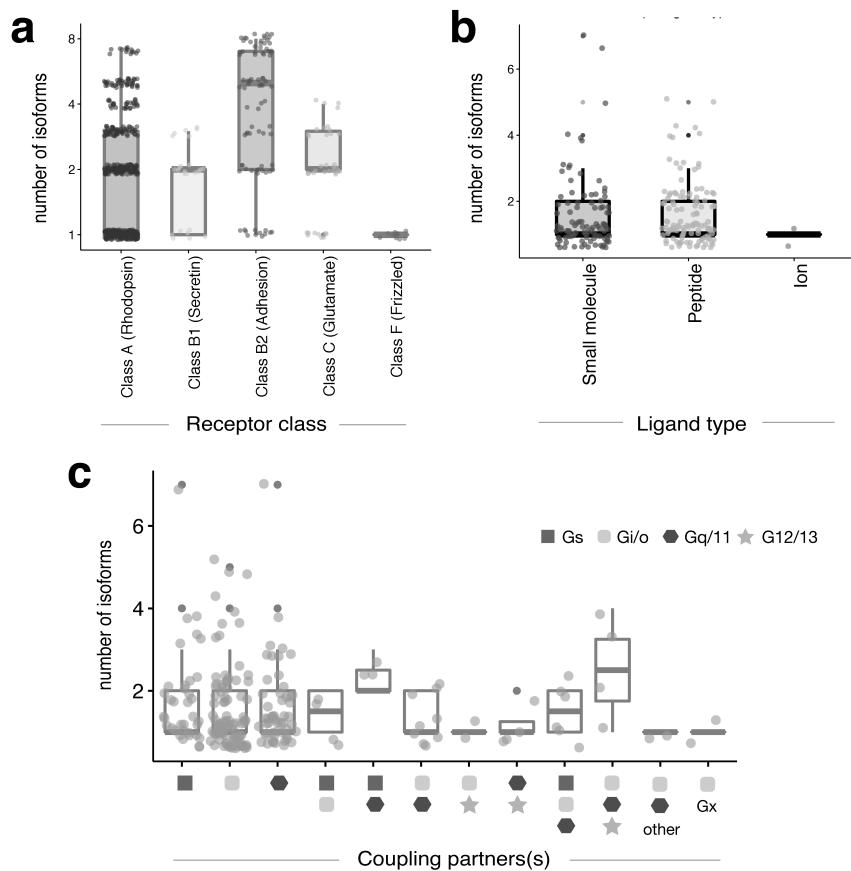
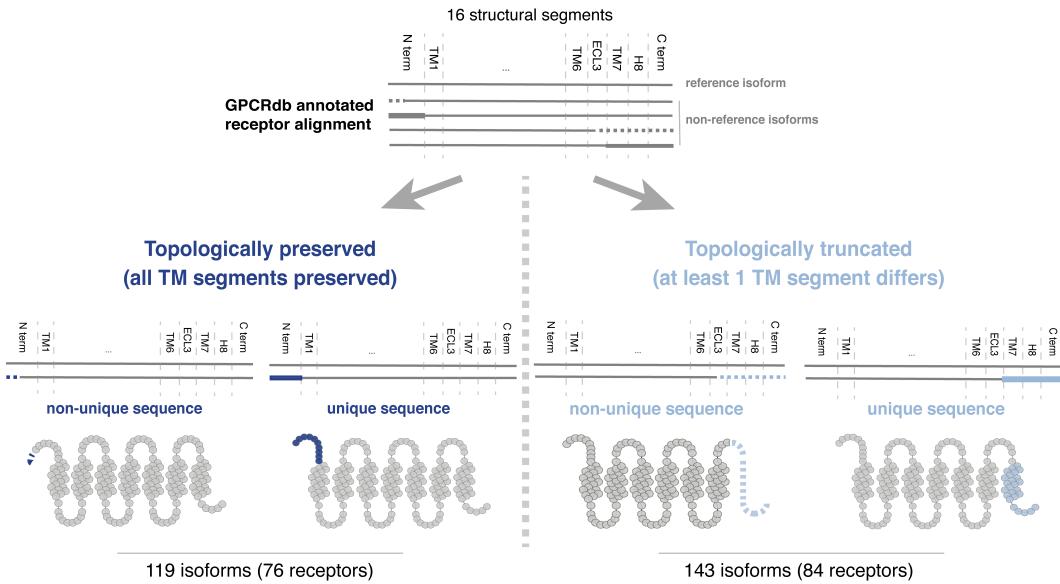


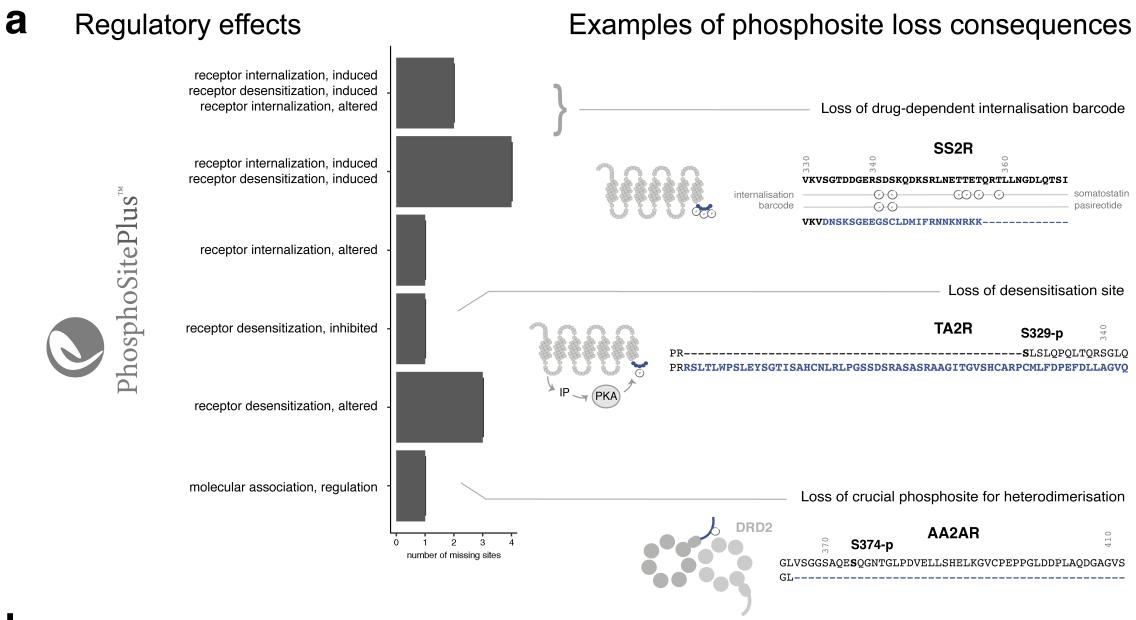
Extended data: Figures



Extended Figure 1. Isoform diversity in GPCRs. Number of isoforms **a**, for different receptor classes **b**, grouped by the type of natural ligand of each receptor; and **c**, grouped by each receptors' coupling partners as annotated in IUPHAR/BPS the Guide To Pharmacology database (www.guidetopharmacology.org).



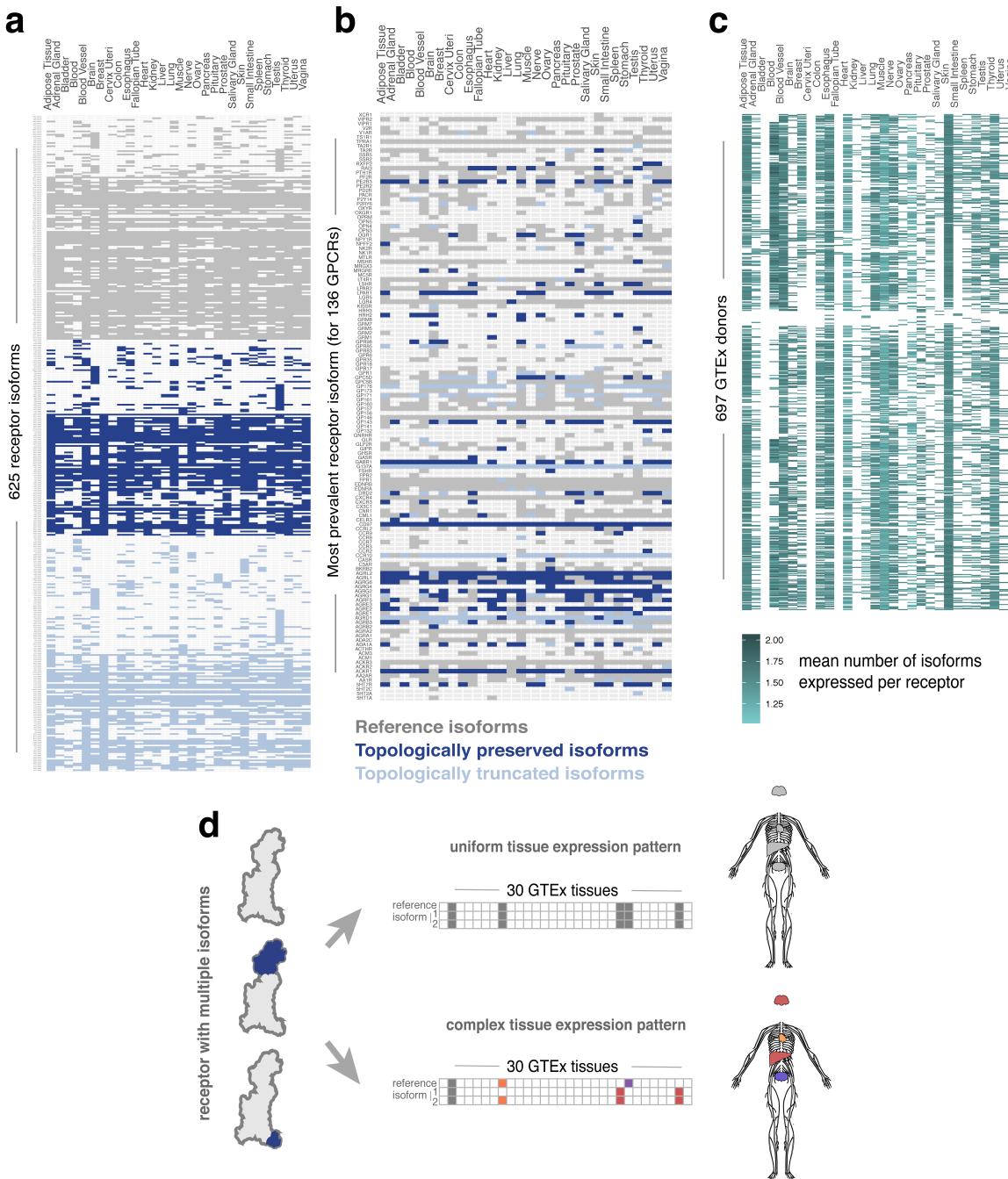
Extended Figure 2. Classification of receptor isoforms according to topology. Structural classification of non-reference isoforms: isoforms are considered topologically preserved if all transmembrane (TM) segments are conserved in sequence with respect to the reference isoform; this includes receptor isoforms with changes (including deletions and unique isoform sequences) in the receptor N- and C-terminus, as well as in intracellular and extracellular loops and the helix 8 (H8) segment. If at least one TM receptor segment changes (i.e. is partially or totally missing or has a different sequence than the reference isoform), isoforms are classified as topologically truncated. Both topologically preserved and truncated isoforms can either have an alternative sequence in the segments that differ from the reference (unique sequence, represented as solid coloured lines) or lack parts of those segments altogether, thus representing a shorter version of the reference sequence (non-unique, represented as dotted coloured lines).



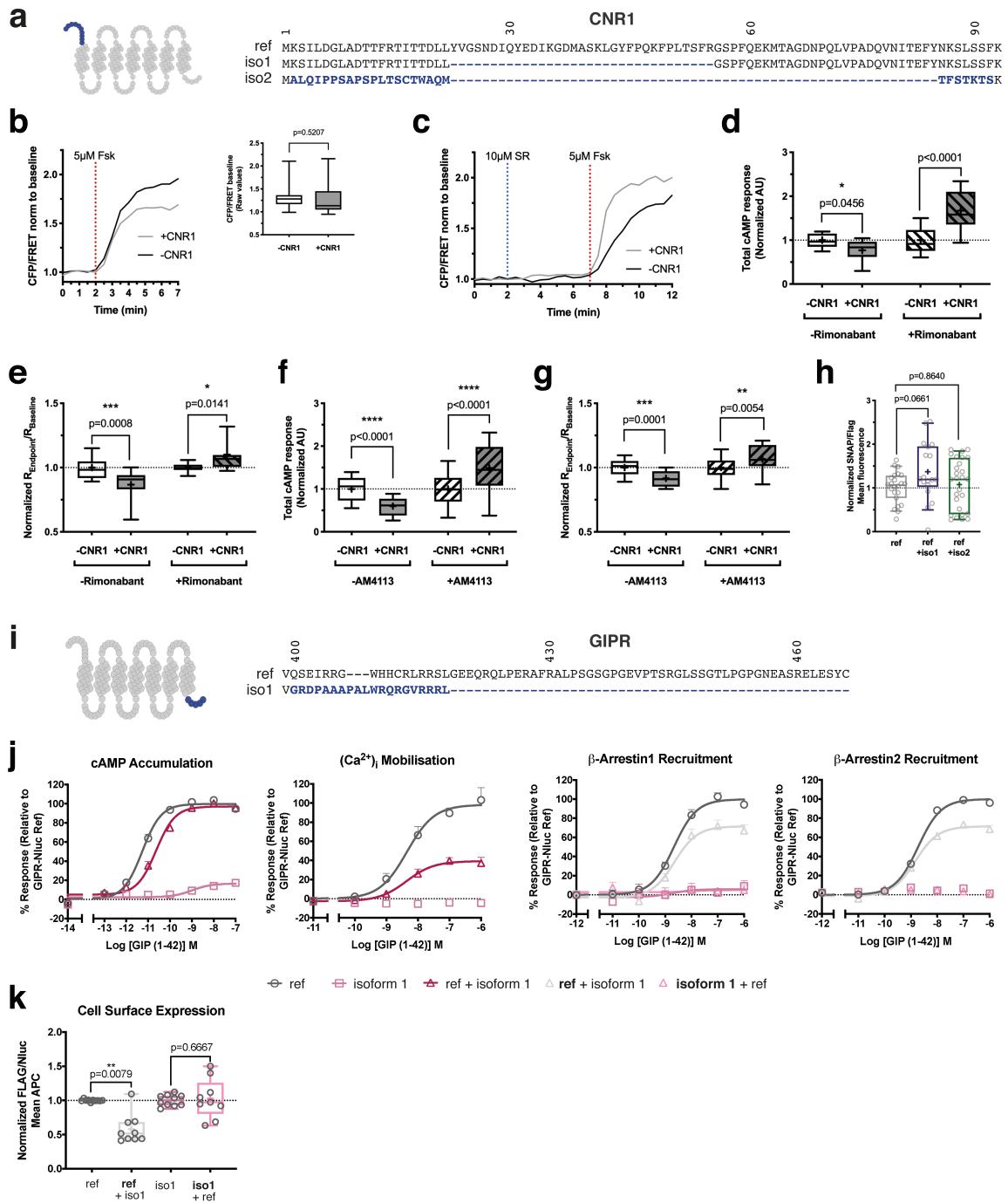
b

Receptor	Affected Isoform	Modification	Motif	Effect	PubMed ID
SSR2	ENST00000315332	S341-p	GTDDGERSDSKQDKS	receptor internalization, induced;	21330405; 21343287; 24565897; 25012983
SSR2	ENST00000315332	S343-p	DDGERSDSKQDKKSRL	receptor desensitization, induced;	21330405; 21343287; 24565897; 25012983
SSR2	ENST00000315332	T359-p	ETTETQRLLNGLQ	receptor internalization, altered	24565897
SSR2	ENST00000315332	T356-p	RLNETTETQRLLNG	receptor internalization, induced;	24565897
SSR2	ENST00000315332	T353-p	DKSRLNETTETQRQL	receptor desensitization, induced	24565897
SSR2	ENST00000315332	T354-p	KSRLNETTETQRLL		24565897
MSHR	ENST00000555427	S316-p	LKEVLTCWS	receptor internalization, altered	17130136
OPRM	ENST00000229768	T396-p	LENLEAETAPLP	receptor desensitization, inhibited	24304623
TA2R	ENST00000411851	T337-p	LSLQPQLTQRSGLQ		15100160
TA2R	ENST00000411851	S331-p	STRPRSLSLQPQLQ		12370102
TA2R	ENST00000411851	S329-p	RLSTRPRSLSLQPQL	receptor desensitization, altered	10827090
AA2AR	ENST00000444262	S374-p	SGGSAQESQGNTGLP	molecular association, regulation	20197060

Extended Figure 3. Non-reference isoforms with missing phosphorylation motifs. **a**, Analysis of PhosphoSitePlus phosphorylation motifs with annotated regulatory function in non-reference receptor isoforms. Frequency of phosphosites related to receptor trafficking and interaction that are found to be missing in non-reference isoforms (left) and examples of potential consequences of this lack of regulation on receptor function for somatostatin receptor type 2 (SSR2), thromboxane A2 receptor (TA2R) and adenosine receptor A2a (AA2AR) (right). **b**, Table of all detected non-reference receptor isoforms where a phosphorylation motif can be lost, together with details about the function of that particular phosphosite in the reference isoform and its associated literature evidence (examples in **a** are highlighted in grey).

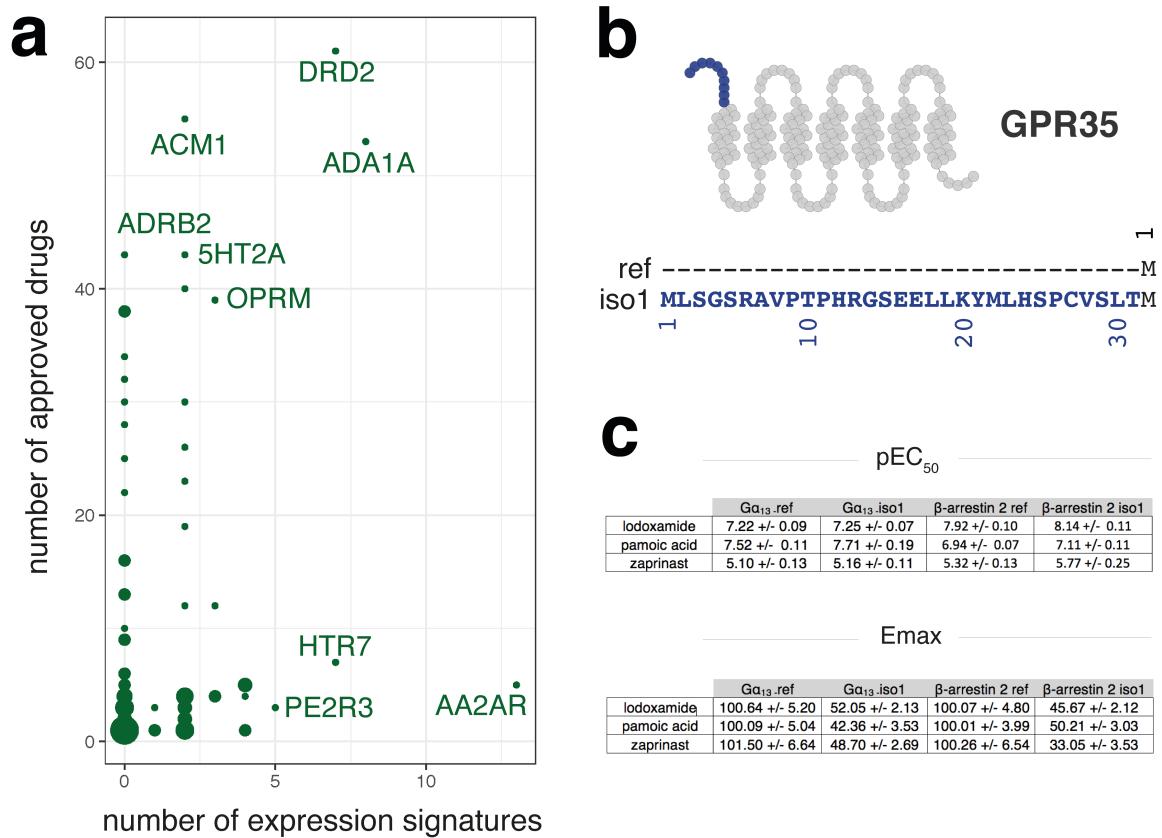


Extended Figure 4. Isoform tissue distribution. **a**, Binary representation of presence / absence of particular receptor isoforms in the 30 different GTEx tissues according to their classification as reference isoform (grey), topologically preserved isoform (dark blue) or topologically truncated isoform (light blue). This figure shows that every tissue expresses all these types of isoforms and that each non-reference isoforms can be consistently found in multiple tissues. **b**, Heatmap showing whether the most prevalent isoform for a receptor in a particular tissue is a reference (grey), topologically preserved (dark blue) or topologically truncated (light blue) receptor isoform. The most prevalent isoform in a particular tissue was considered to be the one with the highest median expression in TPM. **c**, Mean number of isoforms expressed per receptor for every GTEx donor-tissue combination. All available combinations of donor-tissue expression had a mean number of isoforms per receptor over 1 (i.e. reference and non-reference isoforms are co-expressed in each GTEx donor-tissue combination). Empty cells (white) represent donor-tissue combinations with no expression data. See underlying data for **a**, **b** and **c** in **Supplementary Table 2**. **d**, For receptors with multiple isoforms two scenarios are possible. In one scenario, all receptor isoforms could all be consistently expressed in human tissues resulting in a uniform expression pattern. In a second scenario, different isoform combinations may be expressed in different tissues resulting in a complex expression pattern. In order to assess this, we calculated the number of tissue expression signatures per receptor as presented in **Figure 4**.

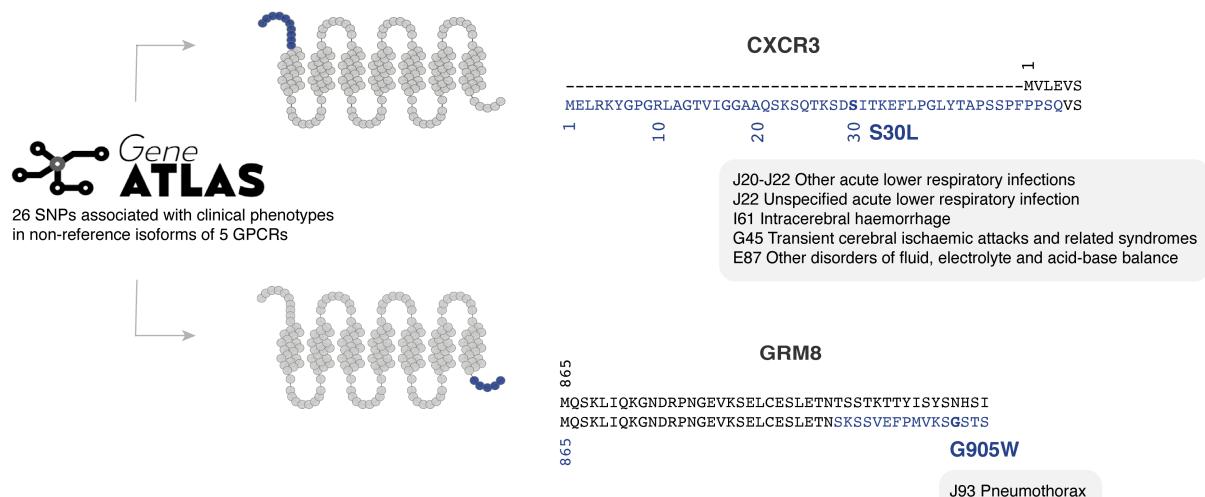


Extended Figure 5. Characterization of the pharmacological and signalling effects of isoform co-expression for CNR1 and GIPR. **a**, Alignment of the N-terminal segment of the reference and non-reference CNR1 isoforms. Varying regions in isoforms 1 and 2 are highlighted in blue. **b**, Change in mean CFP/FRET ratios induced by 5 μM Forskolin (Fsk) in HEK293 cells. Cells from the same coverslip expressing (grey line) and not expressing (black line) the CNR1 reference isoform are shown. Cells expressing CNR1 show a lower Fsk response, as expected with Gi signalling (inset). CFP/FRET ratios in HEK293 cells expressing the CNR1 reference isoform (+CNR1 n=30 cells) or not (-CNR1 n=40 cells). The ratios are not significantly different ($p=0.527$ by unpaired t-test). **c**, The increased Fsk response in CNR1-expressing cells treated with Rimonabant (SR) is consistent with Adenylyl cyclase super-activation as reported in cell lines stably expressing Gi-coupled receptors like CNR1. **d**, Total cAMP response induced by Fsk in CNR1-expressing cells. When normalised to non-expressing cells, the response is decreased in CNR1-expressing cells compared to non-expressing cells. Pre-treatment with 10 μM Rimonabant reversed the CNR1-dependent decrease. **e**, Endpoint cAMP response induced by Fsk in CNR1-expressing cells, normalised to non-expressing cells, is significantly decreased in CNR1-expressing cells compared to non-expressing cells. Pre-treatment with 10 μM Rimonabant reverses the CNR1-dependent decrease (-CNR1 -Rimonabant n=24

cells; +CNR1 -Rimonabant n=19 cells; -CNR1 +Rimonabant n=27 cells; +CNR1 +Rimonabant n=17 cells). **f**, Pre-treatment with 5 μ M neutral antagonist AM4113 significantly increased Fsk-stimulated total cAMP levels in CNR1-expressing relative to non-expressing cells. **g**, AM4113 pre-treatment significantly increased Fsk-stimulated endpoint cAMP levels in CNR1-expressing cells relative to non-expressing cells (-CNR1 -AM4113 n=55 cells; +CNR1 -AM4113 n=50 cells; -CNR1 +AM4113 n=44 cells; +CNR1 +AM4113 n=48 cells). In e-g p-values were obtained by one-way ANOVA with Sidak's multiple comparisons tests. **h**, Expression ratios of SNAP-tagged CNR1 isoforms to Flag-tagged reference isoform are not significantly different across the combinatorial expression conditions (one-way ANOVA with Dunnett's multiple comparisons test). In all cases, box plots show 10-90th percentile with the median, 25th, and 75th percentile as well as the mean (+). **i**, Alignment of the C-terminal segment of the reference and non-reference GIPR isoforms. Varying regions in isoform 1 are highlighted in blue. **j**, Concentration-response curves for cAMP accumulation and intracellular calcium (Ca^{2+}), mobilisation, in response to GIP (1-42) stimulation, in HEK 293T cells expressing GIPR-Nluc reference (dark grey, n=4), GIPR-Nluc non-reference isoform 1 (pink n=4 and n=5, respectively) or GIPR-Nluc reference and GIPR-Nluc non-reference isoform 1 (magenta, n=4). Experiments were performed in duplicate, data normalized to GIPR-Nluc reference and expressed as mean \pm SEM. GIP (1-42)-stimulated β -arrestin1 and β -arrestin2 recruitment to GIPR-Nluc reference in the absence (dark grey) or presence (light grey) of Nluc-GIPR non-reference isoform 1, or, to GIPR-Nluc non-reference isoform 1 in the absence (pink) or presence (light pink) of Nluc-GIPR reference. Data are the mean \pm SEM of n=4 performed in triplicate and data were normalized to GIPR-Nluc reference alone. **k**, Cell surface expression of FLAG-GIPR reference in the absence (dark grey, n=5) or presence (light grey, n=4) of Nluc-GIPR non-reference isoform 1, or, Nluc-GIPR non-reference isoform 1 in the absence (pink, n=5) or presence (light pink, n=5) of FLAG-GIPR reference. All experiments were performed in duplicate and data normalized to either FLAG-GIPR reference alone, or, Nluc-GIPR non-reference isoform 1 alone. Significance was determined by Mann-Whitney test. Box plots show the max and min with the median, 25th, and 75th percentile as well as the mean (+).



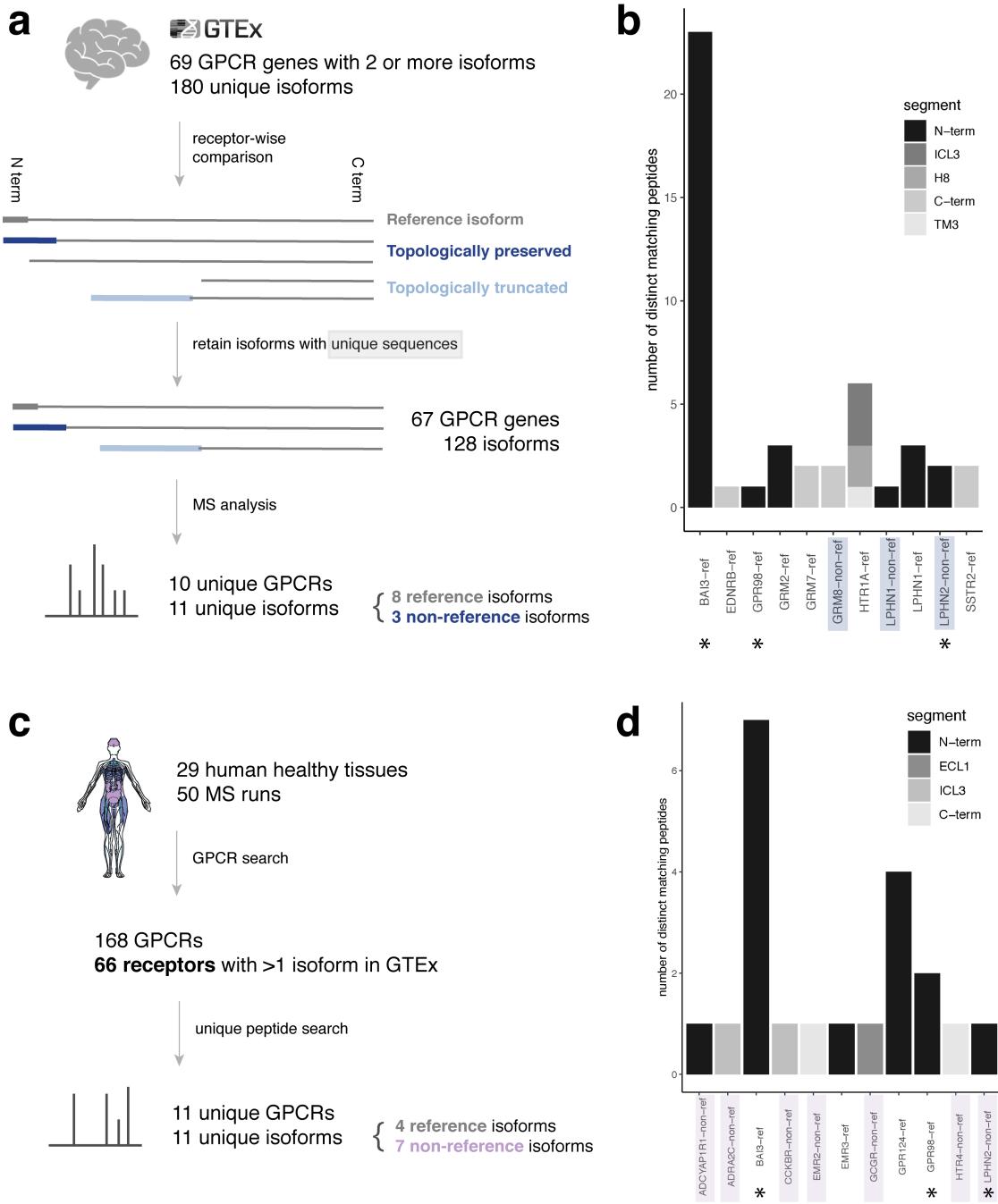
Extended Figure 6. Receptor isoforms and drug targets. **a**, the number of approved drugs greatly varies for receptors with the same number of tissue expression signatures. Count plot of the 111 GPCRs that are targeted by 474 FDA-approved drugs. The radius of the circle indicates the number of GPCR targets for each combination of ‘number of expression signatures’ (x-axis) versus ‘number of approved drugs’ (y-axis). **b**, Alignment of the N-terminal segment of the reference and non-reference GPR35 isoforms. Varying regions in isoform 1 are highlighted in blue. **c**, pEC₅₀ values for Gα_{i3} and β-arrestin 2 coupling to the reference GPR35 isoform and isoform 1 in response to iodoxamide, pamoic acid and zaprinast. **c**, Potency (pEC₅₀) and maximal signal intensity (Emax) of different agonists at GPR35 (error values correspond to standard error of the mean).



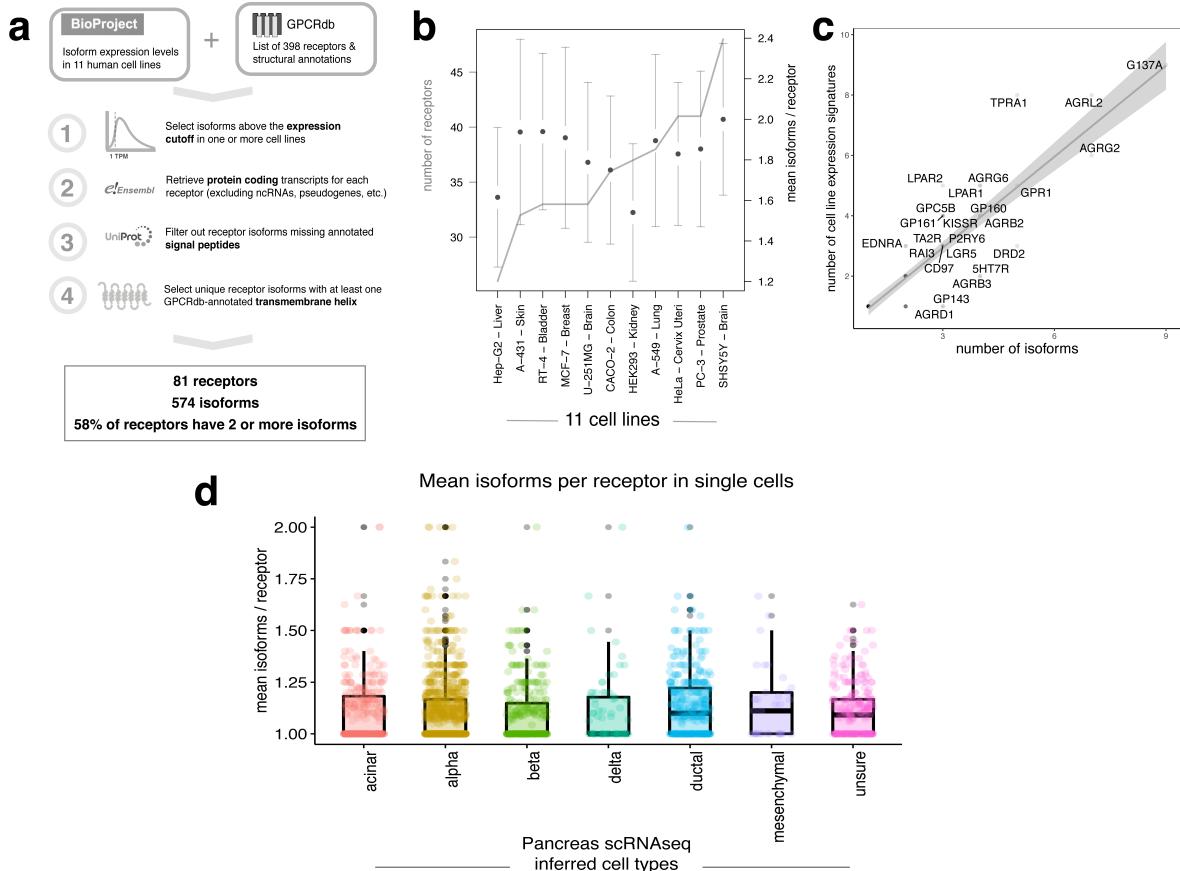
Extended Figure 7. Single Nucleotide Polymorphisms (SNPs) with disease phenotype associations in Gene ATLAS that exclusively map to non-reference receptor isoforms. Disease-related phenotypes that can be exclusively linked to non-reference receptor isoforms were extracted from Gene ATLAS (see **Methods** for further details and **Supplementary Table 3**). Non-reference isoform sequences for the N-terminal segment of the CXCR3 chemokine receptor and C-terminal segment of the metabotropic glutamate receptor 8 (GRM8) are represented in blue while the reference isoform appears in grey. Polymorphisms are indicated in bold together with their associated phenotypes (light grey boxes).



Extended Figure 8. GPCRdb receptor isoform browser. Screenshot of the isoform browser as implemented in GPCRdb. The classification tree (left) depicts the number of unique isoforms for each receptor, receptor family, and GPCR class. In the structural annotation table (right), conserved, partially conserved, and non-conserved structural segments are coloured in green, orange, and red. Purple triangles indicate insertions. Each entry includes the receptor name, isoform number (assigned according to sequence length, with the longest non-reference isoform considered isoform 1) as well as Ensembl transcript IDs, topological preservation or truncation status, tissues in which it has been found to be expressed, an average value of segment completeness with respect to the reference receptor sequence (Ref (%)), its functional annotation (either inferred through structural fingerprint identity with other characterised isoforms, or derived from the literature for isoforms with an associated PubMed ID reference), and a link to its isoform-level proteomics evidence in ProteomicsDB.



Extended Figure 9. Analysis of Mass Spectrometry (MS)-based proteomics data for receptor isoforms. **a**, Analysis pipeline for investigating proteomics data to detect GPCR isoforms that were identified using transcriptomics data at an isoform level: all transcripts from GPCRs with multiple isoforms detected in brain tissue through our GTEx analysis were filtered to identify those with unique peptide sequences. These transcripts where then matched with brain proteomics hits to find how many receptor isoforms with unique protein sequences detected by transcriptomics could be identified by proteomics. **b**, Analysis of the number of matching peptides found in all receptor isoforms considering receptor segments. Non-reference isoforms are highlighted in blue shaded boxes. Consistent with previous studies, non-transmembrane receptor regions are the ones that are often detected by MS. **c**, Proteomic detection of isoforms using MS data from a study of 29 healthy human tissues and 50 MS runs (see Methods). Searching for GPCRs detected in GTEx in this dataset identified a number of receptors with more than one isoform. Of these, we filtered isoforms that had peptides matching one isoform only. **d**, Analysis of the number of matching peptides found in all receptor isoforms in **c** considering receptor segments. Non-reference isoforms are highlighted in violet. As seen in **b**, non-transmembrane receptor regions are the ones that are more often detected by MS. Isoforms with an asterisk are found in both datasets. See all data in **Supplementary Table 5**.



Extended Figure 10. Receptor-centric isoform distribution in cell lines and single cells. **a**, Analysis pipeline combining isoform-level data extracted from transcriptomics experiments of 11 human cell lines deposited in BioProject 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 (**Methods**). **b**, Relationship between the number of GPCRs expressed in the 11 different cell lines (left axis, grey line) and the mean number of isoforms per receptor found in each tissue (right axis, mean and standard error bars shown in black). **c** Relationship between the total number of isoforms per receptor and its number of cell-line expression signatures in the analysed cell lines. The dark grey regression line was obtained using a linear model and the light grey confidence interval represents the standard error. **d**, Mean number of isoforms per receptor expressed in single pancreatic cells as measured by scRNAseq, with each point representing the mean number of isoforms per receptor in a single cell whose transcriptome was sequenced. This analysis shows that, for every inferred cell type, there are cells expressing more than one isoform per receptor. See underlying data for **a**, **b**, **c** and **d** in [Supplementary Table 6](#).