Common coupling map advances GPCR-G protein selectivity

Alexander S. Hauser[1], Charlotte Avet[2], Claire Normand[3], Arturo Mancini[3], Asuka Inoue[4], Michel Bouvier[2*] and David E. Gloriam[1*]

[1] Department of Drug Design and Pharmacology, University of Copenhagen, 2100 Copenhagen, Denmark.

[2] Institute for Research in Immunology and Cancer (IRIC), and Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, Québec, H3T 1J4, Canada.

[3] Domain Therapeutics North America, Montréal, Québec, H4S 1Z9, Canada.

[4] Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Miyagi 980-8578, Japan.

* Corresponding authors:

Data analysis: David E. Gloriam, david.gloriam@sund.ku.dk

Pharmacology: Michel Bouvier, michel.bouvier@umontreal.ca

Abstract

Two-thirds of human hormones and one-third of clinical drugs act on membrane receptors that couple to G proteins to achieve appropriate functional responses. While G protein transducers from literature are annotated in the Guide to Pharmacology database, two recent large-scale datasets now expand the receptor-G protein 'couplome'. However, these three datasets differ in scope and reported G protein couplings giving different coverage and conclusions on GPCR-G protein signaling. Here, we report a common coupling map uncovering novel couplings supported by both large-scale studies, the selectivity/promiscuity of GPCRs and G proteins, and how the co-coupling and co-expression of G proteins compare to the families from phylogenetic relationships. The coupling map and insights on GPCR-G protein selectivity will catalyze advances in receptor research and cellular signaling towards the exploitation of G protein signaling bias in design of safer drugs.

Introduction

G protein-coupled receptors (GPCRs) represent the largest family of proteins involved in signal propagation across biological membranes. They recognize a vast diversity of signals going from photons and odors to neurotransmitters, hormones, and cytokines (1). Their main signaling modality involves the engagement and activation of G proteins. G proteins are heterotrimeric proteins consisting of a α, β and γ subunits that dissociate to α and a βγ upon activation by a GPCR. G proteins are named by their α subunit (16 in human) and are divided into four families which share homology and downstream signaling pathways: G_s (G_s and G_{olf}), $G_{i/o}$ (G_{i1}, G_{i2}, G_{i3}, G_o, G_z, G_{t1}, G_{t2} and G_{gust}), $G_{q/11}$ (G_q, G_{11}, G_{14} and G_{15}) and $G_{12/13}$ (G_{12} and G_{13}). A GPCR's activation of G proteins can be very selective or promiscuous and change upon ligand-dependent biased signaling that alters its profile on the G protein subtype or family levels. The pleiotropic signaling and ligand-dependent bias of GPCRs pose a grand challenge in human biology to map the differential activation of specific G proteins.

The IUPHAR/BPS Guide to Pharmacology (GtP) database contains reference data from expert curation of literature (1). GtP couplings covers 253 GPCRs and the four G protein families. The G protein families have been classified into "primary" and "secondary" transducers without quantitative values. Recently, the Inoue group determined the first large-scale quantitative coupling profiles of 148 GPCRs to the G_q wildtype and 10 G protein chimeras employing a TGF-α shedding assay (2) (NTS_1 and TRH_1 added herein making it 150 receptors). Those chimeras consist of a G_q backbone in which the six most C-terminal Gα residues – a part of the H5 domain inserting to the intracellular receptor cavity – have been replaced to represent all 16 human G proteins (five of which have identical sequences to other G proteins, see below). In a paper accompanying the present analysis, the Bouvier group quantified the couplings of 100 GPCRs to 12 G proteins: G_s, G_{i1}, G_{i2}, G_o (G_{oA} and G_{oB} isoforms), G_z, G_q, G_{11}, G_{14}, G_{15}, G_{12} and G_{13} but not G_{olf} (couples mainly to olfactory receptors), G_{i3} and $G_{t1\text{-}2}$ (Transducin, couples to rhodopsin (visual) receptors) and G_{gust} (Gustducin, couples to taste receptors) (3). The authors used novel enhanced bystander bioluminescence resonance energy transfer biosensors that allow to monitor G protein activation (G protein Effector Membrane Translocation assay; GEMTA) without need to modify the G protein subunits (except for G_s) or the receptors.

Here, we analyze the GtP, Inoue and Bouvier coupling datasets to determine confident couplings supported by at least two independent sources, including novel couplings discovered jointly by the two latter sources. We establish a scalable protocol to normalize quantitative G protein couplings, combine E_{max} and EC_{50} into a common $\log(E_{max}/EC_{50})$ (4) value and aggregate subtypes to allow comparisons across G protein families. On this basis, we develop a unified map of GPCR-G protein couplings that can be filtered or intersected in GproteinDb (5), describe GPCR-G protein selectivity across an unprecedented number of receptors and coupling data points and reveal correlated co-couplings.

Results

Tools and Resources – Common coupling map unifying GPCR-G protein data

Systematic profiling doubles the average number of G protein family couplings of GPCRs

To obtain an overview of the coverages of GPCR-G protein coupling sources, we compared all couplings reported by both the Bouvier (3) and Inoue (2) groups and annotated in the Guide to Pharmacology database (GtP (1)) (Fig. 1). This shows that the three sources together comprise couplings for 265 (67%) out of the 398 non-olfactory GPCRs and that 70 of these receptors are present in all datasets (Fig. 1B). The Bouvier

and Inoue datasets have collectively quantified individual G protein couplings of 178 receptors using one assay, whereas the remaining 87 receptors have so far only been annotated in GtP on the G protein family level from a multitude of publications and assays.

To allow the comparison of coupling densities and distributions across datasets, we selected the E_{max} threshold (1.4 standard deviations above basal) that gives the best agreement between the Bouvier and Inoue dataset. We believe that this is the best possible means to estimate what is correct data (rather than false negative/positive couplings), as large-scale information about what G proteins and GPCRs couple physiologically is not available. This cut-off is more stringent than the minimum of 3% signal above basal used in Inoue's original report (2). We also aggregated G protein subtype couplings of the families (see Methods). This reveals that while GtP covers the largest number of receptors they have relatively few couplings – only 38% of all GPCR-G protein family pairs are 'couplers' compared to 65% in the Bouvier and 75% in the Inoue dataset (average of 1.5, 2.6 and 3.0 G protein families per GPCR, respectively; Fig. 1C). In particular $G_{12/13}$ couplings are underrepresented in GtP where they account for 3% of GPCR-G protein pair datapoints compared to 13% in Bouvier and 17% in Inoue. The Bouvier and Inoue datasets share a more similar overall distribution of couplings, expect for G_s coupling which is twice as frequent within the latter dataset (16% compared to 8%). This demonstrates that the two first systematic coupling profiling studies have substantially expanded the known GPCR-G protein 'couplome' and that their assay platforms are amenable to high-throughput profiling also for the $G_{12/13}$ family for which robust activation assays appeared only recently (6-8).

Given that the Bouvier and Inoue groups used different biosensors and GtP annotates literature reports from very diverse assays, we sought to determine to which extent they report the same couplings for the same GPCRs i.e., their 70 common receptors (data in tab 'BIG-QualComp' in Spreadsheet S3). We find that all three sources agree on an average of 49% of G protein family couplings/non-couplings (distributed as G_s: 55%, $G_{i/o}$: 64%, $G_{q/11}$: 50% and $G_{12/13}$: 25%, Fig. 1D). When instead analyzing the quantitative studies alone (excluding GtP), the agreement increases to 70% across G protein families. This agreement is 68% for individual G protein datapoints, which display an even larger span ranging from at least 53% for G_{15} up to 81% for G_{oA} (Fig. S1). These findings define a sizeable reference set of consensus G protein couplings and show that consistent large-scale profiling studies generate more comparable results than literature.

Common coupling map unifies GPCR-G protein datasets and opens analysis

To enable quantitative correlation of the Bouvier, Inoue and future couplings, we established a data processing protocol giving the highest similarity and correlation of coupling measurements across datasets (Methods). This uses log transformed EC_{50} values and minimum-maximum normalized E_{max} values combined into a unified $\log(E_{max}/EC_{50})$ value and an aggregation of G proteins onto families based on the maximum subtype values (Fig. 2A). Based on this protocol, we created a common coupling map, integrating the Bouvier, Inoue and GtP datasets (Fig. 2B). This unified coupling map establishes that it is possible to obtain comparable quantitative values despite the differences between biosensors and enables quantitative cross-study comparisons herein and in future studies from the field.

To enable any researcher to use the coupling map, we have availed a "G protein couplings" browser (https://gproteindb.org/signprot/couplings) in GproteinDb (7). By default, the coupling map (Fig. 2B) and browser only shows "supported" couplings with evidence from two datasets, but there is an option (first blue button) to changes the level of support to only one (for most complete coverage of GPCRs) or to three (for the highest confidence) sources. An exception, GtP couplings do not require additional support, as they are in most cases supported by multiple independent publications. We propose a standardized terminology

118 to describe couplings based on their level of experimental support from independent groups (Table 1). The
119 criterion of supporting independent data, and the terms "proposed" and "supported", are already used by
120 the Nomenclature Committee of the International Union of Basic and Clinical Pharmacology (NC-
121 IUPHAR) for GPCR deorphanization. Furthermore, the online coupling browser allows any researcher to
122 use only a subset of datasets, or to apply filters to the Log(Emax/EC50), Emax, and EC50 values. Finally,
123 users can filter datapoints based on a statistical reliability score in the form of the number of SDs from basal
124 response.
125

126 Research Advances – Insights on GPCR-G protein selectivity

127 The Bouvier and Inoue datasets jointly support 101 novel G protein couplings

128 We next identified the 'novel' G protein couplings for which a family annotation is missing in GtP but have
129 high confidence from dual support by the Bouvier and Inoue groups (Fig. S1). This revealed 38 receptors
130 with novel couplings to 101 G proteins distributed across all families: $G_{s,}$: 4, $G_{i/o}$: 15, $G_{q/11}$: 10 and $G_{12/13}$:
131 21 (Fig. 3). The largest expansions – an increase by three of G protein families – was obtained for the
132 histamine H_1 and endothelin ET_A receptors which was found to couple to all G protein families but only
133 have $G_{q/11}$-coupling in GtP. Whereas it could be expected that GtP would miss couplings, we also analyzed
134 if its expert curation excluded couplings that may be false positives as they are contradicted by both
135 quantitative studies. This uncovered such G_s-coupling to the α_{2C}-adrenoceptor and cannabinoid CB_{1-2}
136 receptors, $G_{12/13}$ coupling to the purinergic $P2Y_2$ receptor and $G_{i/o}$-coupling to the β_2-adrenoceptor, which
137 however had weak G_z and G_{oB} coupling in the Bouvier study but did not cross the signal threshold (3).
138 Notably, this is only 2% (5/254) of all GtP's GPCR-G protein family pairs. Taken together, these findings
139 serve to quantify the expansion of the GPCR-G protein 'couplome' while also confirming the outstanding
140 accuracy of the expert annotation in the GtP database. Therefore, the large number of new couplings are
141 mainly due to missing investigations in literature rather than incorrect curation.
142

143 Half of GPCRs are selective for a single whereas 5% promiscuously activate all G protein
144 families

145 To gain insight into their levels of coupling selectivity, we intersected the G protein profiles of all receptors
146 and counted the number of coupling partners for GPCRs and G proteins (Fig. 4). On the G protein family
147 level (topmost in Fig. 4), our analysis spans 90 receptors with data only in GtP and 166 GPCRs with data
148 from the Bouvier and/or Inoue groups – totaling 256 receptors. We require couplings from the Bouvier and
149 Inoue groups to be supported by a second dataset (GtP couplings are already typically supported by several
150 publications). We find that these 256 GPCRs couple to on average 1.7 G protein families distributed as 126
151 single-, 83 double-, 34 triple- and 13 all-family activating receptors. The share of fully selective (single-
152 family activating) receptors differs largely across G protein families spanning from 6% for $G_{12/13}$ to 22%
153 for $G_{q/11}$, 26% for G_s and up to 40% for $G_{i/o}$ (tab 'Fig_4' in Spreadsheet S5).
154 Interestingly, all fully promiscuous receptors are class A GPCRs: adenosine A_1, adrenergic $\alpha_{1A,2A}$ and
155 β_1, bradykinin B_2, cannabinoid CB_1, cholecystokinin CCK_1, endothelin ET_A, prostanoid FP, GPR4,
156 histamine H_1, lysophospholipid LPA_4 and orexin OX_2 receptors (216 class A, 11 class B1 and 5 class C
157 GPCRs have been profiled, so far). Conversely, a G protein family has supported couplings to on average
158 112 GPCRs (28% of all receptors) distributed as G_s: 87, $G_{i/o}$ 176, $G_{q/11}$: 134 and $G_{12/13}$: 49 receptors (34%,
159 69%, 52% and 19%, respectively of all receptors, Fig. 4C). Given that 101 of the GPCR-G protein family
160 pairs tested by the Bouvier and Inoue groups represent novel couplings (above) more couplings are expected
161 to be identified as expanding and confirmatory studies emerge. Hence, whereas the results described here

represent the currently known supported couplings, the total 'couplome' will undoubtedly comprise additional yet undetected and unconfirmed couplings, especially among receptors never profiled with a pan-G protein platform.

Three-quarters of GPCR activate all G proteins belonging to the same family

Within each G protein family (rows 2-5 in Fig. 4), on average 73% of GPCRs promiscuously activate all its members ($G_{q/11}$: 73%, $G_{12/13}$: 66%, $G_{i/o}$: 75% and G_s: 80%). In contrast, activation of only one subtype of a G protein family is only observed for 11 G_s, 4 G_z, 1 G_{14}, 10 G_{15}, 5 G_{12} and 10 G_{13}-coupled receptors (Fig. 2 or Spreadsheet S5). Most other receptors activate a subset of G proteins in each family. Strikingly, $P2Y_{1,4}$ (G_{15}), $P2Y_{14}$ (G_z) and GPR55 (G_{13}) are fully selective also when considering G proteins from all families i.e., they only couple to a single of the 16 human G proteins (1.4*SD cut-off applied, 6 G_s-coupling receptors are left out, as G_{olf} has so far only been tested by the Inoue group). However, the three purinergic receptors have additional couplings although not supported by a second dataset ($P2Y_4$ and $P2Y_{14}$) or above the 1.4*SD cut-off ($P2Y_1$ was below) and it is possible that as the characterizations expand even fewer, or no receptors are found to engage only a single G protein.

The abundant coupling of $G_{i/o}$- and $G_{q/11}$ to many receptors with dual (or more) pathways suggests that these G proteins often have more versatile functions. In all, our meta-analysis of GPCR-G protein selectivity point to intriguing differences where some receptors selectively signal via a single effector whereas other GPCRs promiscuously activate all four G protein pathways. Such selective or combined profiles, in interplay with differential spatiotemporal expression (3), can be critical to achieve a specific physiological effect.

G protein co-coupling reflects phylogeny, but all G protein families have an odd member

To assess whether the evolutionary classification dividing G proteins into four families is also representative of their pharmacology, we investigated the correlated coupling of G proteins to receptors. The overall correlation was assessed with Pearson standard correlation coefficients (Fig. 5A-B), and broken down into shared coupling/non-coupling (from Jaccard indices, Fig. 5C) and activation levels (mean log($Emax/EC_{50}$, Fig. 5D). We find that all three comparisons show the strongest correlations for G proteins belonging to the same G protein family (boxed in Fig. 5A, C-D). This demonstrates that the pharmacological relationships, i.e., the coupling selectivity and activation level, do indeed reflect the phylogenetic relationships. This is important, as this is how the human G proteins into four families have been classified traditionally, but it had not been tested if this classification applies to pharmacological profiles of the scales analyzed herein and likely increasingly more common in future studies.

Several G protein pairs stand out with exceptionally high overall correlated coupling (Pearson standard correlation coefficient): G_{i1}-G_{i2} (0.99), G_{oA}-G_{oB} (0.96) and G_q-G_{11} (0.99). Interestingly, the correlated coupling is lower for the only pairs within the G_s and $G_{12/13}$ families, G_s-G_{olf} (0.77) and G_{12}-G_{13} (0.79), respectively. Furthermore, both $G_{i/o}$ and $G_{q/11}$ have an 'odd' member, G_z and G_{15}, respectively (mean of 0.78 and 0.76 to other family members, respectively). The fact that all G protein families have an odd member shows that all signaling pathways have a transducer toolbox allowing them to differentiate signaling.

G_s or $G_{i/o}$ coupling is selective while $G_{12/13}$ is promiscuously activated with $G_{i/o}$ and $G_{q/11}$

The G_s and $G_{i/o}$ families have an inverse correlation in all of the overall, coupling/non-coupling and activation level comparisons (darkest red in Fig. 5A, C-D). This means that G_s and $G_{i/o}$ rarely co-couple to

206 GPCRs and, when they do, they do so with a large difference in their strength of activation. This inverse
207 correlation where only one or the other G protein is activated is in agreement with their function, as G_s
208 stimulates, and $G_{i/o}$ inhibits production of the same cellular second messenger, cAMP. This allows opposite
209 physiological responses to be mediated with temporal selectivity in the same cell by activating different
210 receptors at different times. Differential engagement of the $G_{i/o}$ family could also have functional
211 consequences through other pathways, as G_i was recently shown to be required for scaffolding β-arrestin
212 binding and signaling for some receptors (9).
213 For $G_{12/13}$, we instead find a positive correlation with the $G_{i/o}$ and $G_{q/11}$ families. This is in agreement
214 with Fig. 4 (first Venn) which shows that only 3 GPCRs couple to only $G_{12/13}$ and 3 couple to the rare $G_{12/13}$-
215 G_s family pair, whereas the vast majority of GPCRs, 38 couple to $G_{12/13}$ also couple to the $G_{i/o}$ and/or $G_{q/11}$
216 family. Furthermore, we find an intriguing unique selectivity mechanism for the $G_{i/o}$ and $G_{q/11}$ families.
217 These families have the most frequent co-coupling (light blue in Fig. 5C) but have a high difference in
218 average activation levels (red in Fig. 5D). Together, this shows that selective binding of only $G_{i/o}$ or $G_{q/11}$
219 is uncommon, but selectivity can be achieved by differential levels of activation. This possibility to achieve
220 selectivity via differential activation is likely important, as many GPCRs, 81 couple to both the $G_{i/o}$ and
221 $G_{q/11}$ families. Their high difference in activation levels allow for a selective activation, which could shift
222 upon the presentation of alternative endogenous or surrogate ligands with a signaling bias. Furthermore, it
223 should be pointed out that many reported dual couplings to G_i and G_q may be because PLC, a direct effector
224 of G_q, can also be activated by $G_{\beta\gamma}$ from G_i (i.e. sensitive to the G_i inhibitor pertussis toxin). Thus, the
225 concept of potential scaffolding of PLC by G_q may be required for G_i-derived $G_{\beta\gamma}$ activation (10).
226
227 **Differential tissue expression gives G proteins in the same family large spatial selectivity**
228 To study how tissue expression may influence G protein selectivity, we analyzed consensus transcript
229 expression levels for 50 tissues and 16 organs (here aggregated into 8) from the Human Protein Atlas (11),
230 which also includes data from the Genotype-Tissue Expression (GTeX) project (12) (Fig. 6). We find that
231 the most ubiquitously expressed G proteins are G_s and G_{i2} which are expressed in all 50 tissues and all
232 organ categories at a level that is within the first quartile of all normalized transcripts per million (nTPM)
233 values. In contrast, G_{i3}, G_{t1}, G_{t2}, G_{t3}, and G_{14} have none or only one tissue with a first quartile expression.
234 To facilitate comparison across G proteins, we employed HPA's normalized transcripts per million (nTPM)
235 values after z-score transformation of each G protein across tissues. A z-transformation of the same G
236 proteins visualizing the relative tissue levels of expression, shows a very focused expression of the
237 transducins, G_{t1} and G_{t2}, in the retina and gustducin G_{t3} in the gastrointestinal tract reflecting their role in
238 vision and taste, respectively (Fig. S2A). Notably, each G protein family has at least one subtype that is
239 preferentially expressed in the brain: G_s: G_{olf}, $G_{i/o}$: G_o and G_z, $G_{q/11}$: G_q and $G_{12/13}$: G_{12} (Fig. 6 and Fig. S2A).
240 Given that the tissue expression varies largely for G proteins, we next sought to determine to what
241 extent their tissue expression correlations differ from the traditional four G protein families that represent
242 phylogeny and GPCR coupling (Fig. 5) relationships. To this end, we calculated the co-expression for each
243 G protein pair using Pearson standard correlation coefficients (Fig. S2B). Notably, we find that all four G
244 protein families have members that group apart in the dendrogram (the two members of the G_s family, G_s
245 and G_{olf}, do not cluster adjacently but not as a pair). Instead, we find a coherent cluster for the G proteins
246 that are predominantly expressed in the brain, G_{olf}, G_o, G_z, G_q and G_{12} and cover all four families.
247 Furthermore, among the remaining co-expressing G proteins, three out of the four closest pairs also span
248 families (G_{i3}-G_{15}, G_{i2}-G_{13} and G_{t3}-G_{11} but not G_{t1}-G_{t2}). Taken together, this shows that there is no overall
249 clustering of subtypes within the G protein families, which instead exploit differential expression to gain

spatial selectivity. For example, in the $G_{i/o}$ family G_o and G_z are restricted to brain regions, while the other subtypes have four different peripheral profiles.

Technological considerations – Biosensor sensitivity

Bouvier's and Inoue's biosensors appear more sensitive for G_{15} and, G_s and G_{12}, respectively

To evaluate the sensitivity of the Bouvier and Inoue biosensors for different G proteins, we compared their observed couplings for the 70 common receptors to identify 'unique' and 'missing' couplings (defined in Table 1). For G_{15}, Bouvier reported 30% unique couplings, several of which were validated in Ca^{2+} assays (3), whereas Inoue instead misses 13% of couplings present in GtP (Fig. 7). This indicates that G_{15} couplings are underrepresented in Inoue's data and in literature (annotated in GtP). The few published couplings for G_{15} is likely explained by its lack of expression in the cells most used for *in vitro* experiments, such as HEK293 cells (3), the lack until recently of enough sensitive sensors to directly measuring G_{15} activity, and by the lack or weak (at very high concentration) effect on this subtype by $G_{q/11}$ inhibitor tools YM-254890 and FR900359, respectively (13).

For G_s and G_{12}, 27% and 24%, respectively of Inoue's couplings are unique, whereas Bouvier instead misses 7% of G_s and 6% of G_{12} couplings in GtP. These two G proteins have smaller assay windows which may explain their lower representation in the Bouvier dataset (tab 'DataStats' in Spreadsheet S1). Furthermore, for G_q, G_{11} and G_{14}, 13%, 16% and 16%, respectively of Inoue's couplings unique. In contrast, Bouvier is missing 10%, 10% and 7%, respectively of the known couplings in GtP which indicates an underrepresentation or alternatively that literature has mistaken G_i-derived βγ activation of the PLC-IP3-Calcium signaling pathway for G_q mediated signaling (14-17). Notably, although Inoue used the wildtype protein for G_q, it has a similar or higher frequency of unique couplings compared to all other measured G proteins (G_q: 13% vs. G_{i1}: 7%, G_{i2}: 9%, G_{oA}: 7%, G_{oB}: 6%, G_z: 13%, G_{11}: 16%, G_{14}: 16%, G_{15}: and G_{13}: 17%), except G_s and G_{12}. This suggests that the many unique couplings in Inoue's dataset may be due to high sensitivity rather than the use of chimeric G proteins carrying over of G_q coupling to the other G proteins.

To investigate how unique couplings are affected by weak couplings and biosensor sensitivity, we compared their efficacies and potencies (rows 106-110 in tab 'BIG-QualComp' in Spreadsheet S3). We find that the average E_{max} values are 18% and 15% lower for unique than for supported couplings in the Bouvier and Inoue datasets, respectively. Furthermore, their average pEC_{50} values are 0.4 and 1.5 log units – 2.5-fold and 32-fold, respectively – lower than the supported couplings. This indicates that a part of the differences across datasets is due to weak couplings that can only be detected in the most sensitive assays for the given G proteins and are difficult to distinguish from basal levels (this study used a cut-off of E_{max}>1.4 SDs over basal response). However, couplings may also be overrepresented and determining couplings definitely would require further studies employing independent ('orthogonal') high-sensitivity biosensors. Until then, we recommend the requirement of two independent studies to support of a GPCR-G protein coupling, as is the criterion used herein for the novel couplings and the default in the online G protein couplings resource (1).

Discussion

Given that researchers are now faced with three large G protein coupling datasets varying in coverage and couplings, we established a common normalization protocol making quantified values comparable across data sources and a common coupling map. Previous analyses of GPCR-G protein selectivity have all been

294 based on a single dataset from the GtP (18) database or the Inoue (2) or Bouvier (3) laboratories, whereas
295 the analysis herein combined these datasets to establish couplings supported by two studies (including GtP
296 annotations of multiple literature reports) for a total of 256 GPCRs. This provides a sizeable and reliable
297 reference dataset of supported quantitative couplings suitable for any study.
298 By analyzing the common coupling map, we find that 101 novel couplings are supported by both
299 the Bouvier and Inoue datasets but were not known in GtP which annotates coupling from the literature.
300 We also find that very few receptors, 13 (5%) promiscuously couple to all four G protein families whereas
301 the receptor numbers instead increase several-fold along with selectivity: 34 triple-, 83 double- and 126
302 single-family coupling receptors (Fig. 4B). As more large coupling datasets become available, this is likely
303 to shift somewhat from less selectivity towards higher promiscuity. Furthermore, the finding that on average
304 only 27% of GPCRs activate only a subset of the G proteins in each family (Fig. 4) opens the possibility
305 that signaling bias on the level of specific subtypes may be much more frequent than currently appreciated.
306 G proteins that belong to the same family can have diverse functional outcomes pertaining to effector
307 engagement selectivity and kinetic profiles (19, 20), suggesting that bias within a G protein family may
308 have physiological and/or therapeutic implications (21). The observation that many GPCRs can couple to
309 more than one family of G proteins but still can show selectivity between members of a same family, opens
310 important questions concerning the structural determinants of such multifaceted selectivity profiles.
311 The GPCR-G protein couplings analyzed herein represent the overall couplings that are possible, but
312 may differ largely physiologically across specific tissues. Our analysis showed a tremendous span in how
313 many tissues that express a given G protein. Some G proteins, including the transducins (G_{t1} and G_{t2}) and
314 gustducin (G_{t3}) have a restricted expression in mainly one tissue (retina and the gastrointestinal tract,
315 respectively), whereas the most ubiquitously expressed G proteins, G_s and G_{i2} are highly expressed in all
316 50 analyzed tissues. Notably, G proteins belonging to the same family differ largely in their tissue profiles
317 for example, each G protein family has a single or two subtypes that is/are preferentially expressed in the
318 brain. Given that the activity (here $\log(E_{max}/EC_{50})$ values) can vary substantially across G protein family
319 subtypes (Fig. 2), this provides cells with a mechanism to shift the system bias towards different
320 intracellular signaling pathways. Such system bias can be modulated further through differential expression
321 of also the receptors and multiple downstream intracellular effectors (22). A recent study found a strong
322 spatial correlation of summed mRNA expression levels for G_s-, $G_{i/o}$-, and $G_{q/11}$-linked receptors in humans,
323 macaques, and mice (23). This suggests that the expression patterns of receptors linked to these major G
324 protein families are strongly interdependent and could therefore act together to balance each other in
325 specific tissues. Future studies would be needed to profile the couplings of natively expressed GPCRs and
326 G proteins across different cell types. Meanwhile, a strategy could be to analyze GPCR-G protein selectivity
327 for a subset of G proteins that are selected based on the highest expression in the tissue of interest. For
328 example, analysis of $G_{i/o}$-mediated signaling in the brain could focus on G_o and G_z based on their high
329 expression, and the fact that G_o and G_z have quicker and slower nucleotide exchange, respectively than
330 other member of this family. Such separate meta-analysis of G proteins may be facilitated by the filtering
331 options in the coupling maps in Spreadsheet S5 and online in the G protein couplings browser in
332 GproteinDb (5).
333 The Inoue and Bouvier datasets represent the first steps to systematically characterize large G protein
334 coupling profiles. Other laboratories may use other biosensors to expand the coverage of the coupling map,
335 and our study facilitates this by indicating which of the Bouvier and Inoue biosensors that have the highest
336 sensitivity for each G protein (Fig. 7). However, future studies should also employ different biosensors and
337 systems, as independent support is critical for validation and for distinguishing novel couplings from false

338 positives among the many unique couplings (Fig. 7 and Spreadsheet S3). Of note, several biosensors have
339 recently been published (7, 8, 24) (Gαβγ sensors first described in (25-30)). Whereas a detailed comparison
340 is out of the scope of this study, their pros and cons have recently been reviewed (31) and researchers should
341 strive to use the most native proteins and expression, physiologically relevant tissues/cells (32) and assays
342 with sufficient sensitivity and window. Of note, most studies annotated in GtP have used downstream
343 measurements, which are amplified, and may therefore differ from those obtained by biosensors that
344 measure G protein activation (further discussed in (3)).
345 In all, our cross-dataset analysis has established a protocol and reference set aiding GPCR-G protein
346 coupling studies. The selectivity profiles are the most comprehensive to date spanning 256 receptors with
347 a very diverse activation of a single to all G protein pathways, and presented in a dedicated online G protein
348 couplings browser in GproteinDb (18). The analyses and data presented herein will be very valuable to
349 illuminate undercharacterised pharmacological phenomena such as constitutive activity (33), pre-coupling
350 of G proteins (26, 34) and ligand-dependent biased G protein signaling (35), and to uncover their underlying
351 determinants. They also present the foundation to integrate more coupling data as future studies expand the
352 characterization of the 'couplome'.
353
354
355 # Methods
356 ## Study design
357 The primary objective of this study was to generate a unified map of G protein couplings (Fig. 2) across the
358 three available large datasets from the Bouvier (3) and Inoue (2) groups, and the Guide to Pharmacology
359 database (1), respectively. To do this, we identified the Emax standard deviation cut-off, quantitative
360 normalization protocol and aggregation of G proteins into families giving the best possible agreement
361 between the GPCR-G protein couplings from the Bouvier and Inoue groups. This analysis also involved
362 assessing the agreement of the Bouvier and Inoue group datasets and determining the number of high-
363 confidence novel couplings supported by these two datasets but not reported in the Guide to Pharmacology
364 database. Furthermore, it included a benchmarking of techniques to determine which biosensor produces
365 the most reproducible qualitative coupling determination (coupling vs. non-coupling) for each G protein
366 family. To this end, we compared the three datasets to pinpoint, for each dataset and G protein family, the
367 fraction of couplings supported by another dataset. The scientific analyses based on the map feature the
368 most comprehensive analysis to date of GPCR-G protein selectivity. This was done by intersecting the 11
369 G proteins tested by both Bouvier and Inoue, within and across their families, with respect to the common
370 and unique receptors that they couple to. Finally, the receptor profiles were also used to classify G proteins
371 and to determine any co-correlation between different G protein subtypes and families to GPCRs.
372
373 ## Coupling datasets
374 Updated spreadsheets containing the pEC_{50}, E_{max}, basal signals and standard deviation (SD) values were
375 supplied by Asuka Inoue. Basal signals, spontaneous AP-TGF-α elease (in % of total AP- TGF-α
376 expression) were recalculated from raw data of the previous coupling-profiling campaign (2) and their SD
377 values were computed from independent experiments ($n \geq 3$). This file contains the neurotensin 1 (NTS_1)
378 and Thyrotropin-releasing hormone (TRH_1) receptors not included in the previous publication (2). In the
379 Inoue dataset, the protease-activated receptors PAR3-4 had negative pEC_{50} values (concentration of mU/ml
380 because the ligand, thrombin, was supplied with its enzymatic activity). For the easiness of integration into
381 the coupling map, we added a value of 10 to their pEC50 values. Data qualities (sigmoidal curves) for the

individual GPCR-G protein pairs were manually inspected and concentration-response curves that did not converge nor exceed a threshold (typically, 3% AP-TGF-α release) were regarded as no activity.

The Bouvier dataset ($n \geq 3$) contained some datapoints that were included or excluded based on dedicated analyses. Firstly, we excluded ligand-promoted responses of overexpressed receptors that were equivalent to those of endogenously expressed receptors (yellow fill in tab 'B-EmEC' in Spreadsheet S1). Secondly, couplings with only approximate E_{max} and pEC_{50} values because the dose-response curve did not converge were only included if supported, and not contradicted, by the Inoue and/or GtP datasets (orange fill in tab 'B-EmEC' and analyzed separately in tab 'UnconvergedDRV' in Spreadsheet S1). This is because a coupling with an unprecise quantitative value is better than no coupling, especially when making qualitative comparisons (coupling vs. non-coupling). Based on these criteria seven couplings were included: 5-HT_{1D}-G_z, BLT_1-G_{11}, FFA3-G_{13}, GPR_4-Gi_{1-2}, GPR84-G_{oA} and κ-G_{12}, and six couplings were excluded: CCR5-G_{14}, CXCR5-G_q/G_{11}/$_{G14}$, GPR183-G_{13} and κ-G_{13}.

Standard deviation cut-off

We made a special investigation of couplings that have a full dose-response curve but an E_{max} less than 2 standard deviations (SDs) from the basal signal (red fill in tab 'B-EmEC' and analyzed separately in tab tabs 'B-<2SDs' and 'I-<2SDs' in Spreadsheet S1). To achieve the best possible separation of putative false and real but weak couplings, we identified the threshold value (a number of SDs from basal signal) that gives the best agreement between the Bouvier and Inoue dataset among the common receptors and G proteins tested by both groups. The obtained cut-off, 1.4 SDs was applied as a filter to exclude all Bouvier and Inoue E_{max} and pEC_{50} values for couplings below this cut-off (columns with '>1.4SD' in the heading in tabs 'B-EmEC' and 'I-EmEC' in Spreadsheet S1). As a note, while intra-day measurement error is small for the TGF-α shedding assay (typically, 1-2% AP-TGF-α release), inter-day variability varies widely depending on cell conditions. Since the SD represents inter-day variability, the basal SD cut-off removes more couplings than the SD cut-off of ligand-induced signal or the significance of individual experiments. As a consequence, some of manually annotated couplings in the Inoue dataset are regarded as non-coupling by the basal SD cut-off criteria, including for $P2RY_2$ and $P2RY_6$ that had no couplings above this cut-off.

Generating a subset of comparable GPCR-G protein couplings

To enable qualitative comparison of corresponding datapoints in the datasets from the Bouvier (3) and Inoue (2) groups, we used the subset of 70 GPCRs present in both datasets and belonging to the same class, A (removed only two receptors from class B1). The quantitative comparisons focused on a smaller subset of 51 GPCRs tested with the same ligand and excluded non-coupling GPCR-G protein pairs, as they could not be represented by 0 values (due to e.g., an underrepresentation of couplings in a given datasets and G proteins, see Results). All analyses herein included the 12 G proteins: G_s, G_{i1}, G_{i2}, G_{oA}, G_{oB}, G_z, G_q, G_{11}, G_{14}, G_{15}, G_{12} and G_{13}. G_{olf} and G_{i3} could not be analyzed, as they had not been tested by the Bouvier group (Table S1). The Inoue data for the pairs G_{i1-2}, G_{oA-B} and G_q and G_{11}, respectively, were generated with identical chimera inserting the Gα C-terminal hexamer into a G_q backbone (2). Qualitative analyses compared the presence or absence of each GPCR-G protein coupling while the quantitative analyses were limited to common G protein couplings, i.e., data points in which both the Bouvier and Inoue groups generated a pEC_{50} and E_{max} value.

Validation of comparability when E_{max} or reference agonist differ

To get an overview of the distribution of data in the two datasets, we determined the pEC_{50} and E_{max} mean, median, min, max, span and standard deviation values and plotted box and whiskers plots for each G protein (Spreadsheet S1). This showed that the E_{max} values vary much more across the different G proteins than pEC_{50} values. The E_{max} variation is largest in the Bouvier data wherein G_{15} ranges across three orders of magnitude (16 to 1,067). The Bouvier dataset has low means and spans for G_s (42 and 47) and G12 (71 and 101) indicating a narrow assay signal window (low signal-to-noise). We find that minimum-maximum normalization (to 100%) of each G protein across receptors gives a more uniform distribution (SDs for Bouvier: 19-31 and Inoue: 22-26, rightmost plot pair in Spreadsheet S1).

Whereas both studies tested a majority of receptors with their endogenous ligand, surrogate agonists were used for 15 and 4 GPCRs in the data from the Inoue and Bouvier groups, respectively. Although those ligands were selected for their reference character with similar pharmacology to the endogenous ligand, they could introduce differences in a receptor's G protein profile due to ligand-dependent signaling bias. Analysis of GPCR-G protein couplers and non-couplers (tab 'BI' in Spreadsheet S2) shows that on average 74% and 71% agreeing qualitative couplings (i.e., coupling vs. non-coupling GPCR-G protein pairs) for receptors tested with the same and different agonists, respectively. This is a rather small difference providing confirmation that receptors tested with different ligands can be compared on the qualitative coupling/non-coupling level and be included in the comparison of the different G protein coupling datasets and determination of novel G protein couplings.

Dataset integration into a unified coupling map – Normalization and $\log(E_{max}/EC_{50})$ values

To enable quantitative correlation of the Bouvier and Inoue couplings, we further filtered the 70 common GPCRs to yield 51 common class A GPCRs tested with the same ligand (excluding 29 receptors tested with different ligands and two class B1 GPCRs). To assess the value of normalization, we calculated the average 'similarity' (Bouvier/Inoue ratio) of individual values (tabs ending with '-sim' in Spreadsheet S4) and the 'linear correlation' (r2 value) of each receptor across all G proteins (reported below as averages of individual couplings and G proteins, respectively) (tabs ending with '-r2' in Spreadsheet S4). Linear correlation was only done for receptors with at least three common G proteins/families. Minimum-maximum normalized Emax values were represented as percentage values while decimal values (0-1) were used for the calculation of log(Emax/EC50) values as recommended in (4). For both the Bouvier and Inoue groups, the minimum and maximum represent the signal without (0%) and with (100%) an agonist, respectively, in each experiment replica. The minimum Emax value was therefore set 0 while the maximum was set to the highest value for the given G protein (first, use normalization) or receptor (second, tested but not used normalization).

Minimum-maximum normalization of E_{max} measurements increased the average from 0.22 to 0.61 and the linear correlation r^2 value from 0.28 to 0.30. In contrast, non-normalized pEC_{50} measurements have the most similar values (0.89 compared to 0.66) and an identical linear correlation (0.37). We combined the minimum-maximum normalized E_{max} and non-normalized EC_{50} values into $\log(E_{max}/EC_{50})$ (4). The $\log(E_{max}/EC_{50})$ values have an average similarity that is nearly as high (0.86 compared to 0.89) and an average linear correlation that is better (0.41 compared to 0.37) than for pEC_{50}, the best individual measure. We did not apply a double normalization, i.e., also across G proteins after across receptors, as this worsens the E_{max}, pEC_{50} and $\log(E_{max}/EC_{50})$ value similarity (from 0.61 to 0.39, 0.89 to 0.42 and 0.86 to 0.47, respectively) and linear correlation of each G protein across receptors (from 0.11 to 0.05, 0.49 to 0.20 and 0.46 to 0.23, respectively.

Dataset integration into a unified coupling map – G protein family aggregation

Given that G proteins belong to families that are functionally grouped by sharing downstream signaling pathways, we next investigated the best member-to-family aggregation scheme – specifically, whether the most comparable G protein family values are obtained if using the maximum value from any single subtype or the mean of all subtype members. We found that aggregation using max rather than mean values gives a better similarity for E_{max} (0.69 vs. 0.65), marginally lower similarities for pEC_{50} (0.89 relative 0.90) and identical similarities for $\log(E_{max}/EC_{50})$ (0.88). However, max performs better overall than mean in the comparisons correlating a receptor across G proteins; E_{max} (0.63 and 0.53), pEC_{50} (0.56 and 0.56) and $\log(E_{max}/EC_{50})$ (0.65 and 0.59) or one G protein across receptors; E_{max} (0.29 and 0.20), pEC_{50} (0.71 and 0.74) and $\log(E_{max}/EC_{50})$ (0.69 and 0.71). Notably, the correlation of each receptor across max-aggregated G protein families compared to non-aggregated subtypes is much stronger when considering any of the three pharmacological parameters: E_{max} (0.63 vs. 0.36), pEC_{50} (0.56 vs. 0.30) and $\log(E_{max}/EC_{50})$ (0.65 vs. 0.41). Altogether, this establishes the highest value (max) as the aggregation that yields the most comparable quantitative value for G protein families.

Tissue expression analysis

Consensus transcript expression levels were extracted for 50 different tissues aggregated into 16 organs or systems in the Human Protein Atlas (HPA) (11), which also includes data from the Genotype-Tissue Expression (GTEx) project (12). The G protein expression was quantified in the form of normalized transcripts per million (nTPM), which is a consensus normalized transcript expression value (see https://www.proteinatlas.org/about/assays+annotation#rna). The correlation analysis was performed a default Pearson correlation analysis using Seaborn's clustermap.

Statistical analysis

The aggregated sample size is n=3 or higher for all analyzed GPCR-G protein couplings. The specific sample size for each such coupling is described in the original articles reporting these data and referenced in the present manuscript. The Bouvier dataset contained some datapoints that were included or excluded based on dedicated analyses (see 'Coupling datasets' above). For all figures, the values for *N*, *P*, and the specific statistical test is included in the figure legend or associated manuscript text. For Fig. 5, pairwise correlation of $\log(Emax/EC_{50})$ values was assessed as a measure of the strength of the linear relationship between two pathways. The pairwise distances between pathways were calculated in python scipy package using the *spatial.distance.pdist* method using the *correlation* metric. Statistical significance was determined by Pearson standard correlation coefficient with a two-tailed p-value as implemented in scipy's *stats.pearsonr* method.

References

1. Armstrong JF, Faccenda E, Harding SD, Pawson AJ, Southan C, Sharman JL, et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to MALARIA PHARMACOLOGY. Nucleic Acids Res. 2020;48(D1):D1006-D21. 10.1093/nar/gkz951
2. Inoue A, Raimondi F, Kadji FMN, Singh G, Kishi T, Uwamizu A, et al. Illuminating G-Protein-Coupling Selectivity of GPCRs. Cell. 2019;177(7):1933-47 e25. 10.1016/j.cell.2019.04.044
3. Avet C, Mancini A, Breton B, Gouill CL, Hauser AS, Normand C, et al. Effector membrane translocation biosensors reveal G protein and B-arrestin profiles of 100 therapeutically relevant GPCRs. bioRxiv. 2021:2020.04.20.052027. 10.1101/2020.04.20.052027

4. Kenakin T. A Scale of Agonism and Allosteric Modulation for Assessment of Selectivity, Bias, and Receptor Mutation. Molecular pharmacology. 2017;92(4):414-24. 10.1124/mol.117.108787

5. Pandy-Szekeres G, Esguerra M, Hauser AS, Caroli J, Munk C, Pilger S, et al. The G protein database, GproteinDb. Nucleic Acids Res. 2022;50(D1):D518-D25. 10.1093/nar/gkab852

6. Quoyer J, Janz JM, Luo J, Ren Y, Armando S, Lukashova V, et al. Pepducin targeting the C-X-C chemokine receptor type 4 acts as a biased agonist favoring activation of the inhibitory G protein. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(52):E5088-97. 10.1073/pnas.1312515110

7. Olsen RHJ, DiBerto JF, English JG, Glaudin AM, Krumm BE, Slocum ST, et al. TRUPATH, an open-source biosensor platform for interrogating the GPCR transducerome. Nat Chem Biol. 2020;16(8):841-9. 10.1038/s41589-020-0535-8

8. Maziarz M, Park JC, Leyme A, Marivin A, Garcia-Lopez A, Patel PP, et al. Revealing the Activity of Trimeric G-proteins in Live Cells with a Versatile Biosensor Design. Cell. 2020;182(3):770-85 e16. 10.1016/j.cell.2020.06.020

9. Smith JS, Pack TF, Inoue A, Lee C, Zheng K, Choi I, et al. Noncanonical scaffolding of G_{αi} and β-arrestin by G protein–coupled receptors. Science. 2021:eaay1833. 10.1126/science.aay1833

10. Pfeil EM, Brands J, Merten N, Vogtle T, Vescovo M, Rick U, et al. Heterotrimeric G Protein Subunit Galphaq Is a Master Switch for Gbetagamma-Mediated Calcium Mobilization by Gi-Coupled GPCRs. Mol Cell. 2020;80(6):940-54 e6. 10.1016/j.molcel.2020.10.027

11. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419. 10.1126/science.1260419

12. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5. 10.1038/ng.2653

13. Malfacini D, Patt J, Annala S, Harpsoe K, Eryilmaz F, Reher R, et al. Rational design of a heterotrimeric G protein alpha subunit with artificial inhibitor sensitivity. The Journal of biological chemistry. 2019;294(15):5747-58. 10.1074/jbc.RA118.007250

14. Blank JL, Ross AH, Exton JH. Purification and characterization of two G-proteins that activate the beta 1 isozyme of phosphoinositide-specific phospholipase C. Identification as members of the Gq class. The Journal of biological chemistry. 1991;266(27):18206-16.

15. Tomura H, Itoh H, Sho K, Sato K, Nagao M, Ui M, et al. Betagamma subunits of pertussis toxin-sensitive G proteins mediate A1 adenosine receptor agonist-induced activation of phospholipase C in collaboration with thyrotropin. A novel stimulatory mechanism through the cross-talk of two types of receptors. The Journal of biological chemistry. 1997;272(37):23130-7. 10.1074/jbc.272.37.23130

16. Jiang H, Kuang Y, Wu Y, Smrcka A, Simon MI, Wu D. Pertussis toxin-sensitive activation of phospholipase C by the C5a and fMet-Leu-Phe receptors. The Journal of biological chemistry. 1996;271(23):13430-4. 10.1074/jbc.271.23.13430

17. Camps M, Hou C, Sidiropoulos D, Stock JB, Jakobs KH, Gierschik P. Stimulation of phospholipase C by guanine-nucleotide-binding protein beta gamma subunits. Eur J Biochem. 1992;206(3):821-31. 10.1111/j.1432-1033.1992.tb16990.x

18. Flock T, Hauser AS, Lund N, Gloriam DE, Balaji S, Babu MM. Selectivity determinants of GPCR-G-protein binding. Nature. 2017;545(7654):317-22. 10.1038/nature22070

19. Jiang M, Bajpayee NS. Molecular mechanisms of go signaling. Neurosignals. 2009;17(1):23-41. 10.1159/000186688

20. Ho MKC, Wong YH. Gz signaling: emerging divergence from Gi signaling. Oncogene. 2001;20:1615-25.

21. Anderson A, Masuho I, Marron Fernandez de Velasco E, Nakano A, Birnbaumer L, Martemyanov KA, et al. GPCR-dependent biasing of GIRK channel signaling dynamics by RGS6 in mouse sinoatrial nodal cells. Proc Natl Acad Sci U S A. 2020;117(25):14522-31. 10.1073/pnas.2001270117

563 22. Kolb P, Kenakin TP, Alexander SPH, Bermudez M, Bohn LM, Breinholt CS, et al. Community
564 guidelines for GPCR ligand bias: IUPHAR Review XX. Br J Pharmacol, Accepted. 2021.
565 10.22541/au.163129851.17708337/v1
566 23. Trevino M, Manjarrez E. Balanced Expression of G Protein-coupled Receptor Subtypes in the
567 Mouse, Macaque, and Human Cerebral Cortex. Neuroscience. 2022. 10.1016/j.neuroscience.2022.01.028
568 24. Masuho I, Martemyanov KA, Lambert NA. Monitoring G Protein Activation in Cells with BRET.
569 Methods Mol Biol. 2015;1335:107-13. 10.1007/978-1-4939-2914-6_8
570 25. Gales C, Rebois RV, Hogue M, Trieu P, Breit A, Hebert TE, et al. Real-time monitoring of receptor
571 and G-protein interactions in living cells. Nat Methods. 2005;2(3):177-84. 10.1038/nmeth743
572 26. Gales C, Van Durm JJ, Schaak S, Pontier S, Percherancier Y, Audet M, et al. Probing the activation-
573 promoted structural rearrangements in preassembled receptor-G protein complexes. Nat Struct Mol Biol.
574 2006;13(9):778-86. 10.1038/nsmb1134
575 27. Schrage R, Schmitz AL, Gaffal E, Annala S, Kehraus S, Wenzel D, et al. The experimental power
576 of FR900359 to study Gq-regulated biological processes. Nat Commun. 2015;6:10156.
577 10.1038/ncomms10156
578 28. Breton B, Sauvageau E, Zhou J, Bonin H, Le Gouill C, Bouvier M. Multiplexing of multicolor
579 bioluminescence resonance energy transfer. Biophys J. 2010;99(12):4037-46. 10.1016/j.bpj.2010.10.025
580 29. Bunemann M, Frank M, Lohse MJ. Gi protein activation in intact cells involves subunit
581 rearrangement rather than dissociation. Proceedings of the National Academy of Sciences of the United
582 States of America. 2003;100(26):16077-82. 10.1073/pnas.2536719100
583 30. Janetopoulos C, Jin T, Devreotes P. Receptor-mediated activation of heterotrimeric G-proteins in
584 living cells. Science. 2001;291(5512):2408-11. 10.1126/science.1055835
585 31. Wright SC, Bouvier M. Illuminating the complexity of GPCR pathway selectivity – advances in
586 biosensor development. Curr Opin Struct Biol. 2021;69:142-9. https://doi.org/10.1016/j.sbi.2021.04.006
587 32. Apostolakou AE, Baltoumas FA, Stravopodis DJ, Iconomidou VA. Extended Human G-Protein
588 Coupled Receptor Network: Cell-Type-Specific Analysis of G-Protein Coupled Receptor Signaling
589 Pathways. J Proteome Res. 2020;19(1):511-24. 10.1021/acs.jproteome.9b00754
590 33. Berg KA, Clarke WP. Making Sense of Pharmacology: Inverse Agonism and Functional
591 Selectivity. Int J Neuropsychopharmacol. 2018;21(10):962-77. 10.1093/ijnp/pyy071
592 34. Civciristov S, Ellisdon AM, Suderman R, Pon CK, Evans BA, Kleifeld O, et al. Preassembled
593 GPCR signaling complexes mediate distinct cellular responses to ultralow ligand concentrations. Science
594 signaling. 2018;11(551):1-22. 10.1126/scisignal.aan1188
595 35. Kenakin T. Biased Receptor Signaling in Drug Discovery. Pharmacol Rev. 2019;71(2):267-315.
596 10.1124/pr.118.016790
597 36. Kooistra AJ, Mordalski S, Pandy-Szekeres G, Esguerra M, Mamyrbekov A, Munk C, et al.
598 GPCRdb in 2021: integrating GPCR sequence, structure and function. Nucleic Acids Res.
599 2021;49(D1):D335-D43. 10.1093/nar/gkaa1080
600

601 **Acknowledgments**

602 Nevin Lambert is acknowledged for fruitful discussions. We thank Ayumi Inoue (Tohoku University) for
603 reanalyzing the raw data of the TGF-α shedding assay to obtain SD values for the individual GPCR-G
604 protein pairs.
605

606 **Additional information**

607

608 **Funding**

609 Canada Research Chair in Signal Transduction and Molecular Pharmacology (MB); Canadian Institutes
610 of Health Research grant FDN-148431 (MB); Lundbeck Foundation grants R218-2016-1266 and R313-
611 2019-526 (DEG); Novo Nordisk Foundation grant NNF18OC0031226 (DEG); BINDS program grant

JP20am0101095 (AI); LEAP program grant JP20gm0010004 (AI); Japan Agency for Medical Research and Development (AMED) (AI); Takeda Science Foundation (AI); Uehara Memorial Foundation (AI);

Author contributions

Conceptualization: DEG; Methodology: DEG; Validation: ASH, DEG; Formal Analysis: AI, CA, ASH, DEG; Investigation: ASH, CA, DEG; Resources: CG; Data Curation: AM, CA, CN, DEG; Writing – Original Draft: DEG, MB; Writing – Review & Editing: AI, ASH, CA, DEG, MB; Visualization: ASH, DEG; Supervision: DEG, MB; Project Administration: DEG; Funding Acquisition: AI, DEG, MB.

Data and materials availability

All data are available in GproteinDb (https://gproteindb.or), GitHub (https://github.com/protwis/gpcrdb_data) and Spreadsheets S1-5. No code was developed for this manuscript which instead used the existing resources of GproteinDb (5) and GPCRdb (36) resources. All open-source code can be obtained from GitHub (https://github.com/protwis/protwis) under the permissive Apache 2.0 License (https://www.apache.org/licenses/LICENSE-2.0).

Competing interests

MB is the president of Domain Therapeutics scientific advisory board. AM and CN were employees of Domain Therapeutics North America during part or all of this research.

Figures and Tables

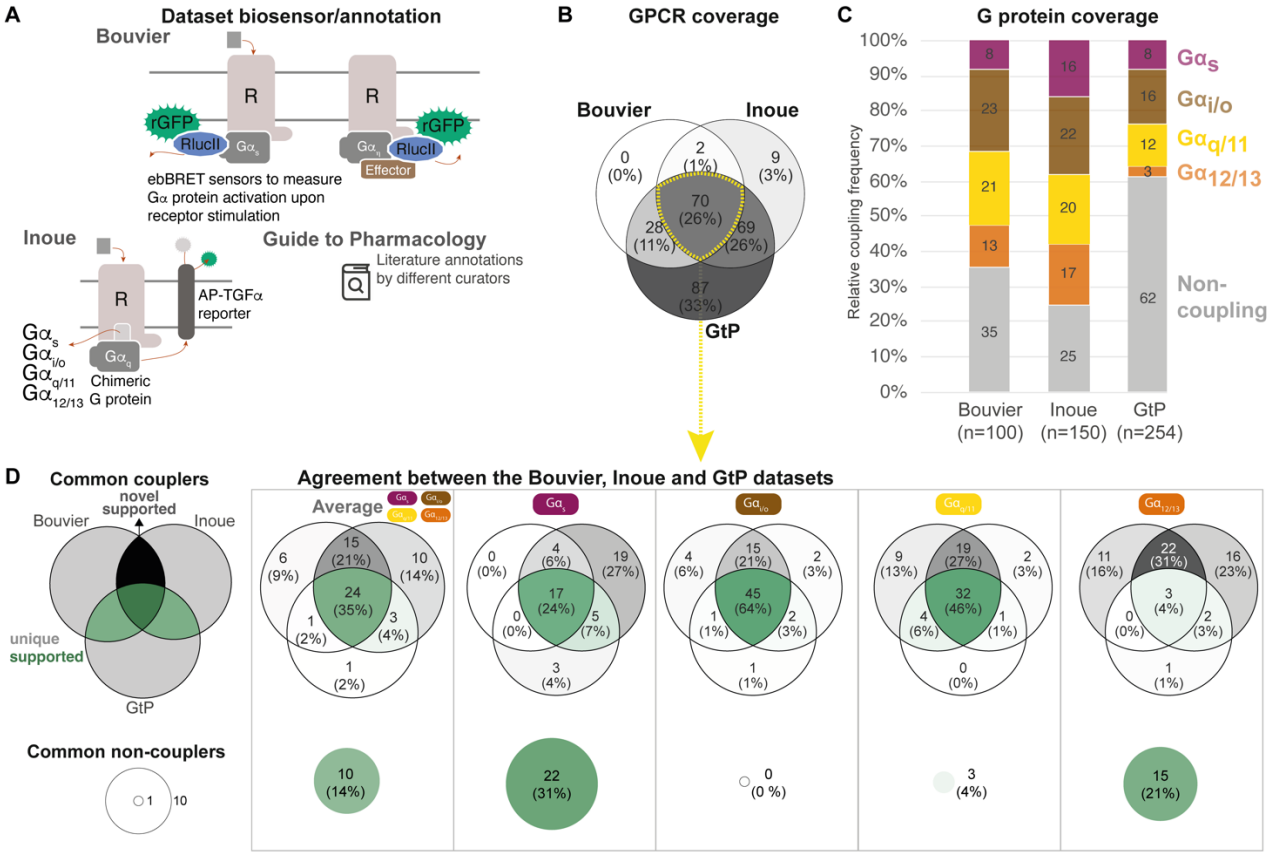
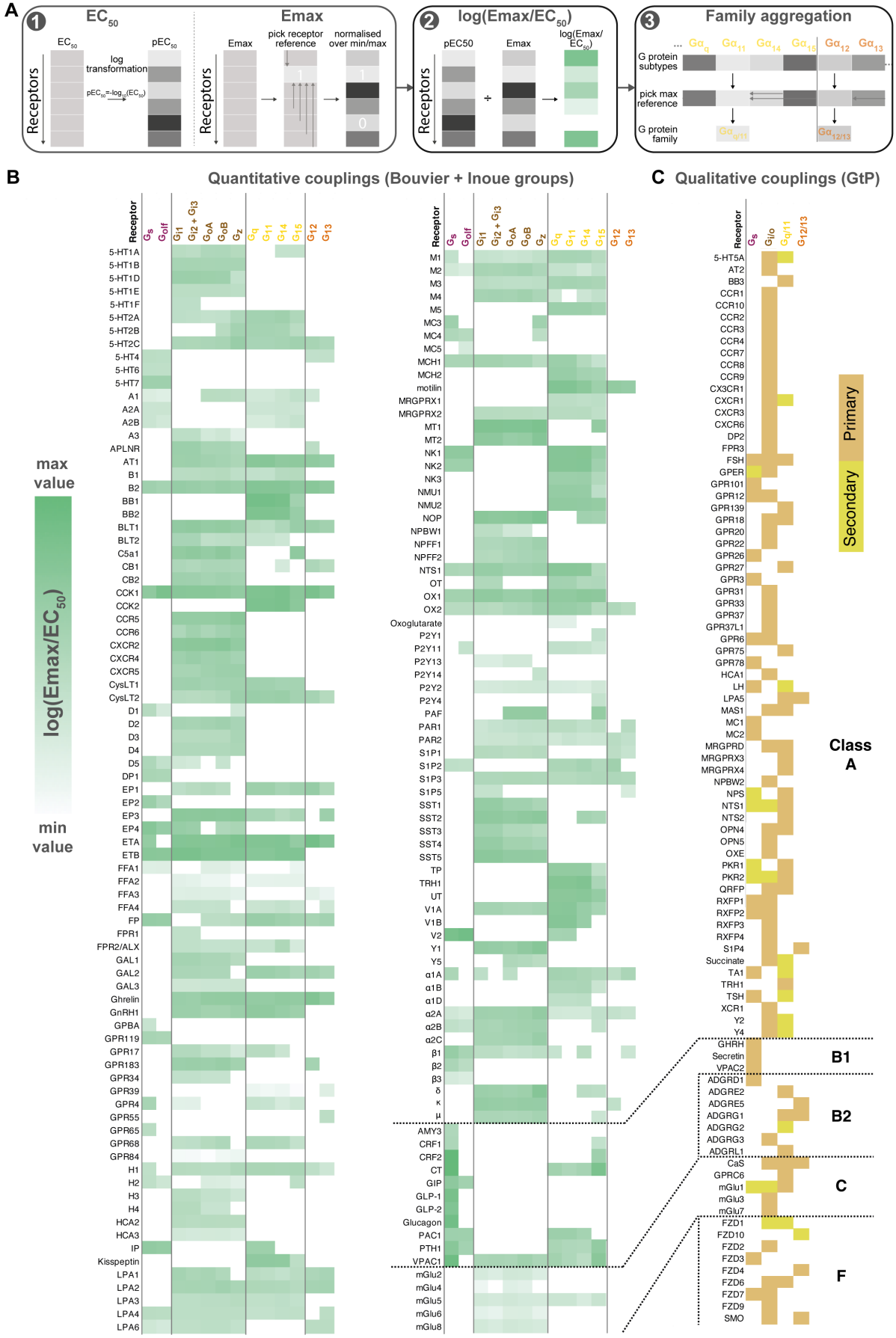


Fig. 1. Coverage and agreement of the Bouvier lab, Inoue group and Guide to Pharmacology datasets. (**A**) Biosensor principles used by the Bouvier and Inoue groups, and literature annotation stored in the Guide to Pharmacology database (GtP). (**B**) Intersection of the GPCRs included in the Bouvier (n=100), Inoue (n=150) and GtP (n=254) datasets. (**C**) Relative family distributions of G protein couplings across datasets. **(D)** Comparison of the G protein family coupling profiles of 70 GPCRs present in all of the Bouvier, Inoue and GtP datasets. More detailed analysis of common G protein couplings for just the Bouvier and Inoue datasets are given in Fig. S1. **(A-D)** Note: All analyses herein cover the 12 G proteins: G_s, G_{i1}, G_{i2}, G_{oA}, G_{oB}, G_z, G_q, G_{11}, G_{14}, G_{15}, G_{12} and G_{13}. G_{olf} and G_{i3} could not be analyzed, as they had not been tested by the Bouvier group. The Inoue data for the pairs G_{i1}-G_{i2}, G_{oA}-G_{oB} and G_q-G_{11} were generated with identical chimera inserting the Gα C-terminal hexamer into a G_q backbone (2) meaning that identical datapoints were used to assess their coverage and agreement with other datasets (Table S1).

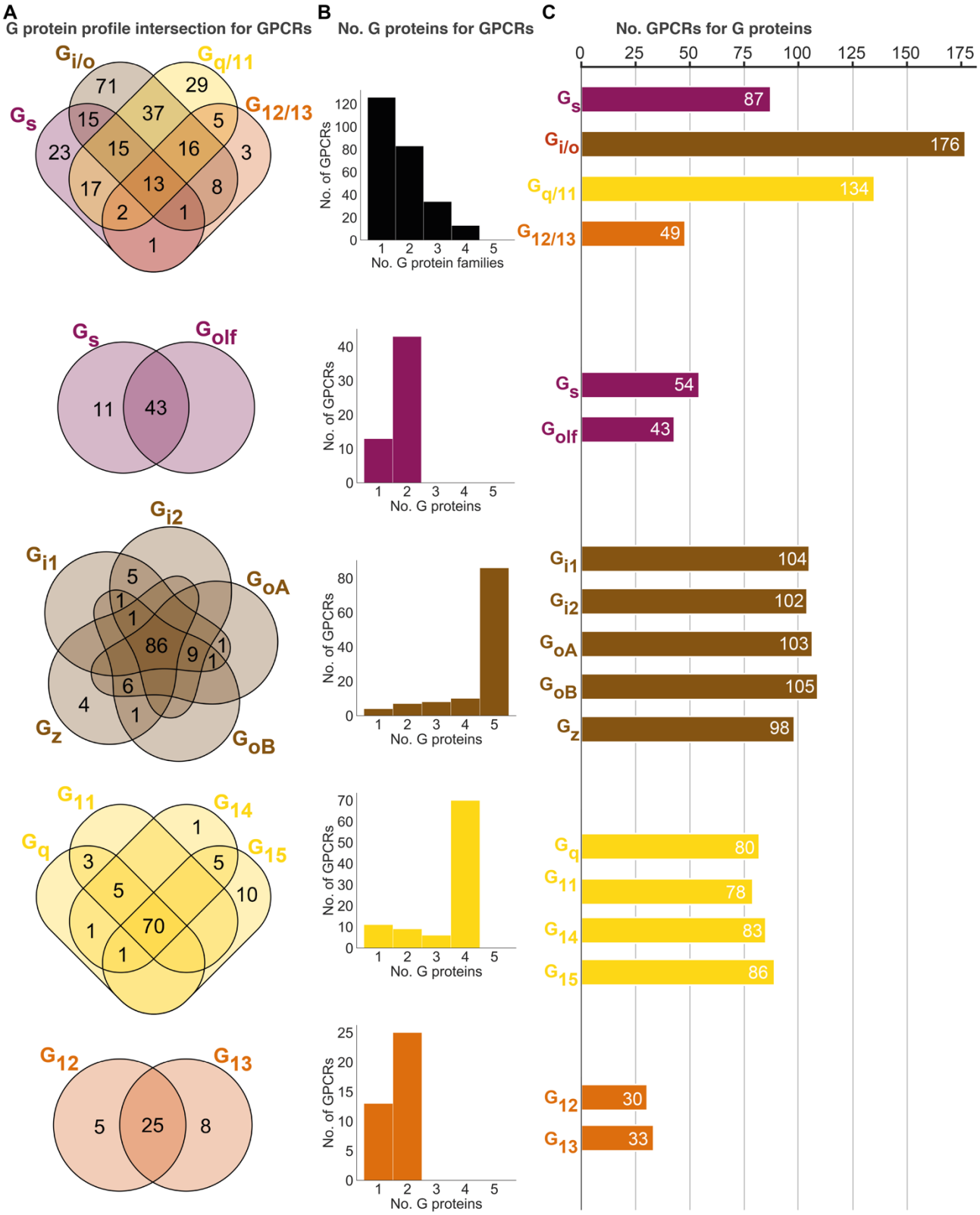

A
1 EC50
Receptors
EC50
pEC50
log transformation
pEC50=-log10(EC50)
Emax
pick receptor reference
normalised over min/max
2 log(Emax/EC50)
pEC50
Emax
log(Emax/EC50)
3 Family aggregation
G protein subtypes
pick max reference
G protein family
B
Quantitative couplings (Bouvier + Inoue groups)
C Qualitative couplings (GtP)
Receptor
max value
log(Emax/EC50)
min value
Primary
Secondary
Class A
B1
B2
C
F

646

647 **Fig. 2. Map of GPCR-G protein couplings supported by at least two studies.** (**A**) Normalization
648 approach. Emax minimum-maximum normalization, EC50 log transformation and use of log(Emax/EC50)
649 as a combined measure with member-to-family aggregation by maximum G protein value. (**B**) Heatmap
650 representation of log(Emax/EC50) values for 166 GPCRs tested by the Bouvier and/or Inoue labs (for
651 couplings with dual data sources a mean is used). For GtP, a G protein subtype is considered supported if
652 the family has a known coupling. G_{i2} and G_{i3} are represented by the same / identical chimeric G protein in
653 Inoue's dataset (Table S1). (**C**) Heatmap representation of primary and secondary transducers for 90 GPCRs
654 which couplings are only covered by the GtP database. (**B-C**) Empty cells (white) indicate no coupling. All
655 source values are available in tab 'Fig_4' in Spreadsheet S5. Note: Researchers wishing to use this coupling
656 map, optionally after applying own reliability criteria or cut-offs, can do so for any set of couplings in
657 GproteinDb (5).

	G_s	G_{i1}	G_{i2}	G_o	G_z	G_q	G_{11}	G_{14}	G_{15}	G_{12}	G_{13}	Sum	G_s	$G_{i/o}$	G_q /11	G_{12} /13	Sum
ET_A												7					3
H_1												7					3
α_{1A}												2					2
CCK_1												6					2
CB_1												3					2
α_{2A}												6					2
M_1												5					2
β_1												3					2
FP												4					2
FFA4												4					2
APLNR												1					1
BLT_1												2					1
B_2												2					1
MC_3												1					1
MC_4												1					1
$CysLT_2$												2					1
$5\text{-}HT_{2B}$												2					1
δ												2					1
NOP												2					1
D_1												1					1
OX_2												2					1
EP_1												2					1
EP_3												1					1
A_1												1					1
AT_1												2					1
α_{2B}												1					1
D_5												3					1
$S1P_1$												2					1
SST2												2					1
V_{1A}												4					1
V_2												2					1
$5\text{-}HT_{1A}$												2					1
FFA3												2					1
M_3												4					1
$5\text{-}HT_{2C}$												2					1
GPR183												1					1
H_2												2					1
M_4												3					1
Sum:	4	9	8	11	12	4	2	8	7	18	18	101	4	15	10	21	50

Fig. 3. Novel G protein couplings identified by both the Bouvier and Inoue groups. Novel couplings identified by the Bouvier and Inoue groups but not in Guide to Pharmacology. G protein families are here considered shared if at least one specific subtype is found to couple in both dataset, which is not the case for the bottommost four receptors.


A
G protein profile intersection for GPCRs
Gi/o
Gq/11
Gs
G12/13
71
29
15
37
5
23
15
16
3
17
13
8
2
1
1
Gs
Golf
11
43
Gi2
Gi1
GoA
5
1
1
86
9
1
1
6
4
1
Gz
GoB
G11
G14
Gq
G15
1
3
5
5
10
1
70
1
G12
G13
5
25
8
B
No. G proteins for GPCRs
No. of GPCRs
No. G protein families
No. G proteins
C
No. GPCRs for G proteins
0
25
50
75
100
125
150
175
Gs
87
Gi/o
176
Gq/11
134
G12/13
49
Gs
54
Golf
43
Gi1
104
Gi2
102
GoA
103
GoB
105
Gz
98
Gq
80
G11
78
G14
83
G15
86
G12
30
G13
33

664

665 **Fig. 4. Unifying the three datasets reveals a large diversity in GPCR coupling selectivity.** (**A**) GPCR-
666 G protein selectivity on the G protein family (top) and subtype levels: Venn diagrams showing the numbers
667 of shared and unique receptors. 0 values (no receptors) are omitted for clarity. (**B**) Receptor coupling
668 promiscuity: Number of receptors that couple to 1 to 4 G protein families (top) and 1 to 5 subtypes. (**C**) G
669 protein coupling promiscuity: Number of receptors that couple to each G protein family (top) or subtype.
670 (**A-C**) Panels **B-C** are based on the couplings from the Bouvier and Inoue groups that are also supported
671 by a second dataset and panel **A** additionally includes GtP couplings. This analysis of the G_s family leaves
672 out 11 receptors tested for coupling to G_s but not to G_{olf}, and G_{olf} couplings are only counted if there is a
673 supported G_s coupling. All source data are available in tab 'Fig_4' in Spreadsheet S5.

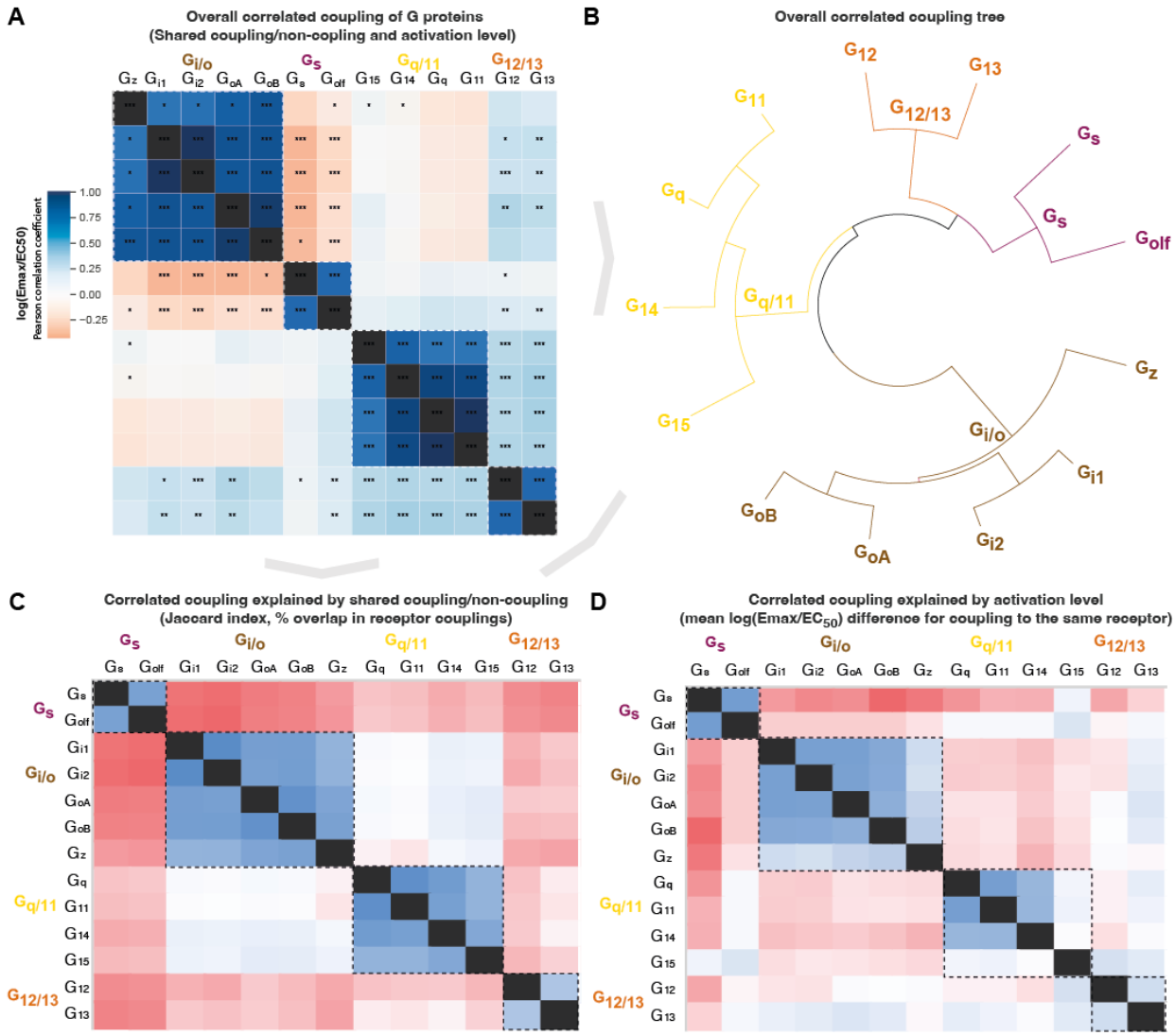


Fig. 5. Correlated coupling of G proteins based on their receptor profiles. (**A**) Overall correlated coupling of G proteins quantified by the Pearson standard correlation coefficient, which gives a measure of the strength of the linear relationship between two G proteins (the log(E_{max}/EC_{50}) value of non-coupling datapoints was set to 0). Statistically significant pairwise correlations are indicated in cells by *$p \leq 0.05$; **$p \leq 0.005$ and ***$p \leq 0.0005$. (**B**) Overall coupling correlation of G proteins shown as a tree. (**C**) Correlated coupling explained by shared coupling/non-coupling quantified as Jaccard indices (% of couplings to the same GPCRs). **(D)** Correlated coupling explained by activation level quantified as the differences in average log(E_{max}/EC_{50}) when a G protein pair couples to the same receptor. The values are averages of the log(E_{max}/EC_{50}) averages of Bouvier and Inoue values, except where data is only available in one dataset i.e., Bouvier only: G_{i1}-G_{i2}, G_{oA}-G_{oB}, G_q-G_{11}, and Inoue only: G_{olf}-all G proteins. (**A, C-D**) All G protein couplings are for class A GPCRs and supported by two datasets (Bouvier or Inoue group or GtP), and their source values are available in tab 'Fig_5' in Spreadsheet S5. (**A-D**) For G proteins (all but G_{olf}) and receptors tested by both the Bouvier and Inoue groups, an average of averages from the two groups is used. The G_{oA} and G_{oB} couplings from Bouvier were compared to the G_o values from Inoue which does not distinguish isoforms.

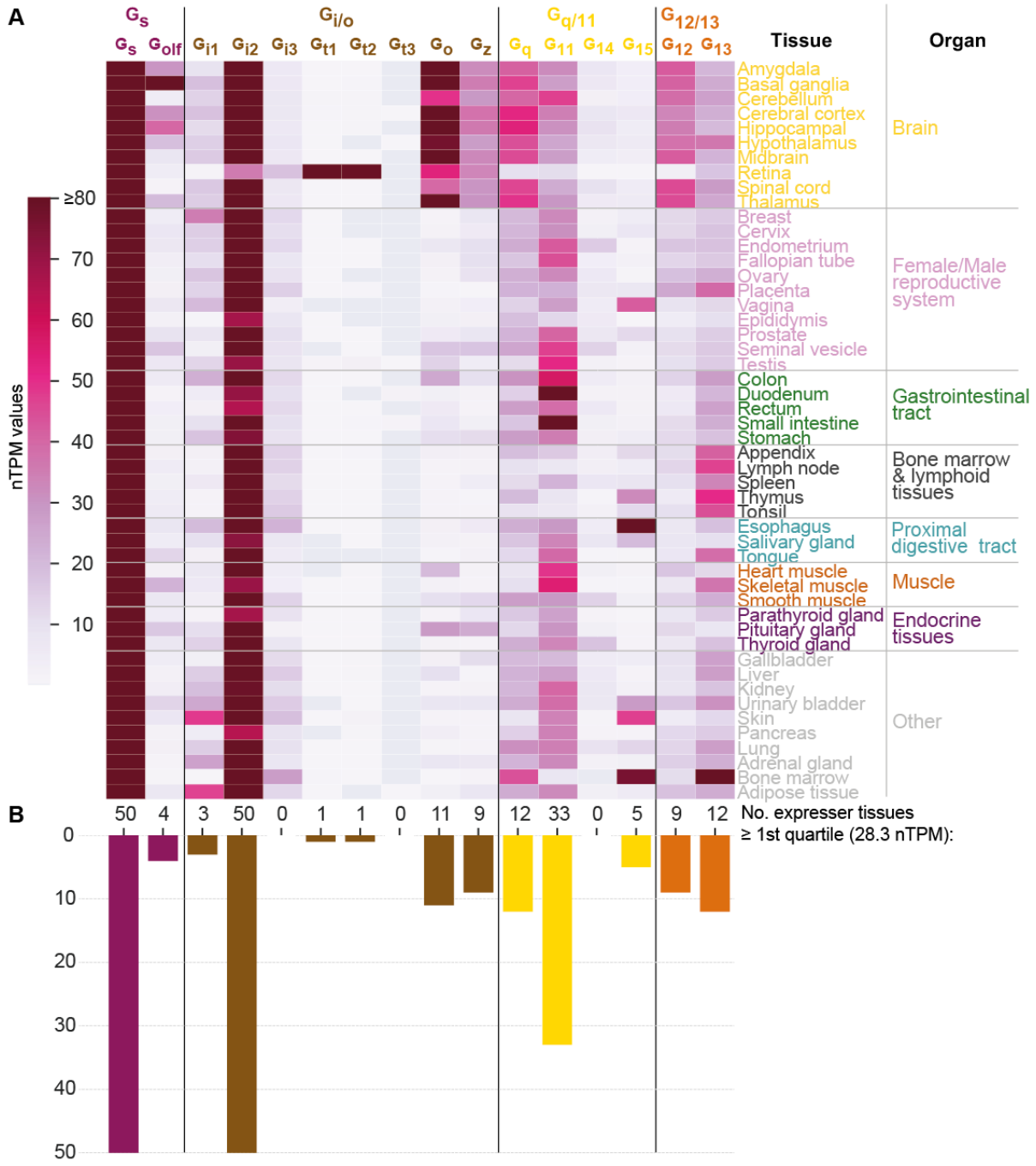


Fig. 6. G protein tissue and organ expression profiles. (**A**) Expression heatmap of all 16 human G proteins across 50 tissues and 16 organs (here grouped in 8 categories) extracted from the Human Protein Atlas (11), which also includes data from the Genotype-Tissue Expression (GTeX) project (12). The coloring denotes the normalized transcripts per million (nTPM) for each G protein and tissue capped at the median of all maximum G protein expression values (G_{t3} at 80 nTPM in retina). (**B**) Number of tissues for which the given G protein has an expression ≥ the first quartile expression threshold (28.4 nTPM) across all 16 G proteins and 50 tissues.

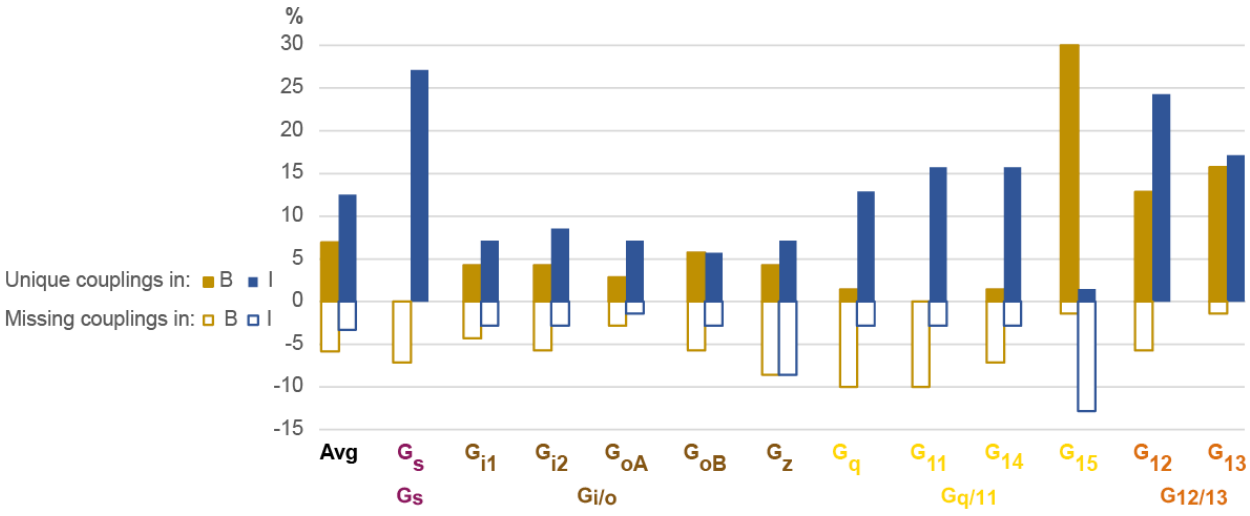


Fig. 7. Unique and missing couplings in the Bouvier and Inoue datasets. Unique and missing couplings (defined in Table 1) among all 70 receptors common to Bouvier, Inoue and GtP (the 100% includes non-coupling receptors). Missing couplings are shown as negative values. For GtP, we consider coupling to a G protein subtype possible if a coupling has been observed for the respective family. Unique couplings are hidden by default in the online G protein couplings browser in GproteinDb, as they await the independent support by a second group (5).

709

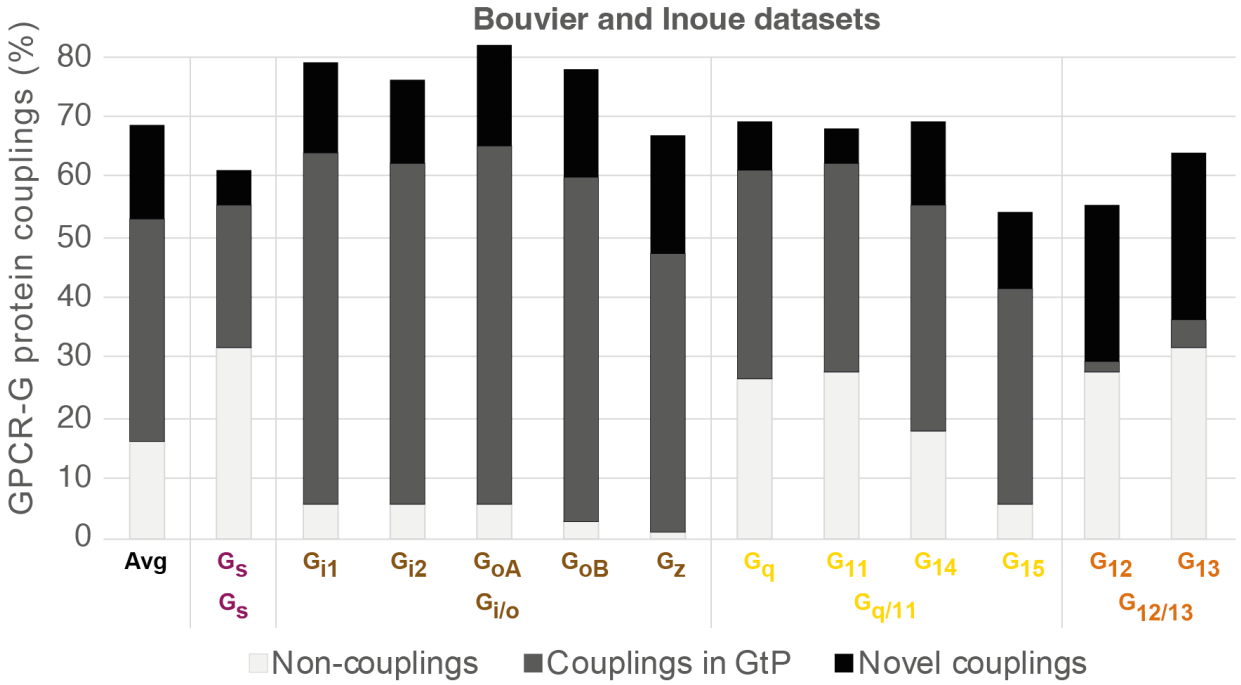
710 **Table 1: Terms and definitions used to classify GPCR-G protein couplings.**

Term	Definition
Supported	Coupling or non-coupling datapoints that is supported by at least one other dataset i.e., at least 2 in total. An exception is made for couplings reported only in the GtP database, and not yet tested in a quantitative dataset, as the couplings in GtP are in most cases supported by multiple independent publications.
Novel	Coupling that is supported by Bouvier and Inoue but not present in GtP.
Proposed	Coupling identified in one but not yet tested in a second quantitative dataset (here Bouvier or Inoue). This refers to mainly receptors, but also G protein subtypes, that have only been tested in one quantitative dataset. Couplings reported only in the GtP database are not considered "proposed" but "supported" because they in most cases are based on the annotation of multiple independent publications.
Unique	Coupling that is unique in one dataset (other datasets have no coupling).
Missing	Non-coupling in the given dataset but coupling in the other compared datasets.

711

712 **Supplementary Information**

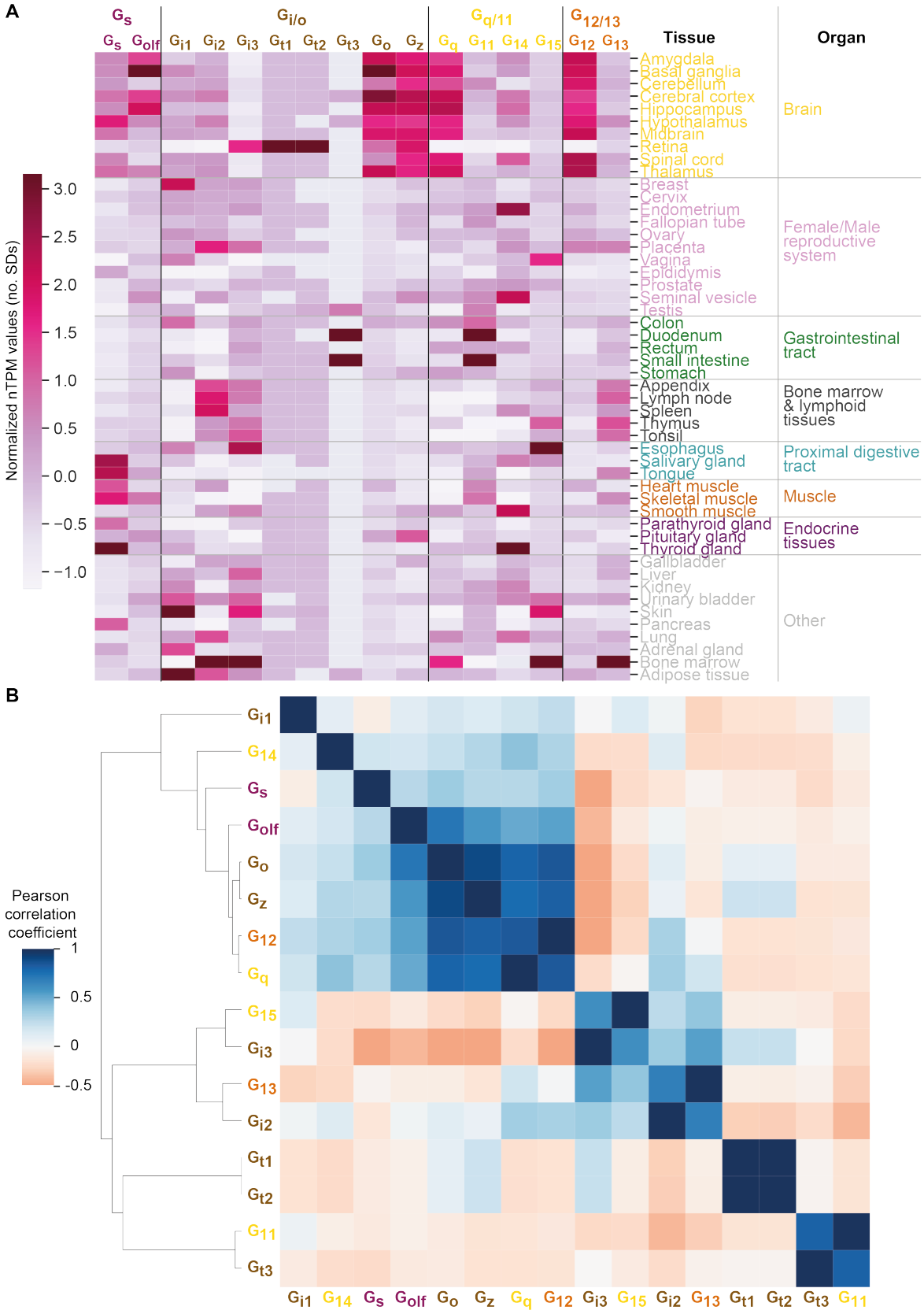
713



714
715

716 **Fig. S1. Common GPCR-G protein couplings in the Bouvier and Inoue datasets.** Black bars represent
717 couplings not in GtP and therefore considered novel (novel couplings are listed in Fig. 3). Frequencies are
718 percent couplings among all receptors common to Bouvier and Inoue (the 100% includes non-coupling
719 receptors). G_{i3} and G_{olf} (chimeras) are not included herein as they were only tested by the Inoue group.
720 Inoue used the same chimeras to represent the pairs G_{oA}-G_{oB}, G_{i1}-G_{i2} and G_q-G_{11} (Table S1).

721
722


A
Gs
Gi/o
Gq/11
G12/13
Gs Golf Gi1 Gi2 Gi3 Gt1 Gt2 Gt3 Go Gz Gq G11 G14 G15 G12 G13
Tissue
Organ
Amygdala
Basal ganglia
Cerebellum
Cerebral cortex
Hippocampus
Hypothalamus
Midbrain
Retina
Spinal cord
Thalamus
Brain
Breast
Cervix
Endometrium
Fallopian tube
Ovary
Placenta
Vagina
Epididymis
Prostate
Seminal vesicle
Testis
Female/Male reproductive system
Colon
Duodenum
Rectum
Small intestine
Stomach
Gastrointestinal tract
Appendix
Lymph node
Spleen
Thymus
Tonsil
Bone marrow & lymphoid tissues
Esophagus
Salivary gland
Tongue
Proximal digestive tract
Heart muscle
Skeletal muscle
Smooth muscle
Muscle
Parathyroid gland
Pituitary gland
Thyroid gland
Endocrine tissues
Gallbladder
Liver
Kidney
Urinary bladder
Skin
Pancreas
Lung
Adrenal gland
Bone marrow
Adipose tissue
Other
Normalized nTPM values (no. SDs)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
B
Pearson correlation coefficient
1
0.5
0
-0.5
Gi1 G14 Gs Golf Go Gz Gq G12 Gi3 G15 Gi2 G13 Gt1 Gt2 Gt3 G11

723

724 **Fig. S2. Correlated expression of G proteins.** (**A)** Expression heatmap like in Fig. 6 but with a z-
725 transformation across tissues, thereby visualizing their relative expression for each G protein. (**B)**
726 Correlated G protein expression across tissues as a function of normalized transcripts per million (nTPM),
727 which is a consensus normalized transcript expression value across the Genotype-Tissue Expression
728 (GTEx) and Human Protein Atlas datasets (see Methods). Pairwise correlation of nTPM expression values
729 by Pearson standard correlation coefficient gives positive and negative correlations between G protein pairs.
730 Expression clusters are distinct from the G protein families, which are denoted with color-coded labels.
731

732 **Table S1. G proteins tested by Bouvier et al. and Inoue et al.**

733 The two published datasets contain 11 common G proteins (G_s, G_{i1}, G_{i2}, G_o, G_z, G_q, G_{11}, G_{14}, G_{15}, G_{12} and
734 G_{13}), two Inoue et al. specific G proteins (G_{olf} and G_{i3}) and two Bouvier specific isoforms (G_{oA} and G_{oB}) of
735 the Go protein that spring from the same gene (GNAO1). Inoue et al. G_q chimera replacing the six C-
736 terminal amino acids have identical sequences for G_{i1-2}, G_{oA-B}, and G_q and G_{11}. All analyses herein used the
737 11 common G proteins, a G_o average of the G_{oA} and G_{oB} isoforms and left out G_{olf} and G_{i3} (not present in
738 Bouvier et al.) while identical chimeras from Inoue et al. were used to represent the both members of the
739 pairs G_{i1-2}, G_{oA-B} and G_q and G_{11}, respectively. Abbreviations: wt = wildtype. RlucII = Renilla luciferase 2.
740

Family	G protein (Protein)	Gene name	UniProt name	UniProt identifier	Bouvier	Inoue
G_s	G_s	GNAS	GNAS2	P63092	wt + RlucII at pos 67	G_{q1-353}-RQYELL
	G_{olf}	GNAL	GNAL	P38405	-	G_{q1-353}-KQYELL
$G_{i/o}$	G_{i1}	GNAI1	GNAI1	P63096-1	wt	G_{q1-353}- KDCGLF
	G_{i2}	GNAI2	GNAI2	P04899-1	wt	
	G_{i3}	GNAI3	GNAI3	P08754	-	G_{q1-353}-KDCGLY
	G_{oA}	GNAO1	GNAO	P09471-1	wt	G_{q1-353}-RGCGLY
	G_{oB}	GNAO1	GNAO	P09471-2	wt	
	G_z	GNAZ	GNAZ	P19086-1	wt	G_{q1-353}-KYIGLC
	G_{t1} (Transducin)	GNAT1	GNAT1	P11488	-	Identical to G_{q1-353}-G_{i1-2} chimera
	G_{t2} (Transducin)	GNAT2	GNAT2	P19087	-	
	G_{gust} (Gustducin)	GNAT3	GNAT3	A8MTJ3	-	
$G_{q/11}$	G_q	GNAQ	GNAQ	P50148-1	wt	wt
	G_{11}	GNA11	GNA11	P29992-1	wt	G_{q1-353}-KEYNLV
	G_{14}	GNA14	GNA14	O95837-1	wt	G_{q1-353}-REFNLV
	G_{15}	GNA15	GNA15	P30679-1	wt	G_{q1-353}-DEINLL
$G_{12/13}$	G_{12}	GNA12	GNA12	Q03113-1	wt	G_{q1-353}-KDIMLQ
	G_{13}	GNA13	GNA13	Q14344-1	wt	G_{q1-353}- KQLMLQ

741

742 # Titles Of Additional Supplementary Material Files
743
744 **Spreadsheet S1. SD cut-off and pEC50 and Emax distributions.**
745 This spreadsheet contains a calculation of a standard deviation (SD) cut-off that gives the best agreement
746 between G protein couplings from the Bouvier and Inoue groups. The cut-off, 1.4 represent the number of
747 standard deviations of a G protein coupling average (at least from triplicates) from the basal signal of the
748 given receptor. The spreadsheet also contains a DataStats tab detailing the distribution of raw and
749 normalized pE_{C50} and Emax values.
750
751 **Spreadsheet S2. Assessment of effect of different ligands.**
752 This spreadsheet contains a comparison of the relative average difference obtained for G protein couplings
753 when determined with and without using the same ligand in the Bouvier and Inoue groups.
754
755 **Spreadsheet S3. Qualitative coupling comparisons.**
756 This spreadsheet contains a qualitative comparison (coupling and non-coupling) of GPCR-G protein pairs
757 from the Bouvier and Inoue groups, and from the Guide to Pharmacology database.
758
759 **Spreadsheet S4. Normalization and aggregation into G protein families.**
760 This spreadsheet contains a comparative analysis of which normalization and aggregation into G protein
761 families that gives the best agreement between G protein couplings from the Bouvier and Inoue groups.
762
763 **Spreadsheet S5. Unified coupling map.**
764 This spreadsheet contains a unification of G protein couplings from the Bouvier and Inoue groups, and from
765 the Guide to Pharmacology database into a common coupling map.