings were taken 5 min after ligands were introduced38.

Guinea pig ileum assays. Male Hartley guinea pigs were anesthetized using isoflurane followed by exsanguination. The myenteric plexus of the ileum was dissected according to the method described by Cowie & al.41. Briefly, a portion of the ileum was removed (10 cm distal to the cecum) into which a glass rod was inserted. The myenteric plexus was removed from the circular longitudinal muscle via gentle scraping with a moist cotton swab and separated from the muscle using forceps. The resulting myenteric plexus tissue was cut into 2.5 mm strips and placed in oxygenated Krebs buffer (37 °C, gassed with 95% O₂/5% CO₂) and tensioned to a baseline tension of 2000 mg. The tissues were washed, equilibrated for 30 min, and subsequently tested for viability with a maximal concentration of Carbachol (300 nM, three times with 10 min of washing, and 10 min of equilibrating in between additions). The final prime was followed by a 20 min wash period followed by a 20 min equilibration period before the start of the experiment. Tissues were continuously stimulated with 0.1 Hertz for 1 ms at 20 volts (producing a stimulation equivalent to 80% of the maximal contractile response). Followings a 10 min baseline stimulation period, the kappa opioid antagonist nor-binaltorphimine was added (10 nM) and incubated for 10 min. Finally, cumulative concentration–response curves were generated to each test ligand or vehicle control (DMSO). Isometric tension data (in mg) were collected. All procedures performed on these animals were in accordance with regulations and established guidelines and were reviewed and approved by Pfizer Institutional Animal Care and Use Committee.

Curve fitting. Concentration response curves describing ligand responses by different receptors (hMOR, hDOR, rMOR, rDOR, and hβ2AR) were analyzed using Graphpad Prism6, using built-in 3 or 4 parameter logistic equations to obtain independent pEC50 and E_max values for each receptor–biosensor pair:

\[
y = a + (b - a) \left( 1 + 10^{(EC50 - x)/C16} \right)
\]

\[
y \rightarrow \text{measured response}; \ a \rightarrow \text{minimal asymptote}; \ b \rightarrow \text{maximal asymptote}; \ c \rightarrow \text{slope}.
\]

Concentration response curves were additionally analyzed with the operational model of Black and Leff52. As above, curves representing responses elicited by the same receptor at each of the ten different biosensors were fit independently. Fitting was done using Graphpad Prism6, using built-in 3 or 4 parameter logistic equations to obtain independent pEC50 and E_max values for each receptor–biosensor pair:

\[
y = a + (b - a) \left( 1 + 10^{(EC50 - x)/C16} \right)
\]

\[
y \rightarrow \text{measured response}; \ a \rightarrow \text{minimal asymptote}; \ b \rightarrow \text{maximal asymptote}; \ c \rightarrow \text{slope}.
\]

Concentration response curves were additionally analyzed with the operational model of Black and Leff52. As above, curves representing responses elicited by the same receptor at each of the ten different biosensors were fit independently. Fitting was done using Graphpad Prism6, using built-in 3 or 4 parameter logistic equations to obtain independent pEC50 and E_max values for each receptor–biosensor pair:

\[
y = a + (b - a) \left( 1 + 10^{(EC50 - x)/C16} \right)
\]

\[
y \rightarrow \text{measured response}; \ a \rightarrow \text{minimal asymptote}; \ b \rightarrow \text{maximal asymptote}; \ c \rightarrow \text{slope}.
\]
Full agonist =
\[
\text{basal} + (\text{Emax} - \text{basal})/(1 + \text{operate1})
\] (4)

Partial agonist =
\[
\text{basal} + (\text{Emax} - \text{basal})/(1 + \text{operate2})
\] (5)

baseline → response observed in the absence of agonist; Emax → maximal response of the system; n → slope of the function which links occupancy to response; KA → association constant; LogE maxim; LogE minimized

When using the logistic model, the fits for three and four parameter curves were compared and the best fit taken. If no fitting was possible without constraints, the minimal asymptote was fixed to zero; if this was unsuccessful, the Hill coefficient was additionally fixed to one (i.e.: only the three parameter fit was considered). If both these constraints proved unsuccessful, and in curves with no inflection point for maximal effect, the highest experimental value was considered Emax. The latter procedure forced the maximal response of very weak partial agonists within the range of experimental data avoiding aberrant predictions due to extrapolation. If no fitting was possible following these constraints, no fitting (NF) status was consigned. If fitting was possible, we made sure that all curves had a Span > 3x SEM, otherwise they were considered as no response (NR).

As used in this study, the operational model does not yield Log(t) or pKa values for full agonists\(^{25}\), which were consigned as not available (NA). In these circumstances, Emax values were used to differentiate these compounds from partial agonists, and differences among full agonists were established through their consolidated Log(t/Ka) coefficients. It should also be noted that by independently fitting curves for different biosensors, the model does not contemplate interconversion among distinct receptor states stabilized by different effectors.

**Feature reduction, and ligand clustering.**

1. Each receptor was represented by a matrix composed of 25 ligands (21 for DRePs × 30 parameters (Emax, LogEmax, and LogE/Ka) for ten assays). This matrix was created by sampling from the normal distribution around each parameter using the mean and standard deviation thereby incorporating the variance associated with each data point and propagating it through the clustering method. In order to correct for scale differences between parameters, we standardized each column to range between 0 and 1 according to:

\[
\text{Standardized value} = \frac{X_{ij} - \text{minimum} \_i}{\text{maximum} \_i - \text{minimum} \_i}
\] (6)

for every ligand \( i \) and every parameter \( j \).

2. Process (1) was repeated 1000 times to create 1000 data matrices each independently put through the following procedure (Supplementary Fig. 2).

3. Nonnegative matrix factorization (NNMF) was used to reduce dimensionality of the data and create the W (ligand × k) and V (k × parameter) basis vectors thereby improving noise and redundancy efficiency. We used sparse NNMF to ignore missing data (“NA”, “NF”, and “NR” curves). Difference between the original matrix V and the product of W × H was minimized to less than 1e-7.

4. K-means clustering was performed on the W basis vector, where the number of clusters equals the number of basis vectors from NNMF (K = k), to assign each parameter cluster to a k.

5. Steps 3 and 4 were repeated 250 times to quantify the fraction of times each compound clustered together resulting in a ligand × ligand frequency matrix ranging from 1 (always clustered together) to 0 (never clustered together). This iterative process quantifies both global and local minima/maxima arising from small variances in clustering resulting from the randomized starting vectors for NNMF and k-means.

6. (a) The entire process including feature reduction and clustering (3–5) was repeated for different values of k (k = 2 to k = 7), providing a frequency matrix for each k. (b) These six frequency matrices were averaged together to quantify ligand similarity independent of the number of features used as each K may extract unique patterns that may be complementary or orthogonal to results from different Ks.

7. Steps 3–6 were independently performed on each of the 1000 sampled data matrices providing 1000 composite similarity matrices.

8. These 1000 matrices were averaged to create a final frequency matrix quantifying how often ligands clustered together and representing total ligand similarity across all concentration response curves.

9. We visualized the similarity matrix using a dendrogram and a heat map using Orange, created from the distance between each compound in the similarity matrix using a Pearson Correlation.

**Simulation of virtual compounds.** We built 16 profiles showing bias and various potencies/efficacies by selecting ranges of KA-τ pairs across six imaginary biosensors. So that our virtual compounds respect these ranges, we invented them by sampling random values of KA and τ within the bounds associated to the imaginary biosensors specified per profile. We used this procedure 20 times for each profile yielding 320 virtual ligands. As for curve fitting, simulations were conducted under the assumption of independence across biosensors. Using the operational model equation, we generated corresponding concentration response curves (CRC) to which we added 10% noise using the flat distribution. Noisy CRCs were then fitted to both the logistic equation and the operational model equation using the Bayesian inference engine STAN\(^{26}\) to yield values of Emax, pEC50, τ, and KA and their associated distributions (from which we computed a standard error of the mean to use in the NNMF pipeline) (values in Supplementary Data 12). The best estimate for τ/Ka ratio and its distribution of draws were also computed directly within the fitting process as a transformed parameter.

Selections of subsets of parameter estimations and associated SEM were used in NNMF/K-means clustering. The resulting matrices of frequency of co-occurrences were used to compute distributions, hierarchical clustering trees, and TSNE plots whose leaves and data points were colored by profile (Fig. 1, Supplementary Fig. 3, Supplementary Fig. 4).

**Clustering of pharmacological parameters.** The 25 (1 per ligand) values for each parameter array (P) were distributed into four smaller arrays corresponding the ligand clusters. We then utilized a two-sample Kolmogorov–Smirnov test to compare each sub-array to the original array (P) to measure if these were randomly sampled from the original array. This provides p-values for each parameter. A significant p-value indicates ligands in that cluster are biased toward a specified response for that parameter. This process was repeated for each of the 30 parameters. These p-values were then sorted according to (i) the type of parameter considered (e.g., pEC50, Emax, or LogE(Ka)) or (ii) measurement similarity acquired from the similarity matrix obtained by the NNMF/K-means method detailed above using the H basis vector instead of the W. The procedure is summarized in Supplementary Fig. 7.

**Comparing clusters among complete data sets.** To compare clustering similarity between two different data sets we implemented two approaches: (a) Directly comparing the two similarity matrices using pairwise differences and (b) quantifying the overall difference between the two matrices using random simulation to obtain a difference threshold and to establish significance.

Direct comparison: We calculated the difference between every paired value in similarity matrix A and B (representing the similarity between compounds i and j):

\[
\text{Difference} = |A_{ij} - B_{ij}|
\] (7)

The resulting difference matrix is of equal dimensionality to A and B, ranging from 1 (compounds i and j are always clustered together in A but never in B) to −1 (always in B but never in A).

Thresholding and random simulation: We compared the difference in Euclidian distance for every pair of ligands i and j between similarity matrices for data set A and B:

\[
\text{Difference} = |L_{A_{ik}} - L_{A_{jk}}| - |L_{B_{ik}} - L_{B_{jk}}|
\] (8)

Where L_{A_{ik}} and L_{A_{jk}} are row vectors representing the similarity of I_k and L_j to all other ligands in matrix A. We then used random clustering replicates (detailed below) to identify a cutoff value to determine which difference values corresponded to a significant variation between A and B. The final comparison between data set A and B was represented as a proportion:

\[
\text{Fraction change} = \frac{\# \text{significant differences}}{\text{Total}\# \text{of comparisons}}
\] (9)

Only ligands tested in both data sets were used (e.g., comparing hMOR to hDOR only used the 21 shared ligands).

Random clustering: For each data set, we created 50 random input data matrices by permuting all mean-standard deviation pairs of data points within the original data matrix. Each random matrix was therefore specific to and equal in size and shape to the original data (e.g., hMOR: (25 * 30); hMOR-jbrr: (25 * 15)). We then repeated the entire NNMF/K-means clustering method on each data-shuffled random matrix resulting in 50 random clustering frequency matrices for each data set type.

To determine a cutoff value representing significant variation between any two data sets (Difference threshold), we calculated the Euclidian Distance between pairs of compounds in the same cluster and compounds in different clusters for each of the 50 trials. The threshold is the mean value of the overlap between the “same cluster” distance and the “different cluster” distribution (see Supplementary Fig. 9). Threshold values range between 0.95 and 1.5. Using these thresholds, it was possible to calculate the proportion of significant variation between two matrices. To quantify if this change was significant, we calculated the fraction of significant changes (using thresholds) between the clustering from the 50 randomized data sets (e.g., 50 random hDOR) compared with reference cluster (e.g., hMOR). The resulting distribution of 50 values represented the proportion of random changes from the reference. This distribution was used to calculate a z-score for the
Comparing clusters generated with complete data and subsets. In order to calculate whether clustering from data subset 1 (e.g., hMOR-Barr) changed more than data subset 2 (e.g., hMOR-G protein) compared with the complete data set clusters (e.g., hMOR), we compared the 50 random similarity matrices to the reference (e.g., hMOR) and calculated the fraction of significant changes using the method detailed above. As a result, we obtained an array of 50 values representing the random change from reference. This array was created for both subsets (e.g., hMOR-Barr and hMOR-G protein). We then iteratively, with replacement, randomly sampled 1 value from each of these arrays and calculated the difference to create a distribution of 1000 values indicating the random expected difference between these two subsets of data. We then calculated a z-score using the mean and standard deviation of this distribution and the actual observed difference.

Clustering ligands according to structural similarities. Each ligand was represented using three standard fingerprint representations: (ECFP-6) Extended-Connectivity Fingerprints (ECFPs) (http://accelrys.com/products/collaborative-science/biovia-pipeline-pilot/), Functional-Class Fingerprints (FCFPs) and MDL MACCS keys. A similarity matrix for each different fingerprint was generated for the 25 ligands in the data set, where each value in the matrix \(s_{ij}\) corresponds to the Tanimoto similarity value between compound \(i\) and compound \(j\) and ranges from a value 0 to 1 (1 being most similar). We combined these three matrices into a single matrix of dimensions (25 compounds × 75 comparisons), and repeated the NNMF/k-means clustering algorithm on the data to yield a structural similarity matrix.

Correlating signaling data to side effect report frequency. A list of all MOR-active compounds was created by searching DrugBank for all approved drugs which activate MOR. The resulting list was intersected with the list of drugs in the FDA’s Adverse Event Reporting System data for which a standardized gamma (SD gamma) score could be generated at the preferred term (PT) level according to the method of Johnson et al.\textsuperscript{33}. Briefly, SD gamma scoring is a statistical approach to identify disproportionately high, or low, numbers of drug-event occurrences by normalizing for number of drugs and number of event reports. SD gamma scores for each event were averaged for all resulting MOR compounds, and PT events were sorted by average score to produce a listing of high-scoring events most clinically relevant to opioid therapy (80 highest scores were considered). A similar procedure was completed to find the 80 side effects associated with \(\beta2\)AR-active compounds.

Individual drug SD gamma scores for frequently reported events were then correlated to Euclidian distances separating prescription opioids (tramadol, buprenorphine, oxycodone, morphine, fentanyl, and loperamide) in hMOR and structural similarity matrices. SD gamma scores were additionally correlated to transduction coefficients for BRET or guinea pig contractility responses respectively normalized to Met-ENK (\(\log(K_{IC50})\)) or loperamide (\(\log(t/K_{IC50})\)). Note that LOP and Met-ENK are balanced ligands that co-cluster in every data set such that differences due to normalization are simply scalar. Individual drug SD gamma scores clinically prescribed \(\beta2\)AR ligands used in the study (isoproterenol, norepinephrine, salbutamol, salmeterol, pindolol, carvedilol, and metoprolol) were similarly correlated to Euclidian distances separating these ligands in the \(\beta2\)AR similarity matrix or to transduction coefficients for BRET responses where isoproterenol was the standard.

Statistical analysis. Correlation analysis: GraphPad Prism\textsuperscript{5} was used to evaluate correlation between drug distance in cluster and the frequency of reports of adverse events. All statistical comparisons were two-sided except when contrasting partial and whole similarity matrices.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. All data generated or analyzed in this study are included in the article and supplementary materials or provided as source data files.

Code availability. All clustering and cluster comparisons were conducted using Python 2.7.6. Complete source code is available for download at http://github.com/JonathanGallion/Benredjem-Gallion.

References

1. Lafferty-Whyte, K., Mormenoe, D. & Del Fresno Marimon, M. Trial watch: opportunities and challenges of the 2016 target landscape. Nat. Rev. Drug. Disc. 16, 10–11 (2017).
2. Lohse, M. J., Nuber, S. & Hoffmann, C. Fluorescence/bioluminescence resonance energy transfer techniques to study G-protein-coupled receptor activation and signaling. Pharm. Rev. 64, 299–336 (2012).
3. Huang, W. et al. Structure insights into micro-opioid receptor activation. Nature 524, 315–321 (2015).
4. Kenakin, T. & Miller, L. J. Seven transmembrane receptors as shapeshifting proteins: the impact of allosteric modulation and functional selectivity on new drug discovery. Pharm. Rev. 62, 265–304 (2010).
5. Kenakin, T. Signaling bias in drug discovery. Expert Opin. Drug Disc. 12, 321–333 (2017).
6. Wacker, D., Stevens, R. C. & Roth, B. L. How ligands illuminate GPCR molecular pharmacology. Cell 170, 414–427 (2017).
7. Luttrel, L. M., Maudslay, S. & Gesty-Palmer, D. Translating in vitro ligand bias into in vivo efficacy. Cell Signal 41, 46–55 (2018).
8. Costa-Neto, C. M., ESLT, Parreiras & Bouvier, M. A pluridimensional View of Biased Agonism. Mol. Pharm. 90, 587–595 (2016).
9. Pang, P. S. et al. Biased ligand of the angiotensin II type 1 receptor in patients with acute heart failure: a randomized, double-blind, placebo-controlled, phase IIb, dose ranging trial (BLAST-AHF). Eur. Heart J. 38, 2364–2373 (2017).
10. Hill, R. et al. The novel mu-opioid receptor agonist PZM21 depresses respiration and induces tolerance to anticoagulation. Br. J. Pharm. 175, 2653–2661 (2018).
11. Singla N. et al. A randomized, Phase Ib study investigating olodaterol (TRV130), a novel µ-receptor G-protein pathway selective (µ-GPS) modulator, for the management of moderate to severe acute pain following abdominoplasty. J Pain Res. 10, 2413–2424 https://doi.org/10.2147/JPR.S147982 (2017).
12. DeWire, S. M. et al. A G protein-biased ligand at the mu-opioid receptor is potently analgesic with reduced gastrointestinal and respiratory dysfunction compared with morphine. J. Pharm. Exp. Ther. 344, 708–717 (2013).
13. Manligik, A. et al. Structure-based discovery of opioid agonists with reduced side effects. Nature 537, 185–190 (2016).
14. Schmid, C. L. et al. Bias factor and therapeutic window correlate to predict safer opioid agonists. Cell 171, 1165–1175 e1113 (2017).
15. Onaran, H. O. et al. Systematic errors in detecting biased agonism: analysis of current methods and development of a new model-free approach. Sci. Rep. 7, 44247 (2017).
16. Benredjem, B., Dallaire, P. & Pineyro, G. Analyzing biased responses of GPCR ligands. Curr. Opin. Pharmac. 32, 71–76 (2017).
17. Kenakin, T., Watson, C., Muniz-Medina, V., Christopoulos, A. & Novick, S. A simple method for quantifying functional selectivity and agonist bias. ACS Chem. Neurosci. 3, 193–202 (2012).
18. Rajagopal, S. et al. Quantifying ligand bias at seven-transmembrane receptors. Mol. Pharm. 80, 367–377 (2017).
19. Lane, J. R., May, L. T., Parton, R. G., Sexton, P. M. & Christopoulos, A. A kinetic view of GPCR allosteric and biased agonism. Nat. Chem. Biol. 13, 929–937 (2017).
20. Kenakin, T. Quantifying biological activity in chemical terms: a pharmacology primer to describe drug effect. ACS Chem. Biol. 4, 249–260 (2009).
21. Raechal, K. M., Walker, J. K. & Bohn, L. M. Morphine side effects in beta-arrestin 2 knockout mice. J. Pharm. Exp. Ther. 314, 1195–1201 (2005).
22. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791 (1999).
23. Forgay, E. & Clusker, C. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. Biometrica. 21, 768–769 (1965). JSTOR 2528559.
24. Kenakin, T. & Christopoulos, A. Measurements of ligand bias and functional affinity. Nat. Rev. Drug Disc. 12, 483 (2013).
25. Kenakin, T. & Christopoulos, A. Signalling bias in new drug discovery: detection, quantification and therapeutic impact. Nat. Rev. Drug. Disc. 12, 205–216 (2013).
26. Gales, C. et al. Probing the activation-promoted structural rearrangements in preassembled receptor-G protein complexes. Nat. Struct. Mol. Biol. 13, 778–786 (2006).
27. Richard-Lalonde, M. et al. Conformational dynamics of KIR3.1/KIR3.2 channel activation via delta opioid receptors. Mol. Pharm. 83, 416–428 (2013).
28. Tsubaki, K., H., Robertson, D. N., Schiller, P. W. & Pineyro, G. Endocytic profiles of delta-opioid receptor ligands determine the duration of rapid but not sustained cAMP responses. Mol. Pharm. 85, 148–161 (2014).
29. Salahpour, A. et al. BRET biosensors to study GPCR biology, pharmacology, and signal transduction. Front Endocrinol. (Lausanne) 3, 105 (2012).
30. Charki, L. et al. Ligand- and cell-dependent determinants of internalization and cAMP modulation by delta opioid receptor (DOR) agonists. Cell Mol. Life Sci. 71, 1529–1546 (2014).
31. Winpenny, D., Clark, M. & Cawkill, D. Biased ligand quantification in drug discovery: from theory to high throughput screening to identify new biased mu opioid receptor agonists. Br. J. Pharmac. 173, 1393–1403 (2016).
32. Black, J. W. & Leff, P. Operational models of pharmacological agonism. Proc. R. Soc. Lond. B Biol. Sci. 220, 141–162 (1983).
33. Johnson, K., Guo, C., Gosink, M., Wang, V. & Hauben, M. Multinomial modeling and an evaluation of common data-mining algorithms for identifying pig-ialls at disp. Pharmaceut. reporting in pharmacovigilance databases. Bioinformatics 28, 3123–3130 (2012).
34. Kalso, E., Edwards, J. E., Moore, R. A. & McQuay, H. J. Opioids in chronic non-cancer pain: systematic review of efficacy and safety. Pain 112, 372–380 (2004).
35. Ballantyne, J. C. Opioid analgesia: perspectives on right use and utility. Nature Communications 10:4075 (2019) https://doi.org/10.1038/s41467-019-11875-6 www.nature.com/naturecommunications
36. Rajagopal, S. Quantifying biased agonism: understanding the links between affinity and efficacy. Nat. Rev. Drug Disc. 12, 483 (2013).
37. Onaran, H. O., Rajagopal, S. & Costa, T. What is biased efficacy? Defining the relationship between intrinsic efficacy and free energy coupling. Trends Pharmac. Sci. 35, 639–647 (2014).
38. Pilgrim, F. J. Binding of buprenorphine to opiate receptors. Regul. guanyl nucleotides Met. Ions. Neuropharmacal. 23, 373–375 (1984).
39. Brown, S. M., Holtzman, M., Kim, T. & Kharasch, E. D. Buprenorphine metabolites, buprenorphine-3-glucuronide and norbuprenorphine-3-glucuronide, are biologically active. Anesthesiology 115, 1251–1260 (2011).
40. Cowie, A. L., Kosterlitz, H. W., Lydon, R. J. & Waterman, A. Regulating myocardial calcium channels by cyclic AMP metabolism. J. Stat. Softw. 76, 1–32 (2017).
41. Galligan, J. J. & Sternini, C. Insights into the role of opioid receptors in the GI tract: experimental evidence and therapeutic relevance. Handb. Exp. Pharmac. 239, 363–378 (2017).
42. Cowie, A. L., Kosterlitz, H. W., Lydon, R. J. & Waterfield, A. A. The effects of morphine-like substances and their antagonists on transmission at the neuro-effector junction of the myenteric plexus-longitudinal muscle preparation of the rat. J. Pharm. Pharmacol. 36, 639–640 (1983).
43. Yamanishi, Y., Pauwels, E. & Kotera, M. Drug side-effect prediction based on the integration of chemical and biological spaces. Discov. Med. 12, 948–952 (2013).
44. Willett P. Similarity-based virtual screening using 2D fingerprints. Drug Discov Today 11, 1046–1053 (2006).
45. Cowie, A. L., Kosterlitz, H. W., Lydon, R. J. & Waterfield, A. A. The effects of morphine-like substances and their antagonists on transmission at the neuro-effector junction of the myenteric plexus-longitudinal muscle preparation of the rat. J. Pharm. Pharmacol. 36, 639–640 (1983).
46. Cummings, J. G., Davis, A. M., Muresan, S., Haebelerlein, M. & Chen, H. Chemical predictive modelling to improve compound quality. Nat. Rev. Drug Disc. 12, 948–962 (2013).
47. van der Westhuizen, E. T., Breton, B., Christopoulos, A. & Bouvier, M. Quantification of ligand bias for clinically relevant beta2-adrenergic receptor ligands: implications for drug taxonomy. Mol. Pharm. 85, 492–509 (2014).
48. Caffrey, M. & Johnson, K. Opioid tolerance to a 24-hour dose of morphine: a meta-analysis of randomized controlled trials. J. Clin. Pharmacol. 46, 378–390 (2006).
49. Finney, P. A. et al. Albuterol-induced downregulation of Gsalpha accounts for pulmonary beta(2)-adrenoceptor desensitization in vivo. J. Clin. Invest. 106, 125–135 (2000).
50. Charkoudian, N. & Rabbitts, J. A. Sympathetic neural mechanisms in human cardiovascular health and disease. Mayo Clin. Proc. 84, 822–830 (2009).
51. Hove-Madsen, L., Mery, P. F., Jurevicus, J., Skeberdis, A. V. & Fischmeister, R. Regulation of myocardial calcium channels by cyclic AMP metabolism. Basic Res. Cardiol. 91 Suppl. 2, 1–8 (1996).
52. Prins, S. M., Hoogendijk, A. & Kosterlitz, H. W. The role of opioids in chronic non-cancer pain: a systematic review of Cochrane Reviews. Pain 151, 373–379 (2017).
53. Brown, S. M., Holtzman, M., Kim, T. & Kharasch, E. D. Buprenorphine metabolites, buprenorphine-3-glucuronide and norbuprenorphine-3-glucuronide, are biologically active. Anesthesiology 115, 1251–1260 (2011).
54. Ballantyne, J. C. Opioid analgesia: perspectives on right use and utility. Nature Communications 10:4075 (2019) https://doi.org/10.1038/s41467-019-11875-6 www.nature.com/naturecommunications
55. M.B., P.D., and M.G. wrote the paper. B.B., J.C., J.G., K.N., M.G., C.L., S.J., V.L., C.S., D.P., B.M., Y.R., P.D., and E.V.W. performed experiments and/or data analysis.
56. Raffa, R. B. et al. The clinical analgesic efficacy of buprenorphine. J. Clin. Pharm. Ther. 39, 577–583 (2014).
57. Robertson, D. N. et al. Design and construction of conformational biosensors to monitor ion channel activation: a prototype FIAH/BRET-approach to Kir3 channels. Methods 92, 19–35 (2016).
58. Bousif, O. et al. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethyleneimine. Proc. Natl Acad. Sci. USA 92, 7297–7301 (1995).
59. Audet, N. et al. Differential association of receptor-G betaarrestin complexes with beta-arrestin2 determines recycling bias and potential for tolerance of delta opioid receptor agonists. J. Neurosci. 32, 4827–4840 (2012).
60. Carpenter, B. et al. Stan: a probabilistic programming language. J. Stat. Softw. 76, 1–32 (2017).

Acknowledgements
This research was supported by a research contract from Pfizer Inc. and grants from the Natural Sciences and Engineering Research Council of Canada (Grant 311997 to G.P.) and the Canadian Institutes of Health Research MOP 124876 (to G.P.), MOP 102630 (to M.B. and O.L.) and Foundation grant (FDN-148431) to MB. MB holds a Canada Research Chair in Signal Transduction and Molecular Pharmacology. Dr Lichtarge’s research was supported by National Institutes of Health (NIH 2R01 GM066999; NIH 5R01 GM079656). B.B. was supported by a studentship from Fonds de Recherche en Santé du Québec. P.D. was supported by a MITACS fellowship. This study was supported by the Quebec Pain Research Network (QPRN) of the Fonds de recherche du Québec Santé.

Author contributions
G.P., M.B., B.B., J.G., A.S., D.C., C.L., P.D., and O.L. conceived the study. B.B., J.G., G.P., M.B., P.D., and M.G. wrote the paper. B.B., J.G., J.K., N.K., M.G., C.L., S.J., V.L., C.S., D.P., B.M., Y.R., P.D., and E.V.W. performed experiments and/or data analysis.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-11875-6.
Competing interests: M.B., C.L., and G.P. have a patent on BRET biosensors licensed to Domain Therapeutics North America and Pfizer Inc. The remaining authors declare no competing interests.

Peer review information: Nature Communications thanks Stephen Duffull and other anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access
This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019