Combinatorial expression of functionally distinct GPCR isoforms can diversify receptor signalling response

Maria Marti-Solano[1,*], Stephanie E. Crilly[2], Duccio Malinverni[1], Christian Munk[3], Matthew Harris[4], Abigail Pearce[4], Tezz Quon[5], Amanda E. Mackenzie[5], Xusheng Wang[6,7], Junmin Peng[6,8], Andrew B. Tobin[5], Graham Ladds[4], Graeme Milligan[5], David E. Gloriam[3], Manojkumar A. Puthenveedu[2,9], M. Madan Babu[1,*]

[1]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK.
[2]Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI 48109, USA
[3]Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark.
[4]Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1PD, United Kingdom.
[5]Centre for Translational Pharmacology, Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, Scotland, UK.
[6]Center for Proteomics and Metabolomics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.
[7]Department of Biology, University of North Dakota, Grand Forks, ND 58202, USA.
[8]Departments of Structural Biology and Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.
[9]Department of Pharmacology, University of Michigan Medical School, Ann Arbor, MI 48109, USA.

*Email: mmarti@mrc-lmb.cam.ac.uk OR madanm@mrc-lmb.cam.ac.uk

Summary

G protein-coupled receptors (GPCRs) are transmembrane proteins that modulate physiology across diverse tissues in response to extracellular signals. GPCR signalling can differ due to variation in the sequence (e.g. polymorphisms) or in the expression of receptors in different tissues. The resulting differences in response are an important source of physiological signalling bias. An underexplored source of such bias is the generation of functionally diverse GPCR isoforms that can have distinct patterns of expression in human tissues. Here, we report the findings from a comprehensive study, integrating data from human tissue-level transcriptomes, GPCR sequences and structures, functional annotations, proteomics, single-cell RNA sequencing, population-wide genetic association studies, and pharmacological experiments. Our results show how a single GPCR gene can diversify into multiple isoforms with distinct structural and signalling properties, and how unique combinations of these isoforms can be expressed in different human tissues, contributing to differences in physiological signalling. Based on their structural changes and expression patterns, some of the detected isoforms may also influence drug response and represent new drug targets with improved tissue selectivity. Our findings highlight the need to move from a canonical to a context-specific view of GPCR signalling, in which one considers how the combinatorial expression of receptor isoforms in a specific system (i.e. a particular cell type, tissue, or organism) collectively impacts receptor signalling. These observations pave the way for understanding the impact of isoform variation on GPCR signalling response and have implications for exploiting such variation as a source of GPCR selectivity in drug development.

Introduction

G protein-coupled receptors (GPCRs) are a diverse family of transmembrane receptors that are widely expressed across human tissues. They govern multiple physiological responses by binding to extracellular ligands and transducing their signal to a variety of intracellular coupling partners. Our understanding of GPCRs as modulators of human physiology has been complicated by the phenomenon of signalling bias, which results in variation in downstream responses due to the differential coupling of receptors with their associated intracellular signalling partners[1,2]. For instance, comparing receptor activation by endogenous and non-endogenous ligands, or in different systems such as cell types, tissues or organisms (i.e., in diverse cellular and physiological contexts), has revealed differences in pharmacological responses[1,2]. While this phenomenon was first described through the characterisation, and later development, of GPCR ligands capable of stabilising different receptor conformations to bias downstream signalling[1,3], the mechanisms that drive physiological signalling bias are less well understood. Two key contributors to this signalling bias are changes in receptor sequences arising from genetic variation (e.g. receptor polymorphisms[4,5]), and the system-specific expression of GPCR signalling components[6].

An important source of functional diversity is the generation of protein isoforms by mechanisms such as tissue-specific alternative splicing, and through the use of alternative transcription start and termination sites[7,8] (**Fig. 1a**; reference vs. non-reference isoforms). Some GPCRs are known to have multiple isoforms with distinct signalling properties[9–12] and alternative splicing events are known to affect different receptor regions[13]. However, the extent to which receptor isoforms generate structural and functional diversity, their pattern of expression in different human tissues, and how combinatorial isoform expression can generate diverse signalling states, thus contributing to physiological signalling bias, is fundamentally unexplored. Considering over one-third of the drugs in the clinic target this receptor family[14,15], understanding how isoform diversity can influence function is not only important to understand receptor physiology, but also to assess how this diversity may affect GPCR drug response.

In this work, through a comprehensive data analysis approach, we have obtained a receptor-wide view of GPCR isoform diversity in humans. Our results show how non-olfactory human GPCRs of different classes can exhibit diversity in receptor structural segments involved in interaction with extracellular ligands and with intracellular signalling partners. They also provide an isoform expression map across human tissues, that reveals how tissue-specific combinations of GPCR isoforms with distinct ligand binding and pharmacological properties could contribute to system-specific receptor signalling. The resulting characterisation of GPCR isoform diversity has been made accessible through the GPCRdb web-resource[16] (currently available at https://gpcrdb.org/protein/isoforms). This can serve as a roadmap to assess the isoform diversity for any given receptor and understand how combinatorial receptor isoform expression can determine signalling response and contribute to physiological signalling bias.

Structural changes in receptor isoforms diversify receptor function

Isoform diversity exists across different GPCR classes

To assess the prevalence of multiple receptor isoforms in all non-olfactory GPCRs, we combined transcript-level mRNA expression data from the Genotype-Tissue Expression (GTEx) Consortium[17,18], which includes gene expression information from 30 human tissues of ~700 donors, with sequence and structural annotations of receptors from the GPCRdb[16] (**Fig. 1b**). We then filtered transcripts by expression level, by annotation as protein-coding in Ensembl[19], and, where applicable, by the integrity of the signal peptides required for GPCR synthesis in the endoplasmic reticulum (ER), as annotated in UniProt[20]. In order to ensure that all isoforms in our study were transmembrane proteins, we created receptor-specific multiple sequence alignments of all the isoforms found for every GPCR. These alignments were then used to select only those receptor isoforms with one or more conserved transmembrane segments, as annotated in GPCRdb (**Fig. 1b**, **Methods**). In this manner, we identified 363 different receptors and 625 distinct GPCR isoforms in the GTEx dataset (that is, 363 reference isoforms corresponding to GPCRdb receptor sequences, and 262 non-reference isoforms). 136 receptors (38%) have two or more isoforms that are expressed in four or more GTEx donors (see **Methods**). Notably, this observed isoform diversity can be found in multiple GPCR classes and is not restricted to receptors that couple to particular G protein subtypes or that bind specific types of endogenous ligands (**Extended data Fig. 1**).

Receptor isoforms generate structural diversity

Considering the prevalence of GPCRs with multiple isoforms across receptor classes, we assessed how isoforms contribute to variation in receptor structure. To this end, using our previously-created receptor-specific multiple sequence alignments (see **Methods**), we assessed which receptor structural segments – as annotated in GPCRdb[21] – were more frequently altered in GPCRs with more than one isoform. To better characterise their structural changes, all non-reference isoforms preserving at least one transmembrane segment from the GPCRdb reference isoform (262 non-reference isoforms from 136 receptors) were classified in two groups: topologically preserved (**Fig. 2a**, dark blue) when they preserved all transmembrane (TM) segments found in the reference isoform; and topologically truncated (**Fig. 2a**, light blue), when one or more TM receptor segments were altered (See **Extended data Fig. 2** for more details on this classification).

To understand which structural segments are simultaneously altered in non-reference isoforms, we created structural fingerprints, which are a representation of the structural segments that are preserved or changed with respect to the reference isoform (**Fig. 2a**). Each of the 16 receptor segments forming the fingerprint is shown as a circle; with preserved segments shown in black and varying or missing segments shown in grey (**Fig. 2a;** A typical GPCR has 16 structural segments, which include the N- and C-terminal segments, the 7 transmembrane helices, helix 8 and the 3 intra- and 3 extra-cellular loops). Assessing all the detected structural fingerprints makes it possible to determine the prevalence of different structural changes across the GPCR isoforms, and how commonly they appear in topologically preserved (dark blue) and truncated (light blue) isoforms (**Fig. 2b**). Analysis of the resulting 55 distinct structural fingerprints reveals that after the reference isoform (containing all 16 structural segments), the most frequent structural fingerprint represents changes in the receptor N-terminal segment. This structural fingerprint appears in 65 different receptor isoforms (from 47 receptors), which all preserve the receptor seven transmembrane helix (7TM) topology. The second most prevalent structural fingerprint corresponds to 32 non-reference isoforms (from 23 receptors) with alterations in the C-terminal segment, again preserving the 7TM topology. Following these, we find three structural fingerprints corresponding to truncated receptor isoforms and a structural fingerprint that includes receptors with an altered helix 8 and C-terminus (**Fig. 2b**).

Interestingly, we find an enrichment in class B2 (adhesion) receptors for isoforms with N-terminal variation and in class C receptors for isoforms with C-terminal changes (**Fig. 2b**, inset). This could, perhaps, be related to the different functional relevance of these segments in these classes. For instance, as the large N-terminal region of class B2 adhesion receptors tends to interact with diverse matricellular and membrane-exposed proteins, receptor isoforms with variation in their N-terminus could possibly rewire interactions with extracellular proteins[22]. Similarly, as C-terminal segments can mediate oligomerisation in class C receptors, receptor isoforms with different C-terminal sequences are likely to show different oligomerisation patterns[23]. While these are likely functional effects of changes in the N- and C-terminal segments in receptors from these classes, the exact effect on signalling response will vary depending on the receptor and cellular context and needs to be validated experimentally (see next section). Together, these observations suggest that a large number of non-reference receptor isoforms can generate structural diversity that could affect functional hotspots, while preserving the 7TM architecture.

Structural diversity and functional impact

The structural variability as captured by the structural fingerprints raises the question as to how many of the observed fingerprints correspond to functional isoforms. This is particularly pertinent for the highly truncated receptor isoforms (i.e. those missing many transmembrane segments), which are less likely to productively bind and couple to intracellular transducers. To link structural fingerprints with likely functional impact, we first determined which receptor isoforms have been previously identified and functionally characterised by performing an extensive literature search. We obtained evidence of functional changes in a number of non-reference receptor isoforms belonging to 12 of the structural fingerprints (marked with asterisks in **Fig. 2b**). In this manner, we could link these structural fingerprints to different functional changes reported in the literature (**Fig. 2c**, **Supplementary Table 1**).

This analysis highlighted that isoforms with an alternative N-terminus are the most frequently characterised (**Fig. 2c**) and that variation of this receptor region tends to alter either ligand binding and/or efficacy. For

instance, the cannabinoid receptor (CNR1), has two non-reference isoforms with decreased binding affinity for endocannabinoids, and for which both endogenous and synthetic ligands display altered potency[24]. Similarly, for the chemokine receptor CXCR3, two non-reference isoforms with different N-termini display unique efficacy profiles towards $G\alpha_i$ activation, β-arrestin recruitment, and ERK1/2 phosphorylation for the chemokines CXCL4, CXCL9, CXCL10, and CXCL1[25]. For receptors with an altered C-terminus, the observed functional outcomes are more diverse, but are mostly related to changes in receptor coupling, internalisation and membrane trafficking. For example, in the thromboxane A2 receptor (TA2R), a non-reference C-terminal splice variant retains the capacity of the reference isoform to signal via phospholipase C, but inhibits cAMP production instead of stimulating it[26]. In this manner, TA2R isoforms rewire G protein coupling preferences. Given that post-translational modifications can drive multiple steps of the GPCR life cycle[27], we also analysed whether experimentally validated C-terminal phosphorylation motifs were conserved in non-reference isoforms. We observed that, for a number of receptors, phosphorylation sites that were analysed by mass spectrometry and further characterised in functional studies, are not retained in the non-reference isoform (**Extended data Fig. 3**). Experimental perturbation of such functional sites in C-terminal regions has been shown to result in altered desensitization and recycling of the reference receptor isoforms (**Extended data Fig. 3**) and suggests possible differences in trafficking in the non-reference isoforms lacking these sites.

In highly truncated isoforms, such as those missing all segments from the third or fourth transmembrane helix, the truncated receptor does not always reach the plasma membrane but can still have a dominant negative effect on the full-length reference isoform by sequestering it in the endoplasmic reticulum. This has been seen in isoforms of the serotonin 2C (5HT2C) and the gonadotropin-releasing hormone (GNRH) receptors[28,29]. In other cases, truncated receptors could modulate the function of other GPCRs. This seems to be the case for an isoform of the μ-opioid receptor (OPRM), which lacks both the N-terminus and helix 1. Knocking out this particular isoform in mice can lead to a loss of analgesic action for delta and kappa opioids, as well as those mediated by α2-adrenergic ligands[30].

We observed that many, as yet uncharacterised, receptor isoforms detected in our analysis (shown in grey in **Fig. 2c**), have the same structural fingerprints as receptor isoforms where literature evidence points to specific functional changes (shown in black in **Fig. 2c**). This topological similarity of uncharacterised isoforms to the annotated isoforms (i.e. sharing the same structural fingerprint) can suggest the potential functional impact of structural alterations in these unstudied receptor isoforms. This knowledge could help prioritising experiments to elucidate the functional consequences of structural variation in specific receptor isoforms. However, as we observe in **Fig. 2c**, the same fingerprint can be related to multiple outcomes for different receptor isoforms, stressing that functional inference through structural fingerprint similarity should only be used as a guide. The exact functional effect of a receptor isoform variation will need to be experimentally studied in a relevant cellular or physiological context to understand the real impact of structural changes on receptor signalling (see below).

Combinatorial expression of isoforms can generate physiological signalling bias

All tissues show combinatorial expression of isoforms

Since many of the observed isoforms may impact the way in which extracellular cues are detected and transmitted into intracellular signals, we analysed whether the expression of the 625 previously detected receptor isoforms (reference and non-reference) was ubiquitous or confined to particular tissues (see **Extended data Fig. 4a** and **Supplementary Table 2**). Interestingly, comparing the median expression of all isoforms of a particular receptor in each of the GTEx tissues reveals that the reference isoform is not always the most highly expressed one (**Extended data Fig. 4b**). The number of receptors that are expressed in a given tissue differs widely: with pancreas expressing the lowest number of different receptors (149 GPCRs), and the pituitary expressing the highest number of receptors (236 GPCRs; **Fig. 3a**). However, the mean number of expressed isoforms per receptor in each tissue remains consistently over 1 ranging from 1.42 to 1.72 (conversely, if only one isoform was expressed for each receptor in a tissue, the mean number of expressed isoforms would be 1). Our results mean that, in every tissue, one or more receptors express multiple isoforms simultaneously (i.e. we always observe combinatorial expression). This is also the case if we analyse the data for each donor independently (**Extended data Fig. 4c**), with the mean number of isoforms per receptor in different tissues ranging from 1.05 to 2.04. This points to the existence of GPCRs with multiple isoforms in all of the 30

analysed tissues. In other words, reference and non-reference isoforms are extensively co-expressed in all these tissues.

Receptor isoform combinations can influence signalling response

Taking this into account, we then investigated whether all isoforms of a receptor are present in all tissues where a receptor is expressed, or whether distinct combinations of receptor isoforms can exist in different tissues (see **Extended data Fig. 4d**). To address this, we computed the number of unique combinations of receptor isoforms that are co-expressed in all GTEx tissues (referred to as the number of tissue expression signatures; **Fig. 3b**). If receptor isoforms have different signalling properties, different isoform combinations in human tissues could result in unique signalling outcomes in response to the same input signal. Thus, representing the number of tissue expression signatures of a receptor (**Fig. 3b**, right) can provide insights into how tissue-specific isoform expression can contribute to complex responses upon systemic receptor activation. As an example, both the cannabinoid receptor 1 (CNR1) and the Adhesion G protein-coupled receptor E2 (CD97) have three different isoforms. However, we find four different tissue expression signatures for CNR1 but only one for CD97. In other words, we find that the three CNR1 isoforms can be expressed in four different combinations across tissues, while only one combination exists for CD97, in which all isoforms are consistently expressed in all tissues. This indicates that the CNR1 may present more diversity in signalling responses depending on what combination of isoforms are expressed in the different tissues compared to CD97. These observations collectively suggest that it may not be enough to characterize individual receptor isoforms independently or to analyse how many isoforms a receptor has but highlights the need to identify how many distinct isoform combinations exist, and how they can collectively influence receptor signalling responses in different tissues.

Receptor isoform combinations diversify signalling outcomes and result in signalling bias

In order to explore this further, we focused on CNR1 and the gastric inhibitory polypeptide receptor (GIPR) as two case studies. As mentioned above, the CNR1 receptor has two non-reference isoforms that are topologically preserved. They vary in their N-terminal segments with respect to the reference isoform and this results in altered ligand binding and potency when compared to the reference isoform (see **Extended data Fig. 5a** and **Supplementary Table 1**). If we consider the tissue distribution of these two topologically preserved non-reference isoforms together with that of the reference isoform, we observe that different tissues express different isoform combinations (**Fig. 3c**). Several of these tissues are known to be physiologically relevant for cannabinoid action such as brain, stomach, pituitary or adipose tissue. To assess whether distinct isoform combinations could result in system-specific signalling, we characterised the contribution of reference and non-reference isoform combinations to CNR1 pharmacology in HEK-293 cells.

Because the primary effect of CNR1 is to inhibit adenylate cyclase through $G\alpha_i$ (**Fig. 3d**), we measured cAMP levels using a FRET sensor in live cells[31]. The initial FRET ratios were comparable in cells expressing and not expressing a FLAG-tagged reference CNR1 isoform (**Extended data Fig. 5b, inset**). In cells expressing CNR1, forskolin-stimulated cAMP levels were lower than in cells not expressing the receptor (**Extended data Fig. 5b**). The lower cAMP levels are consistent with a high basal activity of this receptor, which is likely related to the production of endogenous cannabinoids in line with previous studies[32–34]. This trend was reversed by both Rimonabant, a CNR1 inverse agonist that was withdrawn from the clinic by the European Medicines Agency as an anti-obesity drug due to life threatening psychiatric side effects[35] and the neutral antagonist AM4113 (**Extended data Fig. 5c-g**). We then tested whether co-expression of the non-reference isoforms 1 and 2, at equivalent relative expression levels (**Extended data Fig. 5h**), influenced forskolin-stimulated cAMP levels in cells already expressing the reference isoform. Co-expression of a SNAP-tagged isoform 1, but not an equivalent level of a SNAP-tagged reference isoform, together with a FLAG-tagged reference isoform, led to a significant decrease in endpoint cAMP (**Fig. 3d and e**). In contrast, co-expression of isoform 2 caused a faster forskolin response and higher total cAMP levels (**Fig. 3d and f**). These observations suggest that both the extent and rate of CNR1 signalling could be modulated by co-expression of different CNR1 isoforms. Notably, in cells expressing the reference isoform along with non-reference isoform 2, Rimonabant caused higher cAMP levels in a larger fraction of cells, when compared to cells expressing only the reference isoform (**Fig. 3g**).

In the case of GIPR, we detected one non-reference isoform with a different C-terminus (**Extended data Fig. 5i**). An analysis of the tissue distribution of GIPR (**Fig. 3h**) revealed that adipose tissue, adrenal gland, pancreas or brain, where this receptor has regulatory functions, express either the reference isoform alone or both

isoforms in combination. We did not observe any tissue that only expressed the non-reference isoform. Given that the C-terminal region is important for coupling to G proteins and recruitment of β-arrestin, we measured the pharmacological response of the receptor to the gastric inhibitory polypeptide (GIP) by considering 4 different experimental measurements (production of cAMP, mobilisation of intracellular calcium $(Ca^{2+})_i$, and recruitment of β-arrestin 1 and 2) in HEK-293T cells. Interestingly, although receptor signalling is consistently reduced in the presence of the non-reference isoform alone (see **Extended data Fig. 5j-k**), co-expression of both isoforms has a different effect on the engagement of different signalling transducers as inferred through the second messenger measurements (**Fig. 3i**). Specifically, we measured and compared the response to GIP in cells expressing only the reference isoform and cells co-expressing the reference and non-reference GIPR isoforms. We observed that the efficacy (E_{max}) of GIP is most affected while measuring $(Ca^{2+})_i$ release, while its potency (pEC_{50}) is most affected while measuring cAMP signalling. These differences in downstream signalling response, mediated by different G protein subtypes, suggest that the expression of only one or both receptor isoforms can result in physiological signalling bias in different systems, in the presence of the same GIP concentrations.

Taken together, our observations suggest that, for receptors with multiple isoforms, distinct isoform combinations can diversify signalling outcomes in response to the same ligand. In other words, systems expressing different isoform combinations can vary in the overall pharmacological response to the same ligand. These observations also highlight the need to characterise which receptor isoforms are expressed in a system of interest, so that any model used for characterising receptor pharmacology or for assay development (e.g. for drug compound screening) will allow measurements to reflect the combined signalling effects of receptor isoforms. This can be important for two reasons: on the one hand, if the experimental system does not mimic the one we intend to understand (e.g. because we only overexpress a single receptor isoform), it may not provide a translatable readout; on the other hand, ignoring the isoforms already expressed in that experimental system could contribute to biased measurements with respect to the biological system that is being characterised. Collectively, the observations on CNR1 and GIPR show how the expression of distinct combinations of GPCR isoforms in different tissues has the potential to generate system-specific signalling states, which, in turn, can impact ligand and drug response and can contribute to physiological signalling bias.

Receptor isoforms in GPCR drug targets

Tissue expression signatures of GPCR drug targets

Since more than one-third of all approved drugs target a GPCR[14,15], we analysed how many of these drug targets had multiple isoforms and how complex their tissue expression signatures were. To this end, we analysed 111 GPCRs that have been described as targets of 474 FDA approved drugs[36] (**Fig. 4a**). We found that 42% of the known GPCR drug targets have more than one isoform. Interestingly, we found extensive variation in the number of approved drugs for receptors with the same number of tissue expression signatures (**Extended data Fig. 6**). This highlights that the currently approved drugs target GPCRs with multiple isoforms that are expressed in different combinations in different tissues. The drug target with the highest number of tissue expression signatures, the adenosine 2A receptor (AA2AR with 13 unique combinations of isoform expression), exemplifies the possible impact that isoform diversity could have in tissue-specific drug response. These tissue expression signatures include different combinations of a series of highly truncated isoforms, as well as a topologically preserved non-reference isoform. The latter has a shorter C-terminus lacking a phosphorylation site that, in the reference isoform, is crucial for heterodimerisation of AA2AR with the dopaminergic D2 receptor[37] (see **Extended data Fig. 3**). Formation of this dimer, which results in negative allosteric modulation of ligand binding in the partner receptor[38], is believed to be related to the antiparkinsonian effects of AA2AR antagonists[39]. Taken together, these observations suggest that changes in distribution of particular receptor isoforms could not only result in a tissue-specific response to systemically administered drugs, but also may affect the efficacy of drugs targeting other interacting GPCRs.

Non-reference receptor isoforms as potential drug targets

Although the observed structural diversity and tissue-specific distribution of isoforms may confound our understanding of the systemic effects of drugs targeting some GPCRs, it can also present an opportunity to specifically target particular receptor isoforms with desired tissue expression patterns. To assess whether clinical phenotypes can be uniquely associated with the non-reference isoforms, we analysed the recently generated data

from the UK Biobank and available through Gene ATLAS[40]. Gene ATLAS presents associations between polymorphisms (assessed through genotyping specific single nucleotide polymorphisms, SNPs) and phenotypes from ~500,000 individuals in the UK population. Analysis of the statistically significant associations in this dataset revealed the existence of SNPs within the non-reference isoforms that were uniquely associated with disease-related phenotypes (see **Extended data Fig. 7** for examples and **Supplementary Table 3)**. This genetic evidence of isoform-specific phenotypes implies that (a) whole genome sequencing might allow discovery of new associations between non-reference isoform variants and clinical phenotypes and (b) phenome-wide association studies accounting for multiple receptor isoforms may reveal previously underappreciated isoform-specific therapeutic indications.

Achieving isoform-level selectivity for new drug candidates will likely require isoform-specific structural features to be ligand-accessible (**Fig. 4b**). Such unique isoform structural regions should ideally correspond to one or more extracellular receptor segments with the capacity to affect ligand binding affinity or kinetics. By filtering our dataset using this criterion, we obtained 75 topologically preserved non-reference receptor isoforms with potentially different ligand binding interfaces. In terms of their tissue expression, and to allow for isoform selectivity, non-reference isoforms should display a different tissue distribution as compared to the reference isoform or show a more constrained tissue expression with the potential to avoid on-target side effects in tissues unrelated to the drug treatment. By further applying these criteria, we obtained 62 isoforms from 40 different receptors that could be further assessed for pharmacological interest (**Supplementary Table 4**).

To highlight the challenges and opportunities that drugging such isoforms could present, we chose one of the isoforms detected for further characterisation. GPR35, still considered an orphan receptor, is currently the target of the approved mast cell stabilizer lodoxamide, which is administered topically for the treatment of allergic keratoconjunctivitis[41]. It has also been described as the target of pamoic acid, a partial agonist that has been used as an excipient in different drug formulations (from antihelminthic to antipsychotic drugs), and has been found to attenuate visceral pain in mice[42]. Therefore, GPR35 agonists could be of therapeutic interest not only in allergy, but also in neuropathic pain. In addition to the reference isoform, we detected a non-reference isoform that contains an additional 31 amino acids at its N-terminus (**Fig. 4c**; top and **Extended Fig. 6b**). The reference isoform is more ubiquitously expressed, whereas the non-reference isoform is expressed only in some tissues. The two isoforms have different patterns of expression in tissues of potential interest such as in lungs, where only the reference isoform is expressed, or in nerve, where both isoforms are co-expressed. Pharmacological characterisation of the efficacy of different agonists for these isoforms in HEK-293T cells (**Fig. 4c**) reveals that the non-reference isoform does not change $G\alpha_{13}$ or β-arrestin 2 signalling when compared to the reference isoform. However, for each agonist tested, the maximal signal intensity generated was markedly lower (**Fig. 4c**), while potency remained comparable (**Extended Fig. 6c**)

This observation points to the possibilities and challenges that some of the detected N-terminal receptor isoforms may present: when designing receptor agonists, exploiting the structural differences found between isoforms could allow activating the less ubiquitously expressed GPCR isoforms in a tissue of interest, while sparing receptor activation in others where the more ubiquitously expressed isoform of the receptor will be present. This would require drug screening to be conducted considering non-reference isoforms, rather than focussing on reference isoforms only, which is currently largely the case. For GPR35, this could lead to the detection of new drug candidates specifically targeting the extended N-terminus of the non-reference isoform. These compounds could still act as agonists in tissues such as nerve to treat visceral pain, while leaving many other tissues (where only the reference isoform is expressed) unaffected. On the contrary, when trying to block receptor signalling, all receptor isoforms expressed in a tissue of interest should be considered, to avoid the risk of not fully inhibiting GPCR signalling or, at least, of encountering unexpected pharmacodynamics issues. Beyond the receptor isoforms with changes in ligand accessible regions, developments in the design of new allosteric modulators, intra-cellular compounds that influence receptor signalling[43] and RNA based approaches[44] that can selectively silence the expression of specific receptor isoforms might expand the collection of receptor isoforms that can be targeted by drugs to confer improved tissue selectivity.

Further evidence for GPCR isoforms and considerations

While our study reveals extensive structural and combinatorial diversity arising from receptor isoforms, an important next step is a proteome-wide characterisation to gain insights into relevant receptor isoforms in different tissues. Full coverage of membrane proteins such as GPCRs by techniques such as mass spectrometry (MS)-based shot-gun proteomics is still challenging, due to their low abundance and their membrane spanning nature[45]. Furthermore, characterization of isoforms is even more difficult, as protein segments that discriminate isoforms represent a small fraction of detectable peptide fragments of a receptor in MS experiments. Despite these challenges, an analysis of multiple proteomics datasets revealed the existence of peptides that can uniquely discriminate both reference and non-reference isoforms from the GTEx dataset (see all the available isoform-specific MS evidence from ProteomicsDB in https://gpcrdb.org/protein/isoforms (**Extended data Fig. 8**), and further analyses in **Extended data Fig. 9** and **Supplementary Table 5**). As the changes in the relative expression levels of receptor isoforms across tissues could be a source of system bias[46], another direction for further characterisation of receptor isoforms is to accurately quantify their abundance in specific systems. This would not only ensure characterisation of the receptor isoforms at the protein level, but also allow the detection of some lowly expressed but functionally relevant isoforms that may have not been accounted for in our analysis. In this context, further efforts for protein quantification in the GTEx project[47], together with technical developments in focused proteomics approaches[48], should allow for a deeper characterisation and quantification of receptor isoforms in the near future.

The GTEx data analysed here currently provides tissue-level spatial resolution of isoform expression. This implies that some of the isoforms we find in a tissue may be confined to particular cell types. It also means that, whenever possible, the multiple isoforms detected for the different receptors should be further characterised at the cell type level, and ideally in individual cells. To address this, we first investigated whether homogeneous cell populations still express multiple isoforms per receptor by characterising 11 widely-used human cell lines. Consistent with our tissue-level analysis, we find receptors with multiple isoforms in all cell lines (**Extended data Fig. 10a-c**, **Supplementary Table 6**). The observed isoform diversity in cell lines has important implications for the pharmacological characterisation of new ligands in these commonly used systems. Because the collection of isoforms that is expressed in each cell line may vary, this might greatly impact receptor pharmacological readouts. Immortalised cell lines, however, are characterised by genomic and expression heterogeneity, and may not mimic physiological expression patterns of GPCR isoforms. For this reason, we also analysed whether receptor isoforms are co-expressed in primary cells from human tissues. An analysis of GPCR expression in healthy human pancreatic cells, obtained by SmartSeq single cell RNA sequencing[49], revealed the presence of multiple isoforms per receptor in each of the analysed cell types (α, β, δ, acinar, ductal and mesenchymal pancreatic cells) confirming the prevalence of isoform combinatorial expression at the single cell level beyond cancer cell lines (**Extended data Fig. 10a-c**, **Supplementary Table 6**). On-going international projects such as the Human Cell Atlas (https://www.humancellatlas.org/), combined with *ad-hoc* investigations on the detection of protein isoforms described in our study, will allow a more comprehensive investigation in the future into the precise pattern of receptor isoform co-expression in individual cells and cell types from different tissues.

In tissues such as the nervous system, it has been shown that different isoforms could be expressed in different parts of the same cell. For example, the long form of the D2 dopamine receptor is classically post-synaptic, while the short form is pre-synaptic[50]. While current large-scale datasets and high-throughput experimental approaches cannot reveal such patterns of expression, we hope that our analysis, and the resource of human receptor isoforms that we present here, will allow individual labs interested in a specific receptor and its isoforms to delve deeper into a cell type of relevance to investigate such questions.

Discussion

From a canonical to a context-specific view of GPCR signalling

Our integrative analysis (**Fig. 5**) highlights how some GPCR genes can give rise to a collection of receptor isoforms with unique structural and signalling properties, and how the combinatorial expression of these isoforms might diversify receptor signalling response in different human tissues. These findings imply that, for receptors with more than one isoform, the observed signalling readout upon receptor activation by a particular

ligand may be a composite effect of that ligand acting on a collection of receptor isoforms present in the system (**Fig. 5**). Furthermore, given that isoform composition can differ between tissues, activation by the same ligand could result in a variety of system-specific downstream signalling responses (**Fig. 5**). Thus, when interpreting the action of natural or synthetic ligands for receptors with multiple isoforms, we may need to move from the canonical view to a context-specific view of GPCR signalling, where one considers how the structural diversity and combinatorial isoform expression collectively drive system-specific signalling response (where a system can consist of individual cells, tissues or an entire organism, **Fig. 5**).

This means that, to extrapolate ligand effects in particular tissues, the choice of a model system may need to closely mimic the composition of receptor isoforms in the key sites of action. In other words, our observations point to the need of a critical assessment of isoform composition in cells, tissues and in other model organisms, as a key consideration that must be incorporated to successfully translate readouts from cell, tissue and animal models to studies on human physiological responses and clinical trials. The existence of non-reference isoforms and their association with specific clinical phenotypes in the human population points to new opportunities to understand human physiology and for drug development. Specifically, non-reference isoforms with unique tissue expression patterns and variation in specific structural segments might represent new drug targets. Future research accounting for context-specific GPCR signalling and exploiting isoform selectivity might therefore open the door for the development of drugs (e.g. small molecules and RNA based approaches) addressing new and unmet clinical needs, as well as potentially minimizing on-target side effects. We hope that, beyond the proof-of-principle characterisation presented here, our work will inspire more detailed research programmes in academia and industry focussed on particular GPCRs. To facilitate this, we have developed a resource that allows exploring GPCR isoform diversity in humans and assessing the potential for isoform-dependent signalling responses in any given receptor (**Extended data Fig. 8** and https://gpcrdb.org/protein/isoforms).

Conclusion

In biology, it has been generally appreciated that gene duplication and generation of isoforms tend to be inversely correlated mechanisms that provide new material for diversifying biological functions[51]. Our observation about the prevalence of GPCR isoforms in humans suggests that gene duplication and isoform diversity in this family have been simultaneously exploited during evolution. The finding that functionally diverse isoforms can be uniquely co-expressed in different tissues points to a strategy by which organisms can finely calibrate the way in which individual tissues respond to the same extracellular stimulus. Unlike differential transcription of gene duplicates, the fact that isoforms can be post-transcriptionally regulated might represent a rapid mechanism to diversify cell response to the same ligand. In this manner, generation of isoform diversity in different tissues could have been exploited by multi-cellular organisms to rapidly mount different physiological responses to the same ligand at the tissue level, leading to a co-ordinated homeostatic response at an organism-level.

Methods

Transcriptomics, protein sequence and pharmacological data collection. In order to collect information for all human GPCRs, we obtained GPCRdb names, Uniprot identifiers and receptor class annotations of all GPCRs in the GPCRdb[16] database via the REST API and with help of the jsonlite R package. We then used those Uniprot identifiers to obtain associated information from Ensembl[19] (GRCh37 assembly) using the R BioMart interface[52]. Specifically, we collected HUGO Gene Nomenclature Committee (HGNC) symbols, gene and transcript IDs, transcript biotype and signal peptide annotations and protein sequences for all GPCR transcripts in Ensembl. Only transcripts belonging to the 'protein coding' biotype (i.e. excluding short and long non-coding and nonsense-mediated decay transcripts, pseudogenes, etc.) and with a preserved signal peptide (for those receptors with one annotated) were kept for further analysis. Using the Ensembl transcript IDs, we then used transcript-level Transcripts Per Kilobase Million (TPM) data (GTEx_Analysis_2016-01-15_v7_RSEMv1.2.22_transcript_tpm.txt) from the GTEx[18] Portal to filter for GPCR transcripts detected by TruSeq.v1 technology (as annotated in the database in GTEx_v7_Annotations_SampleAttributesDS.txt) and expressed at ≥1 TPM in at least 4 different GTEx donors (using donor annotations found in GTEx_v7_Annotations_SubjectPhenotypesDS.txt). Please refer to the original GTEx publications for further information on RNAseq quality control and analysis[18,53]. The expression cut-off was chosen both to match the cut-off used in another publication from the GTEx consortium that focussed on the analysis of splice variants[54], and to ensure that any non-reference isoform studied would have a high likelihood to be translated into protein and have an effect when combinatorially co-expressed with the reference isoform. In parallel, the donor cut-off was chosen after analysing which was the lowest number of donors that would still capture all known reference receptor isoforms (in our case, 4 donors for the reference ADGRB2 receptor). Furthermore, in order to analyse whether the number of isoforms per receptor was associated with any receptor pharmacological features, we used the jsonlite R package to query the REST API from the Guide to Pharmacology[55] database. This allowed the GPCR HGNC symbols to be linked to annotations on receptor natural/endogenous ligands and transduction mechanisms.

Structural fingerprints and functional annotation. For receptors with more than one isoform, we analysed the structural segments that differed between reference and non-reference isoforms. To do so, we used the protein sequences from Ensembl to obtain a sequence alignment using MUSCLE[56] with default parameters (gap extension set at -1.0). In the case of receptors with transcripts coding for the same amino acid sequence (i.e. with just the 5' or 3' UTRs being different), we retained one representative sequence only. We also used GPCRdb receptor names to download information on receptor segments associated with every receptor residue from GPCRdb as detailed in the previous section. For each multiple sequence alignment, we found the Ensembl transcript corresponding to the structurally annotated sequence in GPCRdb and considered it our reference receptor sequence; receptors for which none of the isoforms could be assigned to a GPCRdb reference (FFAR4, GPR111, GPR107, GPRC5C, LPHN3, LRB4R2 and OR51E1) were excluded from further analysis. For non-reference receptor isoforms, we used the multiple sequence alignment to determine which GPCRdb annotated receptor segments were the same as in the reference and which were different or absent to obtain a "structural fingerprint". Each fingerprint contains information about the 16 structural segments that constitute a GPCR, and whether they are conserved with those in the GPCRdb annotated reference or they differ (i.e. have a different sequence or are missing). In this manner, we generated a 16-dimensional vector with 1s and 0s, denoting conserved or altered segments, for each isoform. In the graphical representation of these fingerprints (see **Fig. 2**), conserved segments are shown as dark circles, and varying or missing segments as grey circles. To ensure that our analysis was centred on isoforms that were anchored to the membrane we excluded non-reference isoforms that did not have at least one transmembrane structural segment in common with their reference isoform. The resulting non-reference isoforms were classified as 'topologically preserved' when all receptor transmembrane segments of the reference isoform were preserved and 'topologically truncated' when at least one annotated transmembrane segment was altered (missing or differing) in comparison to the reference isoform. To assess the functional importance of the non-reference isoforms detected in GTEx, we performed a comprehensive literature search for each receptor with multiple isoforms in the PubMed database and collected all references in which a human non-reference isoform matching one of our analysed transcripts had been functionally characterised.

Characterisation of tissue expression and tissue expression signatures. Using tissue annotations for the transcript-level TPM data from the GTEx Portal (GTEx_v7_Annotations_SampleAttributesDS.txt), we first calculated the number of receptors and the mean number of isoforms per receptor in all 30 tissues analysed in the GTEx project, considering expression data from all available donors. We also calculated the mean number of isoforms per receptor and tissue on a donor-by-donor basis. The tissues include: Brain, Nerve, Pituitary, Adrenal Gland, Thyroid, Blood Vessel, Blood, Spleen, Heart, Muscle, Bladder, Kidney, Salivary Gland, Esophagus, Stomach, Small Intestine, Liver, Pancreas, Colon, Adipose Tissue, Skin, Lung, Breast, Ovary, Fallopian Tube, Uterus, Vagina, Cervix Uteri, Prostate, and Testis. In the case of transcripts that code for the same isoform (i.e. synonymous transcripts: same protein sequence but different 5' or 3' UTRs), we considered the protein isoform to be expressed in all tissues where the synonymous transcripts had been detected. To calculate the most prevalent isoform per receptor in each tissue, we selected the isoform with the highest median expression considering all GTEx donor samples from that tissue. For the 136 receptors with more than 1 isoform, we also obtained receptor-centric Nx30 binary matrices representing the presence or absence of receptor isoforms in different tissues considering all donors, where N is the number of unique isoforms of a particular receptor and 30 is the number of tissues analysed in GTEx. These matrices were then used to calculate the number of unique tissue expression signatures per receptor, that is, the number of unique combinations of receptor isoforms that are found across the different human tissues.

CNR1 cloning and pharmacological characterisation. cDNA sequences for human CNR1 isoforms 1 and 2 were used to generate gene blocks (Integrated DNA Technologies) which were cloned into a pcDNA3.1 backbone with a SNAP tag using BamHI and XbaI sites. The human CNR1 reference isoform was cloned into a pcDNA3.1 backbone with an Influenza hemagglutinin signal sequence and a FLAG tag or SNAP tag using AgeI and XbaI sites. A mixed population of stable HEK-293 cells expressing the FLAG-tagged CNR1 reference isoform was generated using Geneticin selection. Cells were cultured in DMEM/High glucose supplemented with 10% Fetal Bovine Serum (FBS) and maintained at 37°C and 5% CO_2. The FLAG-tagged CNR1 cell line was transfected with 300ng SNAP-tagged CNR1 reference isoform as a control or SNAP-tagged CNR1 non-reference isoform 1 or 2 and 300ng cAMP fluorescence resonance energy transfer (FRET) sensor ICUE3, a gift from Jin Zhang (Addgene plasmid #61622[31]) using Effectene according to manufacturer's standards and incubated for 5hrs at 37°C. Cells were imaged 48-72hrs after transfection. Prior to FRET experiments, cells were pre-treated with 500 µM phosphodiesterase inhibitor 3-isobutyl-1-methylxanthine (IBMX) for 2hrs. Before imaging, cells were labelled with Alexa-568-conjugated M1 antibody (Millipore Sigma F3040) and 500nM SNAP-surface Alexa-Fluor 647 (New England BioLabs S9136S), for 10min at 37°C to label FLAG and SNAP tagged receptors, respectively. Cells were imaged in L-15 media supplemented with 1% FBS and 500 µM IBMX at 37°C in a temperature-controlled imaging chamber (In Vivo Scientific). Cells were imaged on a TiE inverted microscope (Nikon) with a 60x NA 1.49 Apo-TIRF objective (Nikon) with an iXon-888 Life EMCCD camera (Andor). Cells were selected for imaging based on expression of both a FLAG tagged and SNAP tagged isoform as well as the FRET sensor. Initial confocal slice images were acquired to record receptor expression. Relative expression of FLAG and SNAP isoforms was quantified using ImageJ. Briefly, the SNAP channel was used to generate a mask defining the membrane receptor region of the cell. This mask was applied to both the FLAG and SNAP channels and the mean fluorescence intensity of these regions was measured and expressed as a ratio, normalized to the mean ratio of the control ref+ref condition on a given day to account for variability in labelling and dye/antibody aliquots between experiments. Cells were imaged in widefield imaging mode for CFP (405 nm excitation, 400 emission filter), YFP or FRET (405nm excitation, 514 emission filter) and for the SNAP tagged isoform (647 nm excitation, 700 emission filter) every 30 sec with 5 frames of baseline before 5 µM Forskolin (Fsk) addition to stimulate adenylyl cyclase activity. For experiments in which cells were pre-treated with Rimonabant or AM4113, cells were pre-treated with antagonist for 5 minutes, and imaged over this time period, before the addition of Forskolin. In experiments addressing the effect of isoform expression on Rimonabant responses, cells were similarly imaged every 30 sec with 5 frames of baseline before 500nM Fsk addition, followed by 10 µM Rimonabant addition after 5 min. Images were exported in 16-bit tiff files and analysed using ImageJ to calculate the change in CFP/FRET ratio in each cell, as previously described[57]. All data were analysed using GraphPad Prism 7. All data points representing individual cells were normalized to the mean of the control condition for each day, to account for experimental variability between days. All data were analysed by one-way ANOVA (alpha=0.05), with pairwise comparisons made using either Sidak's or Dunnett's multiple comparisons. ANOVA p-values and those of associated post-hoc tests are provided in the figure legends. The asterisks denote p-value, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.

GIPR cloning and pharmacological characterisation. PCR amplicons of GIPR reference or GIPR isoform 1 were ligated into either a pcDNA3.1(+)-sigNluc backbone (signal peptide from the murine 5-HT3R), or a pcDNA3.1(-)-Nluc backbone to generate Nluc-GIPR or GIPR-Nluc constructs, respectively. All cloning results were confirmed by Sanger sequencing (University of Cambridge). FLAG-GIPR reference was supplied by Sulieman Al-Sabah (University of Kuwait). β-arrestin1/2-YFP and GRK5 were supplied by Kathleen Caron (University of North Carolina at Chapel Hill). HEK-293T cells (ATCC[58]) were cultured in DMEM/F12 (Gibco), supplemented with 10% heat-inactivated FBS (Sigma) and 1% antibiotic antimycotic solution (Sigma) and maintained at 37 °C with 5% CO_2 in a humidified environment. Transfections were performed using either Fugene HD (Promega) at a 1:3 w:v ratio of DNA:Fugene HD, in accordance with the manufacturer's instructions, or polyethylenimine (PEI, Polysciences Inc.) and 150 mM NaCl at a 1:6 w:v ratio of DNA:PEI. HEK-293T cells were transfected with pcDNA3.1(-), FLAG-GIPR reference + pcDNA3.1(+)-sigNluc, Nluc-GIPR isoform 1 + pcDNA3.1(-) or FLAG-GIPR reference + Nluc-GIPR isoform 1 at a 1:1 ratio. After 48 hours, 350000 cells were washed three times in fluorescence activated cell sorting (FACS) buffer (PBS supplemented with 1% BSA and 0.03% sodium azide) before and after incubation with the appropriate antibody for 1 hr at room temperature in the dark; rat allophycocyanin (APC)-conjugated anti-FLAG monoclonal antibody, diluted 1:100 (BioLegend); rabbit anti-Nluc polyclonal antibody, diluted 1:100 (Promega); or goat APC-conjugated anti-rabbit IgG polyclonal antibody 31984, diluted 1:150 (Thermofisher). APC intensity was measured using a BD Accuri C6 flow cytometer (BD Biosciences) Ex. λ 633 nm and Em. λ 660 nm. Data were normalized to the mean APC intensity of cells transfected with FLAG-GIPR reference + pcDNA3.1(+)-sigNluc (where expression of FLAG-GIPR reference was assessed) or Nluc-GIPR isoform 1 + pcDNA3.1(-) (where expression of Nluc-GIPR isoform 1 was assessed) as 100% and pcDNA3.1(-) as 0%. For cAMP accumulation and $(Ca^{2+})_i$ mobilisation assays, HEK-293T cells were transiently transfected for 48h with GIPR-Nluc reference + pcDNA3.1(-), GIPR-Nluc isoform 1 + pcDNA3.1(-) or GIPR-Nluc reference + GIPR-Nluc isoform 1, at a 1:1 ratio. GIP (GIP (1-42))-mediated cAMP accumulation was measured after 30 minutes stimulation, in the presence of 500 μM IBMX, using the LANCE® cAMP detection kit (Perkin Elmer) and a Berthold Mithras LB 940 multimode microplate reader, as previously described[59,60]. Forskolin (100 μM) was used as a positive control. $(Ca^{2+})_i$ mobilisation was measured as previously described[59]. GIP (1-42) was robotically added using a BD Pathway 855 high-content bio-imager and images captured every ~0.5 seconds for 80 seconds. The image series were processed using Fiji (Is Just) Image J to correct for background fluorescence of the region of interest and determine the maximum intensity to generate concentration-response curves. Ionomycin (10 mM) was used as a positive control. To measure β-arrestin1/2 recruitment, HEK 293T cells were transiently transfected with β-arrestin1/2-YFP, GRK5 and either GIPR-Nluc reference + pcDNA3.1(+)-sigNluc, GIPR-Nluc isoform 1 + pcDNA3.1(+)-sigNluc, GIPR-Nluc reference + Nluc-GIPR isoform 1 or GIPR-Nluc isoform 1 + Nluc-GIPR reference at a 5:4:1:1 ratio and grown overnight. 50000 cells per well were seeded into poly-L-lysine coated white 96-well plates (Perkin Elmer) in reduced serum media (MEM + 2% FBS + 1% antibiotic antimycotic solution). The following day, cells were incubated with coelenterazine h (5mM, diluted in PBS containing 0.49 mM $MgCl_2.6H_2O$, 0.9 mM $CaCl_2.2H_2O$ and 0.1% BSA) for 10 minutes. The bioluminescence resonance energy transfer (BRET) ratio (530 nm/460 nm) was measured every minute for 60 minutes after GIP (1-42) addition, using a Berthold Mithras LB 940 multimode microplate reader and data were corrected to vehicle treated samples. The peak GIP-induced ΔBRET ratio (530 nm/460 nm) for β-arrestin1 and β-arrestin2 recruitment (9 minutes and 8 minutes, respectively) was measured using a Berthold Mithras LB 940 multimode microplate reader and data were corrected to vehicle treated samples. All data were analysed using GraphPad Prism 8.4. For cAMP accumulation, $(Ca^{2+})_i$ mobilisation and β-arrestin1/2 recruitment assays, data were fitted to obtain concentration-response curves using the three-parameter logistic equation (for pEC_{50} and E_{max} values) and were normalized to the GIPR reference response. To construct the radar plots, E_{max} or pEC_{50} values for each condition were subtracted from the GIPR-Nluc reference to produce ΔE_{max} or ΔpEC_{50} values. For cell surface expression assays, data were normalized to the APC intensity of cells transfected with FLAG-GIPR reference + pcDNA3.1(+)-sigNluc (where expression of FLAG-GIPR reference was assessed) or Nluc-GIPR isoform 1 + pcDNA3.1(-) (where expression of Nluc-GIPR isoform 1 was assessed) as 100% and pcDNA3.1(-) as 0%. Significance was determined using a Mann-Whitney test.

GPR35 cloning and pharmacological characterisation. The GPR35 reference - $G\alpha_{13}$ sensor is described fully in Mackenzie et al[61] and an equivalent GPR35 isoform 1-$G\alpha_{13}$ sensor was generated by replacement of GPR35 reference with GPR35 isoform 1, which contains an additional 31 amino acids at the N-terminus but is

otherwise identical in sequence to the reference. These two constructs were transfected transiently into HEK-293T cells using polyethylenimine linear MW-25000 2 days prior to experiments. Thirty minutes before the assay, cells were washed with Hanks Buffered Saline Solution containing 10 mM HEPES and incubated in the same buffer at 37°C. Because BRET provides a ratiometric signal and each sensor is a single polypeptide, outcomes are anticipated to be independent of expression levels, but this was assessed directly by recording levels of substrate-induced luciferase activity. HEK-293T cells were transfected transiently to co-express either GPR35 reference-eYFP or GPR35 isoform 1-eYFP and β-arrestin 2 tagged with *Nano*luciferase[61–63]. Agonist-induced proximity between the eYFP and *nano*luciferase generated BRET upon addition of the luciferase substrate colenterazine-h. Lodoxamide, pamoic acid and zaprinast were purchased from commercial sources.

Analysis of drug targets and genetic association data. To assess how many receptors with more than one isoform are known GPCR drug targets, we used information on 474 FDA approved drugs matched to their GPCR targets as available in the literature[36]. In addition, we analysed phosphorylation sites annotated in reference isoforms were not retained in non-reference receptor isoforms by analysing data from PhosphoSitePlus[64]. To assess whether non-reference isoforms could be associated with clinical phenotypes of pharmacological interest, we analysed SNPs in the human population that were statistically significantly associated with phenotypes in the Gene ATLAS[40] database. We considered only associations having high statistical significance (P < 1e-10) and involving clinically determined phenotypes (excluding self-reported and generic phenotypes). The resulting SNPs were then mapped to non-reference receptor isoform sequences. For all the analysed GPCRs, we only considered phenotype associations if they uniquely mapped to SNPs in the non-reference isoform. Finally, we assessed which alternative isoforms could represent druggable targets and might offer isoform-level tissue selectivity. To do that, we used the multiple sequence alignments from the previous steps to filter for non-reference isoforms with a different sequence in at least one extracellular segment (i.e. in the N-terminus or ECL1, 2 or 3). We then filtered for non-reference isoforms with a different tissue distribution compared to the reference isoform using the previously computed receptor-centric binary expression matrices. To do so, we used the previously obtained binary tissue expression fingerprints (i.e. the strings of 1 (expressed) or 0 (not expressed) of length 30 corresponding to the 30 GTEx tissues) for each receptor isoform to select those non-reference isoforms with expression fingerprints that differed from their respective reference isoform.

Mass spectrometry data analysis. We analysed the available mass spectrometry (MS) data on GPCR isoforms at the protein level by using a brain-derived MS dataset (PRIDE accession number: PXD0079985). Using these data, we mined for peptides uniquely matching one of the isoforms detected in brain tissue in the GTEx dataset. A total of 105 MS runs from brain tissue were used for identifying potential peptides associated with GPCR proteins (FDR < 0.01 using the target-decoy strategy). Detailed parameters for database search were described in our recent publication[65]. In brief, all 105 MS/MS raw data files were searched against a composite target/decoy database[66] using the JUMP hybrid search engine[67]. The target database was generated by combining the Swiss-Prot human core database (20,368 entries) and manually curated GPCR isoforms (1,033 entries), and the decoy database was produced by reversing the target protein sequences. FDR was estimated by the number of decoy matches (nd) and the number of target matches (nt), according to the equation (FDR = nd/nt). Putative peptide-spectrum matches (PSM)s were filtered by mass accuracy and then grouped by precursor ion charge state and further filtered by JUMP scores (Jscore and ΔJn) to reach the expected FDR. All GPCR transcripts with two or more isoforms as detected in the GTEx brain tissue transcriptome dataset were filtered to obtain isoforms with unique protein sequences. These transcripts were then matched with the brain MS hits to identify receptor isoforms at the protein level. We also analysed data from an independent MS study consisting of 50 MS runs from 29 healthy human tissues[48]. Raw MS data were downloaded from the FTP site ftp://ftp.pride.ebi.ac.uk/pride/data/archive/2019/07/PXD010154. All 1,557 MS data files were searched against the same composite target/decoy database using the comet engine to increase search speed[68]. Major parameters included precursor mass tolerance (±10 ppm) and product ion mass tolerance (0.01 Da), full trypticity, dynamic mass shift for Met oxidation (+15.99491), maximal missed cleavage (n = 2), and maximal modification sites (n = 3). Putative PSMs were filtered by mass accuracy and comet scores (Xcorr and ΔCn) to reduce protein FDR < 0.01. A total of 299,600 peptides were identified with FDR < 1%, including 4,182 peptides mapping to the detected GPCR isoforms. These peptides were then filtered to select only those that would uniquely discriminate a particular GPCR isoform by searching the previously obtained isoform sequences for unique hits. Finally, we used ProteomicsDB[69] (www.proteomicsdb.org) to systematically explore all GPCRs for which more than one isoform had been detected using their REST API. To do that, we used isoform-specific Uniprot

identifiers to identify isoforms that had peptides uniquely matching to them (ISUNIQUE = 1).

Cell line and single cell RNA expression analysis. We used RNA-seq data deposited in BioProject[70] (accession number PRJNA183192) to analyse the receptor transcripts expressed in 11 commonly used human cell lines (HEK293, SHSY5Y, RT-4, PC-3, MCF-7, Hep-G2, HeLa, CACO-2, A-549, U-251MG, and A-431) derived from 10 different human tissues. Following the analysis pipeline used for the GTEx transcript-level expression data, SRA files for all cell line experiments were downloaded and converted to FASTQ files with fastq-dump. Next, transcript level information was obtained with RSEM[71] using the GRCh37 human assembly as a reference. GPCR transcripts expressed over the previously defined expression threshold (≥1 TPM) were considered for comparison. The same analysis was repeated for a single cell RNAseq dataset from healthy human pancreas[49] (BioProject accession number PRJNA322355). SRA files for all single cell experiments were downloaded and converted to FASTQ files with fastq-dump. Again, transcript level information was obtained with RSEM[71] using the GRCh37 human assembly as a reference. GPCR transcripts expressed over the previously defined expression threshold (≥1 TPM) were considered for comparison and grouped according to inferred cell types as annotated by the authors in BioProject.

Data visualisation and analysis. All graphics were created using RStudio version 1.1.419 and the ggplot package as well as the UpSetR, ggalluvial and snakeplotter packages for their respective figures. Unless stated otherwise, all calculations, analyses and data processing were performed using custom written scripts in R. All illustrations were made using Adobe Illustrator CC 22.0.1.

References

1. Urban, J. *et al.* Functional selectivity and classical concepts of quantitative pharmacology. *J. Pharmacol. Exp. Ther.* **320**, 1–13 (2007).
2. Kenakin, T. Biased Receptor Signaling in Drug Discovery. *Pharmacol. Rev.* **71**, 267–315 (2019).
3. Bermudez, M., Nguyen, T. N. & Omieczynski, C. Strategies for the discovery of biased GPCR ligands. *Drug Discov. Today* **24**, 1031–1037 (2019).
4. Hauser, A. S. *et al.* Pharmacogenomics of GPCR Drug Targets. *Cell* **172**, 41-54.e19 (2018).
5. Thompson, M. D. *et al.* Pharmacogenetics of the G protein-coupled receptors. *Methods Mol. Biol.* **1175**, 189–242 (2014).
6. Smith, J. S., Lefkowitz, R. J. & Rajagopal, S. Biased signalling: from simple switches to allosteric microprocessors. *Nat. Rev. Drug Discov.* **17**, 243–260 (2018).
7. Ellis, J. D. *et al.* Tissue-Specific Alternative Splicing Remodels Protein-Protein Interaction Networks. *Mol. Cell* **46**, 884–892 (2012).
8. Buljan, M. *et al.* Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. *Mol. Cell* **46**, 871–883 (2012).
9. Kilpatrick, G. J., Dautzenberg, F. M., Martin, G. R. & Eglen, R. M. 7TM receptors: The splicing on the cake. *Trends Pharmacol. Sci.* **20**, 294–301 (1999).
10. Markovic, D. & Challiss, R. A. J. Alternative splicing of G protein-coupled receptors: physiology and pathophysiology. *Cell Mol. Life Sci.* **66**, 3337–3352 (2009).
11. Wise, H. The roles played by highly truncated splice variants of G protein-coupled receptors. *J. Mol. Signal.* **7**, 13 (2012).
12. Chhibber, A. *et al.* Transcriptomic variation of pharmacogenes in multiple human tissues and lymphoblastoid cell lines. *Pharmacogenomics J.* **17**, 137–145 (2017).
13. Zhou, J., Zhao, S. & Dunker, A. K. Intrinsically Disordered Proteins Link Alternative Splicing and Post-translational Modifications to Complex Cell Signaling and Regulation. *J. Mol. Biol.* **430**, 2342–2359 (2018).
14. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2016).
15. Sriram, K. & Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol. Pharmacol.* **93**, 251–258 (2018).
16. Pándy-Szekeres, G. *et al.* GPCRdb in 2018: Adding GPCR structure models and ligands. *Nucleic Acids Res.* **46**, D440–D446 (2018).
17. GTEx Consortium, *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science.* **348**, 648–660 (2015).
18. Mele, M. *et al.* The human transcriptome across tissues and individuals. *Science.* **348**, 660–665 (2015).
19. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).

20. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
21. Isberg, V. *et al.* Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol. Sci.* **36**, 22–31 (2015).
22. Langenhan, T. Adhesion G protein–coupled receptors—Candidate metabotropic mechanosensors and novel drug targets. *Basic and Clinical Pharmacology and Toxicology* **1**, 1–12 (2019).
23. Milligan, G. A day in the life of a G protein-coupled receptor: The contribution to function of G protein-coupled receptor dimerization. *British Journal of Pharmacology* **153**, S216-29 (2008).
24. Ryberg, E. *et al.* Identification and characterisation of a novel splice variant of the human CB1 receptor. *FEBS Lett.* **579**, 259–264 (2005).
25. Berchiche, Y. A. & Sakmar, T. P. CXC Chemokine Receptor 3 Alternative Splice Variants Selectively Activate Different Signaling Pathways. *Mol. Pharmacol.* **90**, 483–495 (2016).
26. Hirata T, Ushikubi F, Kakizuka A, Okuma M, N. S. Two Thromboxane A2 Receptor Isoforms in Human Platelets. *J. Clin. Invest.* **97**, 949–956 (1996).
27. Venkatakrishnan, A. J. *et al.* Structured and disordered facets of the GPCR fold. *Curr. Opin. Struct. Biol.* **27**, 129–137 (2014).
28. Martin, C. B. P. *et al.* RNA splicing and editing modulation of 5-HT2C receptor function: Relevance to anxiety and aggression in VGV mice. *Mol. Psychiatry* **18**, 656–665 (2013).
29. Grosse, R., Schöneberg, T., Schultz, G. & Gudermann, T. Inhibition of gonadotropin-releasing hormone receptor signaling by expression of a splice variant of the human receptor. *Mol Endocrinol* **11**, 1305–1318 (1997).
30. Marrone, G. F. *et al.* Truncated mu opioid GPCR variant involvement in opioid-dependent and opioid-independent pain modulatory systems within the CNS. *Proc. Natl. Acad. Sci.* **113**, 3663–3668 (2016).
31. DiPilato, L. M. & Zhang, J. The role of membrane microdomains in shaping beta2-adrenergic receptor-mediated cAMP dynamics. *Mol. Biosyst.* **5**, 832–7 (2009).
32. Leterrier, C., Bonnard, D., Carrel, D., Rossier, J. & Lenkei, Z. Constitutive endocytic cycle of the CB1 cannabinoid receptor. *J. Biol. Chem.* **279**, 36013–36021 (2004).
33. Howlett, A. C. *et al.* Endocannabinoid tone versus constitutive activity of cannabinoid receptors. *Br. J. Pharmacol.* **163**, 1329–1343 (2011).
34. Bouaboula, M. *et al.* A selective inverse agonist for central cannabinoid receptor inhibits mitogen-activated protein kinase activation stimulated by insulin or insulin- like growth factor 1: Evidence for a new model of receptor/ligand interactions. *J. Biol. Chem.* **272**, 22330–22339 (1997).
35. Padwal, R. S. & Majumdar, S. R. Drug treatments for obesity: orlistat, sibutramine, and rimonabant. *Lancet (London, England)* **369**, 71–7 (2007).
36. Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: New agents, targets and indications. *Nat. Rev. Drug Discov.* **16**, 829–842 (2017).
37. Borroto-Escuela, D. O. *et al.* A serine point mutation in the adenosine A2AR C-terminal tail reduces receptor heteromerization and allosteric modulation of the dopamine D2R. *Biochem. Biophys. Res. Commun.* **394**, 222–227 (2010).
38. Ferre, S., von Euler, G., Johansson, B., Fredholm, B. B. & Fuxe, K. Stimulation of high-affinity adenosine A2 receptors decreases the affinity of dopamine D2 receptors in rat striatal membranes. *Proc. Natl. Acad. Sci.* **88**, 7238–7241 (2006).
39. Schwarzschild, M. A., Agnati, L., Fuxe, K., Chen, J. F. & Morelli, M. Targeting adenosine A2A receptors in Parkinson's disease. *Trends in Neurosciences* **29**, 647–654 (2006).
40. Canela-xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, (2018).
41. MacKenzie, A. E. *et al.* The antiallergic mast cell stabilizers lodoxamide and bufrolin as the first high and equipotent agonists of human and rat GPR35. *Mol. Pharmacol.* **85**, 91–104 (2014).
42. Zhao, P. *et al.* Targeting of the orphan receptor GPR35 by pamoic acid: A potent activator of extracellular signal-regulated kinase and β-arrestin2 with antinociceptive activity. *Mol. Pharmacol.* **78**, 560–568 (2010).
43. Thal, D. M., Glukhova, A., Sexton, P. M. & Christopoulos, A. Structural insights into G-protein-coupled receptor allostery. *Nature* **559**, 45–53 (2018).
44. Shin, H. *et al.* Recent Advances in RNA Therapeutics and RNA Delivery Systems Based on Nanoparticles. *Adv. Ther.* **1**, 1800065 (2018).
45. Adhikari, S., Sharma, S., Ahn, S. B. & Baker, M. S. In Silico Peptide Repertoire of Human Olfactory Receptor Proteome on High-Stringency Mass Spectrometry. *J. Proteome Res.* acs.jproteome.8b00494 (2019). doi:10.1021/acs.jproteome.8b00494
46. Insel, P. A. *et al.* GPCR Expression in Native Cells: 'Novel' endoGPCRs as Physiologic Regulators and Therapeutic Targets. *Mol. Pharmacol.* 1–29 (2015). doi:10.1124/mol.115.098129
47. Stranger, B. E. *et al.* Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nat. Genet.* **49**, 1664–1670 (2017).
48. Wang, D. *et al.* A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol.*

Syst. Biol. **15**, e8503 (2018).
49. Enge, M. *et al.* Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell* **171**, 321-330.e14 (2017).
50. De Mei, C., Ramos, M., Iitaka, C. & Borrelli, E. Getting specialized: presynaptic and postsynaptic dopamine D2 receptors. *Current Opinion in Pharmacology* **9**, 53–58 (2009).
51. Kopelman, N. M., Lancet, D. & Yanai, I. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. *Nat. Genet.* **37**, 588–589 (2005).
52. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
53. The GTEx Consortium *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (80-.).* **348**, 648–60 (2015).
54. Saha, A. *et al.* Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res.* **27**, 1843–1858 (2017).
55. Harding, S. D. *et al.* The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: Updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. *Nucleic Acids Res.* **46**, D1091–D1106 (2018).
56. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
57. Weinberg, Z. Y., Zajac, A. S., Phan, T., Shiwarski, D. J. & Puthenveedu, M. A. Sequence-specific regulation of endocytic lifetimes modulates arrestin-mediated signaling at the m opioid receptor. *Mol. Pharmacol.* **91**, 416–427 (2017).
58. Bailey, S. *et al.* Interactions between RAMP2 and CRF receptors: The effect of receptor subtypes, splice variants and cell context. *Biochim. Biophys. Acta - Biomembr.* **1861**, 997–1003 (2019).
59. Weston, C. *et al.* Receptor activity-modifying protein-directed G protein signaling specificity for the calcitonin gene-related peptide family of receptors. *J. Biol. Chem.* **291**, 21925–21944 (2016).
60. Knight, A. *et al.* Discovery of Novel Adenosine Receptor Agonists That Exhibit Subtype Selectivity. *J. Med. Chem.* **59**, 947–964 (2016).
61. Mackenzie, A. E. *et al.* Receptor selectivity between the G proteins Ga12 and Ga13 is defined by a single leucine-to-isoleucine variation. *FASEB J.* **33**, 5005–5017 (2019).
62. Milligan, G. Orthologue selectivity and ligand bias: Translating the pharmacology of GPR35. *Trends Pharmacol. Sci.* **32**, 317–325 (2011).
63. Jenkins, L. *et al.* Agonist activation of the G protein-coupled receptor GPR35 involves transmembrane domain III and is transduced via Gα 13 and β-arrestin-2. *Br. J. Pharmacol.* **162**, 733–748 (2011).
64. Hornbeck, P. V *et al.* PhosphoSitePlus, 2014: Mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520 (2015).
65. Bai, B. *et al.* Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer's Disease Progression. *Neuron* **105**, 975-991.e7 (2020).
66. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. & Gygi, S. P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *J. Proteome Res.* **2**, 43–50 (2003).
67. Wang, X. *et al.* JUMP: A tag-based database search tool for peptide identification with high sensitivity and accuracy. *Mol. Cell. Proteomics* **13**, 3663–3673 (2014).
68. Eng, J. K., Jahan, T. A. & Hoopmann, M. R. Comet: An open-source MS/MS sequence database search tool. *Proteomics* **13**, 22–24 (2013).
69. Samaras, P. *et al.* ProteomicsDB: a multi-omics and multi-organism resource for life science research. *Nucleic Acids Res.* **48**, D1153–D1163 (2020).
70. Barrett, T. *et al.* BioProject and BioSample databases at NCBI: Facilitating capture and organization of metadata. *Nucleic Acids Res.* **40**, 57–63 (2012).
71. Li, B. & Dewey, C. N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, (2011).

Acknowledgements

We thank Alissa Hummer, Richard Henderson, Franziska Heidenreich and William R. Orchard for reading the manuscript. M.M-S and M.M.B. acknowledge the MRC (MC_U105185859) for support. M.M-S is a Wolfson College Junior Research Fellow and has been supported by a Federation of European Biochemical Societies Long-Term Fellowship and a Marie Skłodowska-Curie Individual Fellowship from the European Union's Horizon 2020 research and innovation programme under grant agreement No 832620. D.M. acknowledges the support of the Swiss National Science foundation (SNF) under grant P2ELP3_18910. J.P. is partially supported by National Institutes of Health grant R01AG053987. X.W. is partially supported by the pilot grant of the NIH Center of Biomedical Research Excellence (COBRE) for epigenomics of development and disease and pilot grant of the Core NIDA Center of Excellence in Omics, Systems Genetics, and the Addictome. M.A.P. was supported by National Institutes of Health GM117425 and National Science Foundation 1935926. M.H. and G.L. were supported by an MRC confidence in concept award (MC_PC_17156). A.P. was supported by a BBSRC-iCase studentship (BB/JO14540/1) co-funded with AstraZeneca. DEG acknowledges the Lundbeck Foundation (R313-2019-526) and Novo Nordisk Foundation (NNF17OC0031226) for financial support. CM acknowledges the Lundbeck Foundation (R218-2016-1266). S.E.C. was supported by a National Science Foundation Graduate Research Fellowship under Grant DGE 1256260. G.M. and A.B.T. were supported by Biotechnology and Biosciences Research Council Grants BB/P000649/1 and BB/P00069X/1, respectively. M.M.B. is a Lister Institute Fellow and is also supported by the ERC (ERC-COG-2015-682414).

Author Contributions

M.M-S collected data, wrote scripts and performed all the computational analysis. S.E.C. and M.A.P. experimentally characterised CNR1 isoform pharmacology. M.H., A.P. and G.L. experimentally characterised GIPR isoform pharmacology. T.Q., A.E.M., G.M. and A.B.T. experimentally characterised GPR35 isoform pharmacology. D.M. collected data and performed the genomics analysis. C.M. and D.E.G. designed and implemented the GPCRdb receptor isoform browser. X.W. and J.P. collected data and performed the mass spectrometry data analysis. D.E.G helped with aspects of data interpretation. M.M-S and M.M.B. designed the project, analysed and interpreted the results, and wrote the manuscript. All authors read and provided their comments on the draft. M.M-S led the project. M.M.B. initiated, managed, and set the direction of research.

Author Information

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.M-S (mmarti@mrc-lmb.cam.ac.uk) or M.M.B. (madanm@mrc-lmb.cam.ac.uk).

Figures

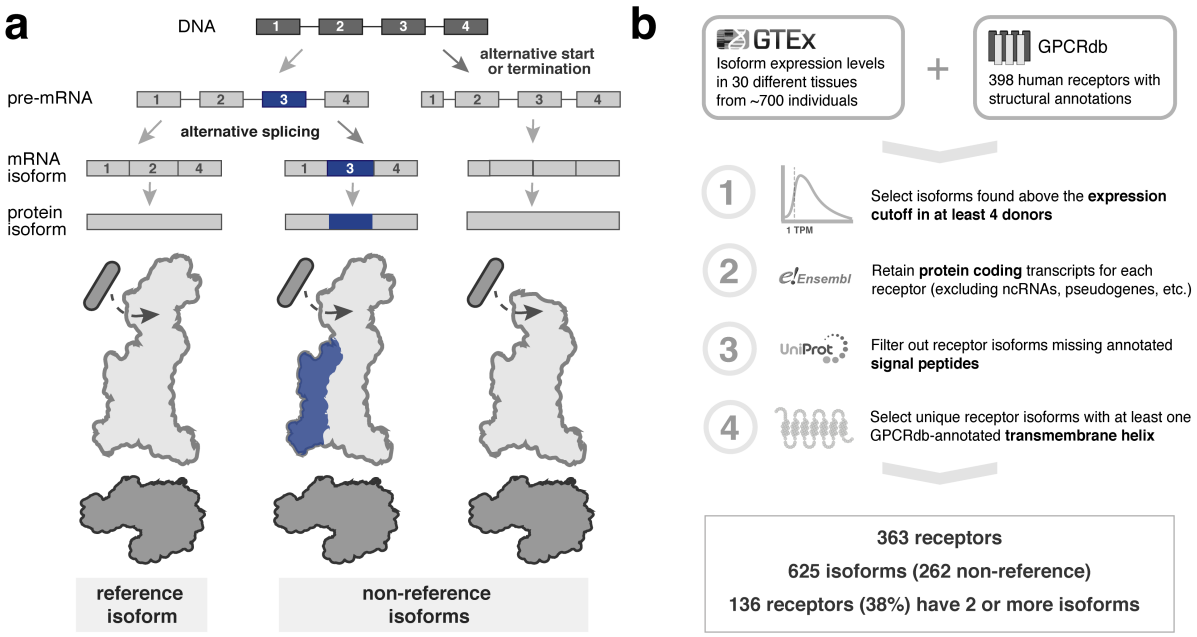

Figure 1. GPCR-wide analysis of isoform expression in humans. a, Alternative transcription start or termination, together with alternative splicing, can generate non-reference receptor isoforms (i.e. isoforms with different sequences than the GPCRdb annotated receptor), with altered structural and signalling properties. **b**, Analysis pipeline combining GTEx isoform-level data with GPCRdb, Ensembl, and UniProt annotations to filter for highly-expressed, protein-coding isoforms; isoforms with truncated signal peptides or those without at least one conserved transmembrane helix as compared to their GPCRdb reference are not considered (see **Methods**).

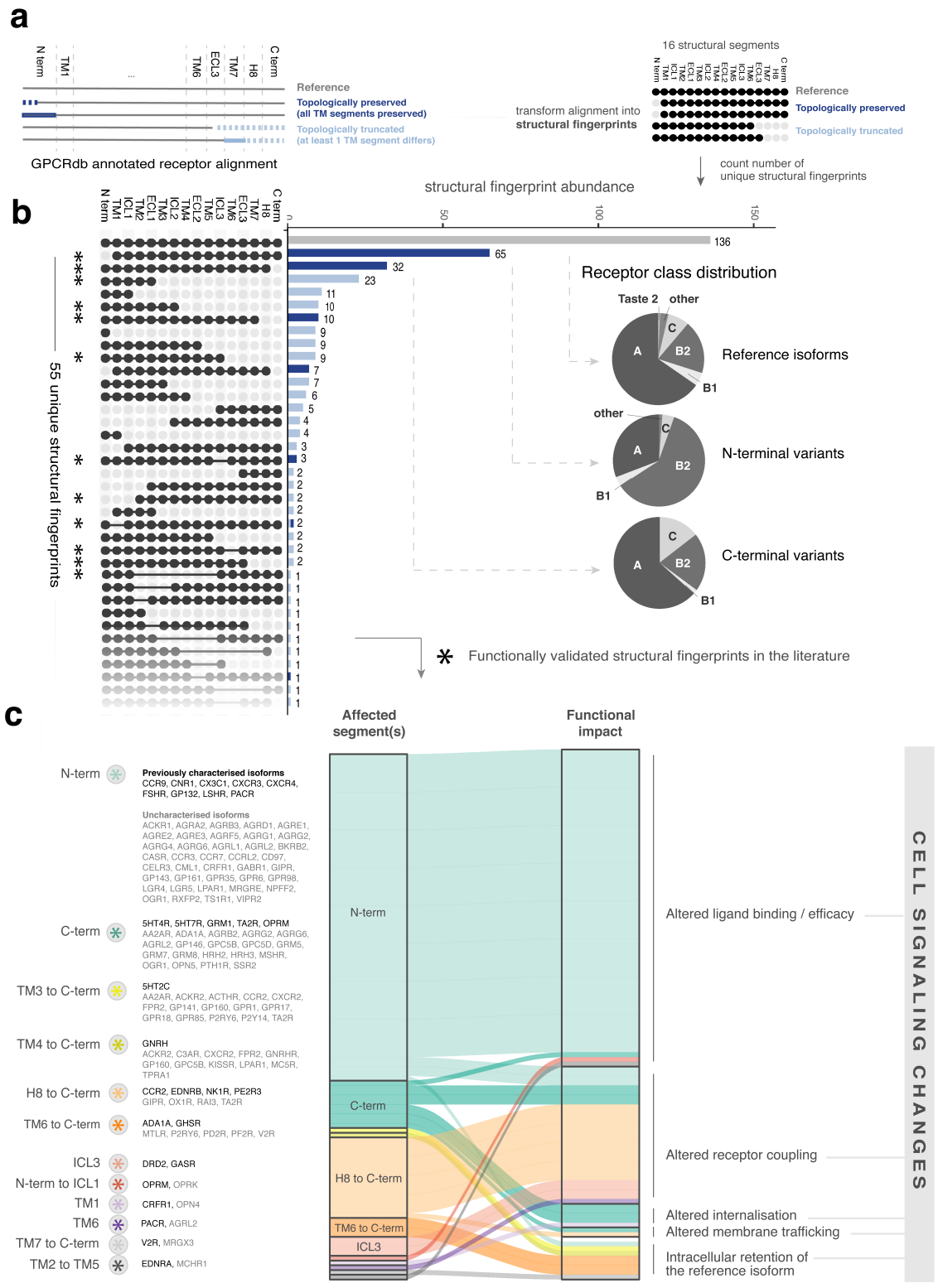

Figure 2. Structural diversity and functional impact of receptor isoforms. a, Annotation of multiple sequence alignments for all GPCRs with more than one isoform according to structural segment definitions in GPCRdb allows discriminating topologically preserved isoforms containing all the transmembrane (TM) segments present in the reference isoform (dark blue), and topologically truncated isoforms with at least one missing or altered TM segment (light blue). This information is used to generate a structural fingerprint for every receptor isoform. **b**, Distribution of abundance of unique structural fingerprints for receptors with two or more isoforms. Dark blue and light blue bars represent the abundance of topologically conserved and truncated versions of a structural fingerprint, respectively. Pie charts show the GPCR class distribution, as annotated in the IUPHAR/BPS the Guide To Pharmacology database (www.guidetopharmacology.org), in

the three most common structural fingerprints. Asterisks indicate fingerprints with at least one functionally characterised receptor in the literature as shown in **c**, Association between structural fingerprints and functional impact as available in the literature. Receptors with functionally characterised isoforms are listed in black. Height of the individual stacked bars represents the number of characterised isoforms in the literature. Receptors listed in light grey have the same structural fingerprint as those shown in black with literature evidence.

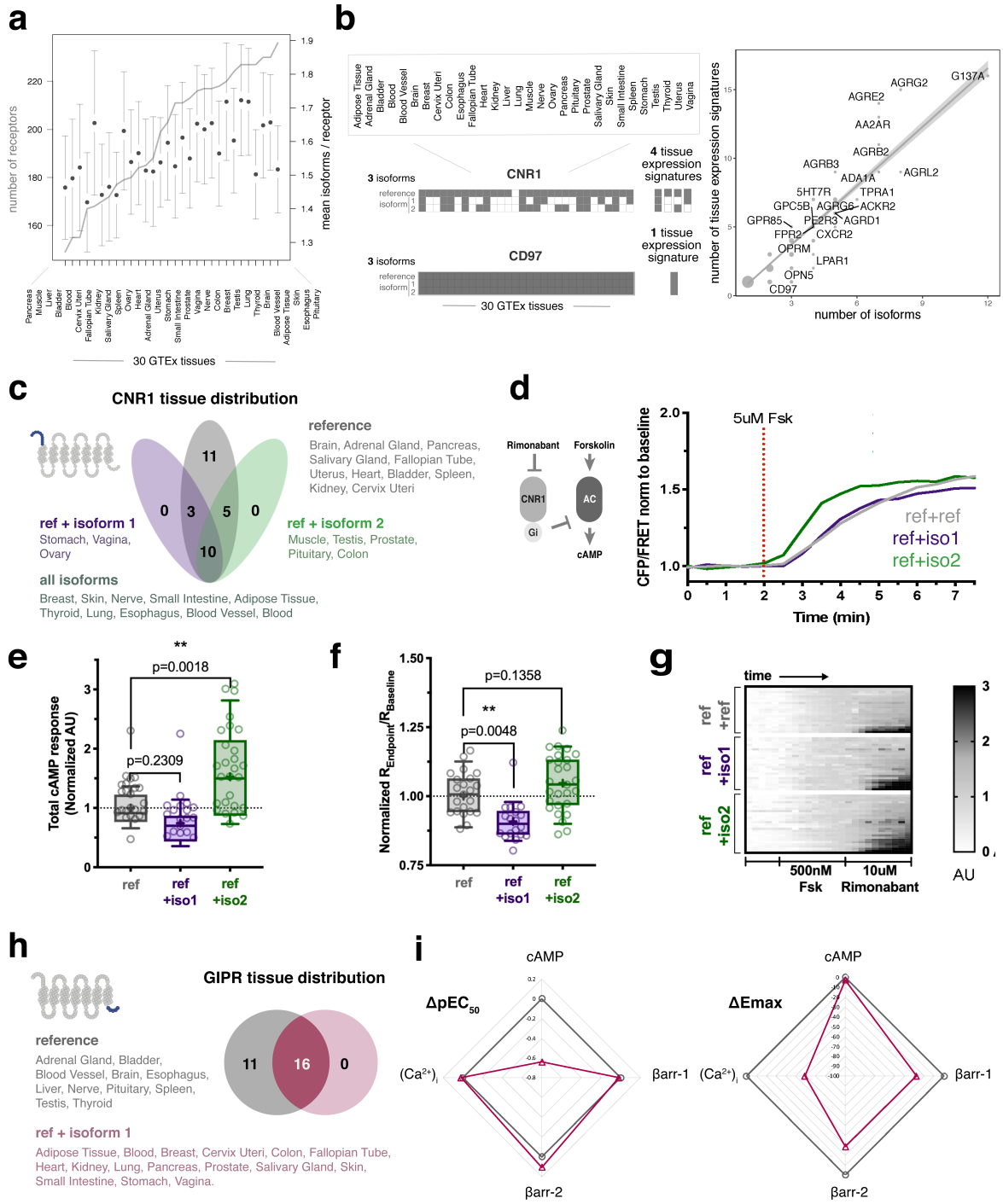

Figure 3. Tissue distribution of receptor isoforms and the impact of combinatorial isoform expression on pharmacological response. a, Relationship between the number of GPCRs expressed in the different GTEx tissues (left axis, grey line) and the mean number of isoforms per receptor expressed in each tissue (right axis, mean and standard error bars shown in black). **b**, Isoform expression matrix for the cannabinoid receptor 1 and Adhesion G protein-coupled receptor E2 (CNR1, CD97). Considering how many unique isoform combinations exist for a particular receptor across all GTEx tissues allows us to determine the number of receptor tissue expression signatures. The relationship between number of tissue expression signatures and the total number of isoforms per receptor is shown on the right. The dark grey regression line was obtained using a linear model and the light grey confidence interval represents the standard error. **c,** Tissue distribution of the reference (grey) and the two topologically conserved CNR1 non-reference isoforms (purple and green) with changes in the N-terminus (see alignment in **Extended Fig. 5a**). **d,** Top right inset: Schematic representation of CNR1-mediated inhibition of cAMP production by adenylate cyclase (AC). Pharmacological modulation can be attained by AC activation by Forskolin (Fsk) or CNR1 inverse agonism by Rimonabant. Representative trace of mean cAMP levels in Fsk

(5μM)-stimulated HEK 293 cells expressing FLAG-tagged CNR1 reference isoform along with SNAP-tagged reference isoform as a control or SNAP-tagged non-reference isoform1 or isoform 2. **e,** Endpoint cAMP levels are significantly decreased in cells co-expressing the reference isoform and non-reference isoform1 compared to cells expressing only the reference isoform. **f,** Total cAMP response, normalized to cells expressing the reference isoform, is increased in cells co-expressing the reference isoform and non-reference isoform2 compared to the control cells (ref+ref, n=23 cells; ref+iso1, n=18 cells; ref+iso2, n=27 cells; by one-way ANOVA with Dunnett's multiple comparisons test). Box plots show 10-90th percentile with the median, 25th, and 75th percentile as well as the mean (+). Individual data points are overlaid on the box plot. **g,** Heatmap representation of cAMP levels after Fsk and Rimonabant treatment over time in individual HEK 293 cells expressing a Flag-tagged reference isoform and the different SNAP-tagged isoforms. CFP/FRET ratios were normalised using the baseline ratio as 0 and the maximum ratio after Fsk, but before Rimonabant, treatment as 1. **h,** Tissue distribution of the reference (grey) and the GIPR non-reference isoform 1 with changes in the C-terminal segment (pink) (see alignment in **Extended Fig. 5i**). **i,** Radar plots summarising differences in potency (ΔpEC_{50}) and maximum response ($\Delta Emax$) for 4 different measurements (cAMP and intracellular calcium $(Ca^{2+})_i$ release, and β-arrestin 1 and 2 recruitment) between cells expressing the GIPR reference isoform (grey line) or a combination of the reference isoform and isoform 1 (magenta).

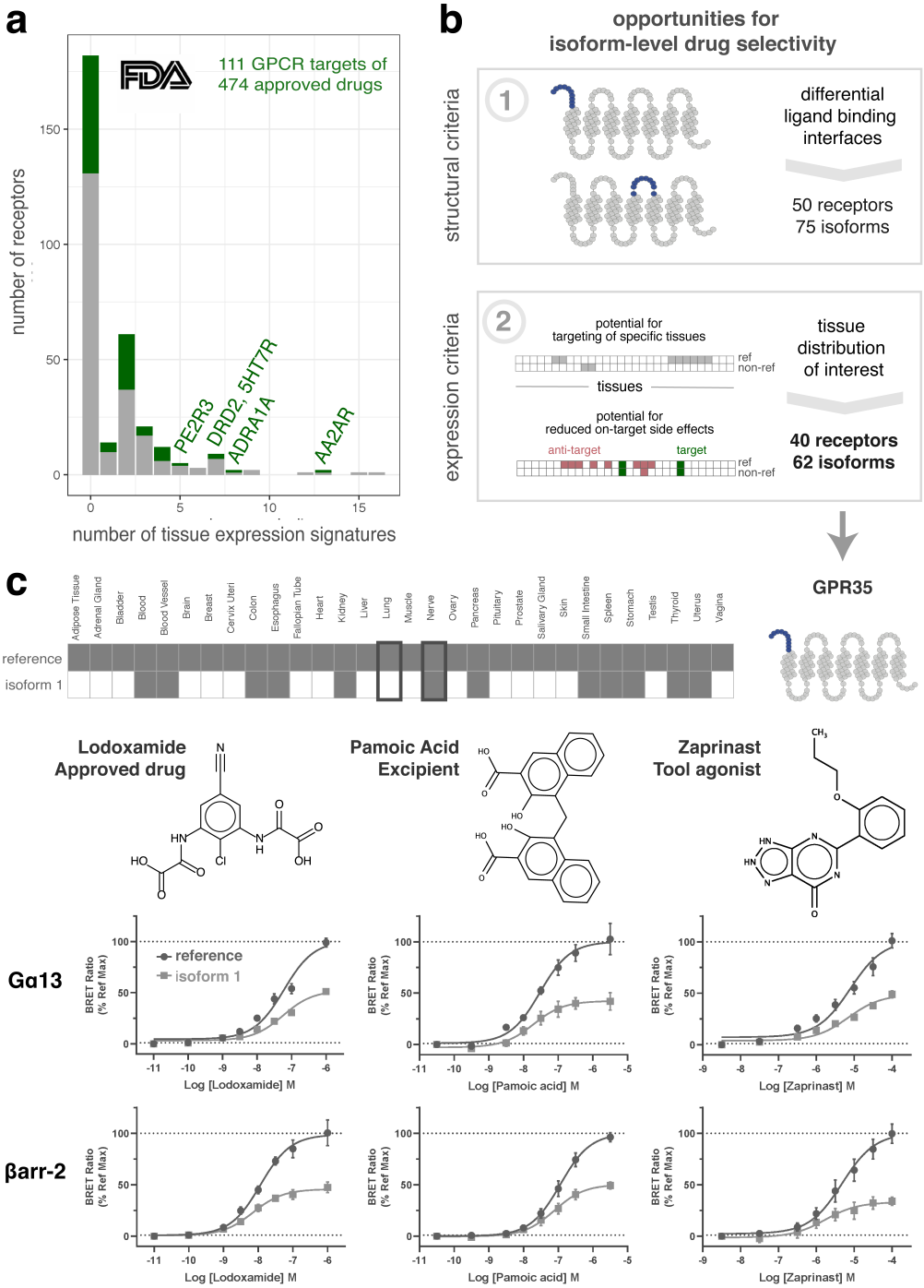

Figure 4. GPCR isoforms of drug targets. a, Number of receptors categorized by the number of tissue expression signatures for all GPCRs (grey) and for those which belong to the 111 targets of 474 FDA-approved drugs (in green). **b**, Filtering based on structural and expression considerations identifies non-reference isoforms with extracellular structural changes and tissue distribution that is different from the reference isoform, potentially allowing the development of ligands that specifically target them. **c**, Tissue distribution of the reference and isoform 1 of GPR35. BRET signals were monitored after treatment of HEK-293T cells with varying concentrations of lodoxamide, pamoic acid, and zaprinast using a GPR35-Gα_{13} SPASM sensor (upper panel; n=3) or eYFP-tagged GPR35 isoforms and β-arrestin 2 tagged with *Nano*luciferase (lower panel; n=3, see **Methods** for further details).

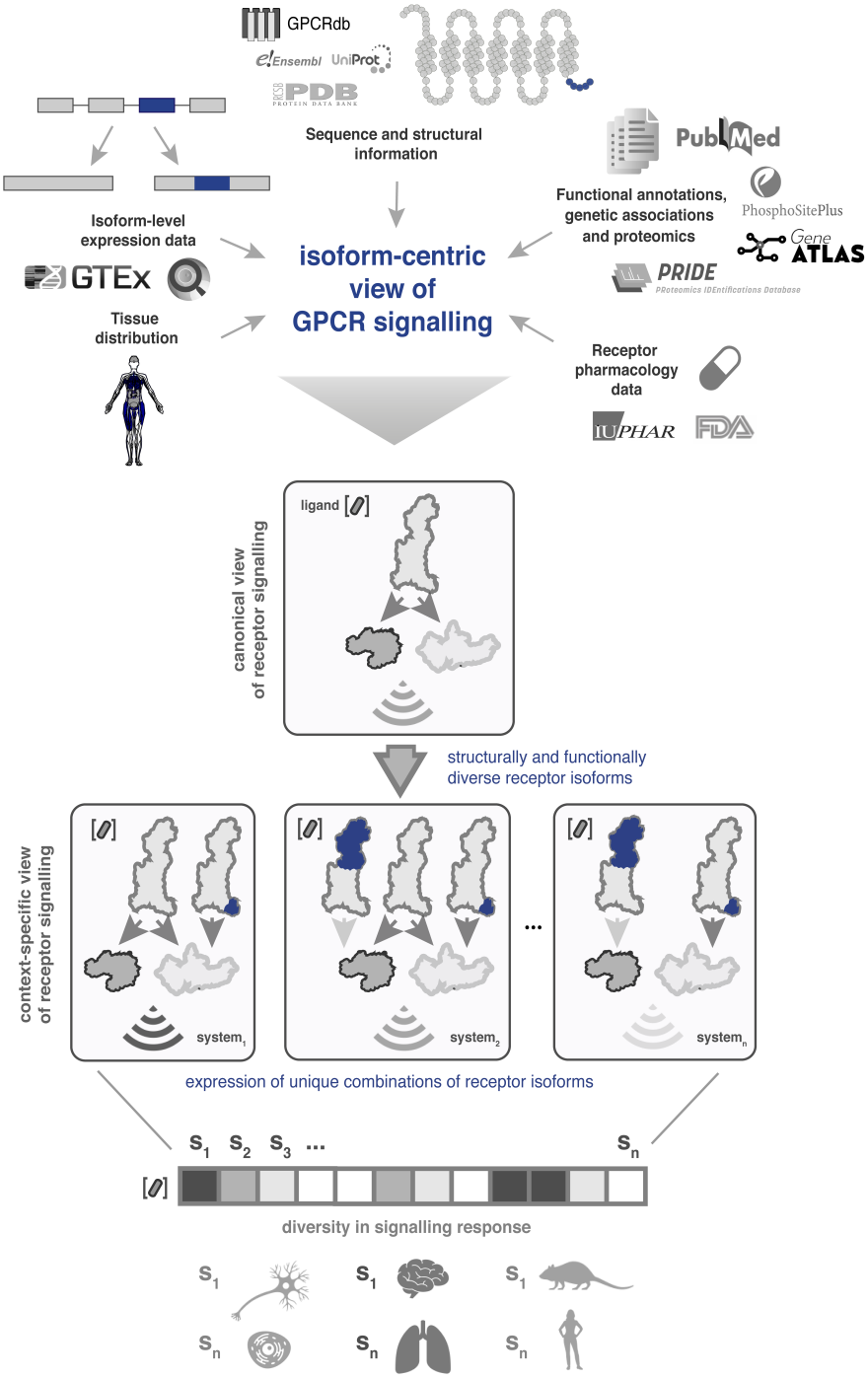

Figure 5. From a canonical to a context-specific view of GPCR signalling. Integration of multiple datasets provides an isoform-centric view of GPCR signalling. In the case of receptors with a single isoform, activation by a particular ligand concentration will result in an observed signalling readout. However, when receptors have multiple isoforms expressed in the same system, a signalling readout may be the collective result of activating structurally and pharmacologically distinct receptor isoforms. If the isoform composition changes in different systems (s_1 to s_n), this can result in differences in signalling in response to the same ligand (left, right; intensity of the box denotes extent of signalling response). The different systems can correspond to different levels of complexity from that of a single cell, to tissues, and entire organisms. In this manner, combinatorial expression of functionally distinct GPCR isoforms can diversify receptor signalling response.