GPCRs show widespread differential mRNA expression and frequent mutation and copy number variation in solid tumors

Krishna Sriram[1], Kevin Moyung[1], Ross Corriden[1], Hannah Carter[2], Paul A. Insel[1, 2 *]
Departments of Pharmacology[1] and Medicine[2]
University of California, San Diego

*: Corresponding Author

Contact Information:

Email: pinsel@ucsd.edu
Phone: 858 534 2298

Mailing Address:
9500 Gilman Drive,
BSB 3076, Mail Code 0636
La Jolla, CA, 92093

Abstract:
G protein-coupled receptors (GPCRs) are the most widely targeted gene family for FDA-approved drugs. To assess possible roles for GPCRs in cancer, we analyzed Cancer Genome Atlas data for mRNA expression, mutations, and copy number variation (CNV) in 20 categories/45 sub-types of solid tumors and quantified differential expression of GPCRs by comparing tumors against normal tissue from the GTEx database. GPCRs are over-represented among coding genes with elevated expression in solid tumors; most tumor types differentially express >50 GPCRs, including many targets for approved drugs, hitherto largely unrecognized as targets of interest in cancer. GPCR mRNA signatures characterize specific tumor types, indicate survival and correlate with expression of cancer-related pathways. Tumor GPCR mRNA signatures have prognostic relevance for survival and correlate with expression of numerous cancer-related genes and pathways. GPCR expression in tumors is largely independent of staging/grading/metastasis/driver mutations and GPCRs expressed in cancer cell lines parallels that measured in tumors. Certain GPCRs are frequently mutated and appear to be hotspots, serving as bellwethers of accumulated genomic damage. CNV of GPCRs while common, does not generally correlate with mRNA expression. We suggest a previously under-appreciated role for GPCRs in cancer, perhaps as functional oncogenes, biomarkers, surface antigens and pharmacological targets.

List of abbreviations/acronyms
CAFs: Cancer associated fibroblasts; CCLE: Cancer Cell Line Encyclopedia; CNV/CNA: Copy number variation/ amplification; DE: Differential Expression; GPCR: G protein-coupled receptor; GTEx: Gene Tissue Expression Project [1]; GtoPdb: IUPHAR/BPS Guide to Pharmacology[2]; MDS: Multi-dimensional Scaling; Nmut: Number of genes with somatic non-silent mutations per tumor genome; OE: Overexpression; TCGA: The Cancer Genome Atlas; Table 1 lists abbreviated cancer names.

Introduction

G protein-coupled receptors (GPCRs), the largest family of cell-surface receptors (>800 in the human genome), mediate the signaling of a wide variety of ligands, including hormones, neurotransmitters, proteases, lipids, and peptides. GPCRs regulate many functions (e.g., metabolism, migration, proliferation) and interactions of cells with their environment. GPCRs are also the largest family of targets for approved drugs [3, 4], interacting with ~35% of FDA approved drugs, but are infrequently targeted in tumors other than endocrine cancers, even though a role for GPCRs has been implicated in features of the malignant phenotype [5-8]. One reason for their limited use is the notion that GPCRs are rarely mutated in cancer [9] although mutations occur in heterotrimeric GTP binding (G) proteins that GPCRs activate [9] and GPCRs regulate pathways, such as Wnt, MAPK and PI3K signaling, with mutations in cancer [10]. The biological relevance of GPCRs for the malignant phenotype and their high druggability imply that GPCRs might be an under-explored class of contributors to and targets in cancer.

To define the landscape of GPCRs in cancer, we undertook an integrated analysis of Differential Expression (DE), mutations, and CNV of GPCRs in 20 types of solid tumors (**Table 1, Supplementary Tables 1, 2**). Using RNA-seq data from TCGA and the GTEx database [1], we performed DE analysis of GPCRs in tumors compared to normal tissue, an analysis facilitated by the TOIL recompute project [11]. We studied GPCRs annotated by GtoPdb [2], including endoGPCRs (which respond to endogenous agonists) and taste receptors but not olfactory GPCRs for which such annotations are unavailable (**Supplement 2)**. Our findings identify many differentially expressed GPCRs in solid tumors and corresponding cancer cell lines but a less important role for mutations and CNV. GPCRs with DE predict survival and are associated with expression of oncogenes and tumorigenic pathways. Overall, these results reveal a largely under-appreciated potential of GPCRs as contributors to cancer biology and as therapeutic targets. Our results from DE (N = 6224 individual tumors), mutation (N = 5103), and CNV analysis (N = 7545) are available as a resource at *insellab.github.io*.

FIGURE 1

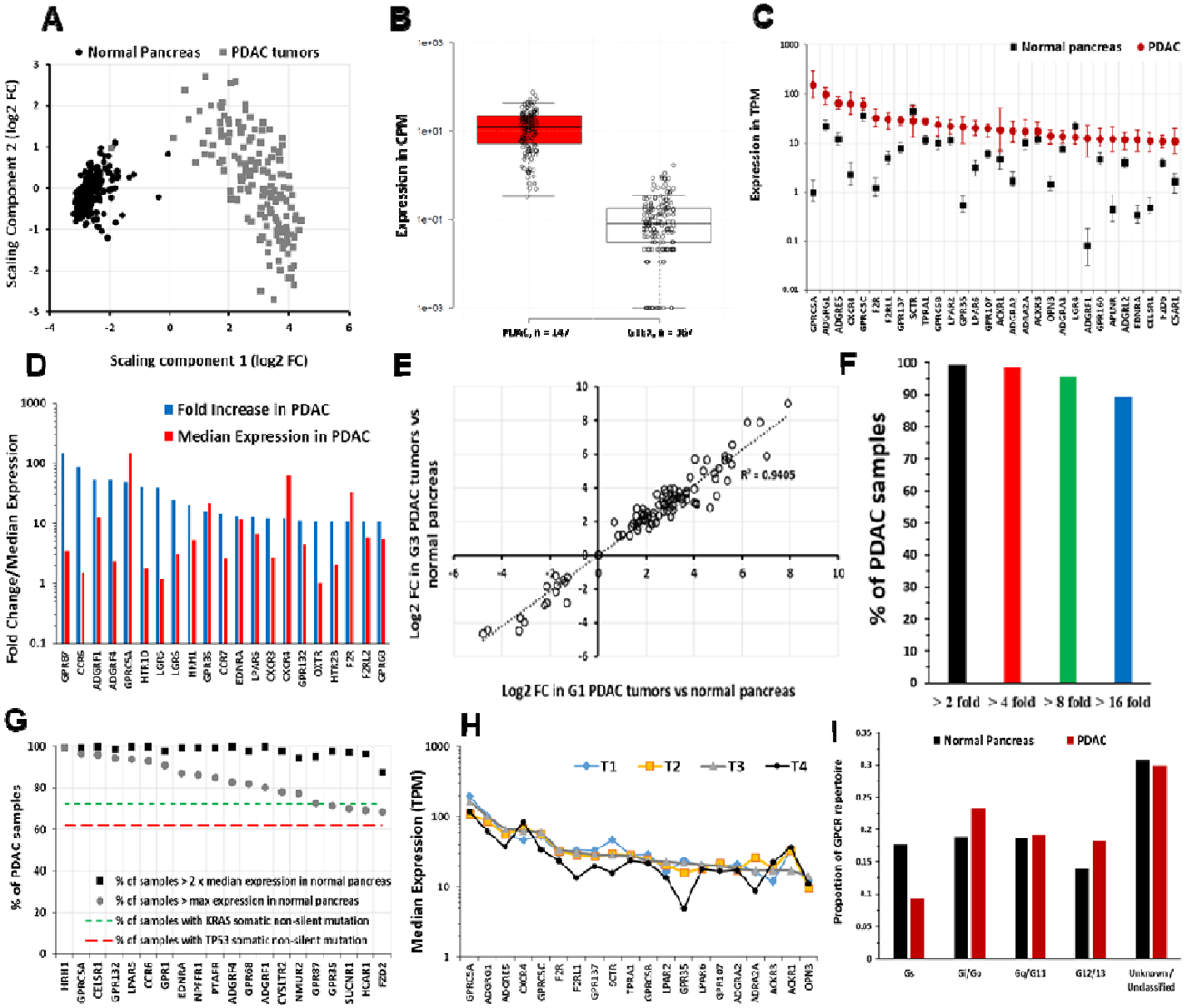

A
Normal Pancreas
PDAC tumors
Scaling Component 2 (log2 FC)
Scaling component 1 (log2 FC)
B
Expression in CPM
C
Normal pancreas
PDAC
Expression in TPM
D
Fold Increase in PDAC
Median Expression in PDAC
Fold Change/Median Expression
E
Log2 FC in G3 PDAC tumors vs normal pancreas
Log2 FC in G1 PDAC tumors vs normal pancreas
R² = 0.9405
F
% of PDAC samples
> 2 fold
> 4 fold
> 8 fold
> 16 fold
G
% of PDAC samples
% of samples > 2 x median expression in normal pancreas
% of samples > max expression in normal pancreas
% of samples with KRAS somatic non-silent mutation
% of samples with TP53 somatic non-silent mutation
H
T1
T2
T3
T4
Median Expression (TPM)
I
Normal Pancreas
PDAC
Proportion of GPCR repertoire
Gs
Gi/Go
Gq/G11
G12/13
Unknown/ Unclassified

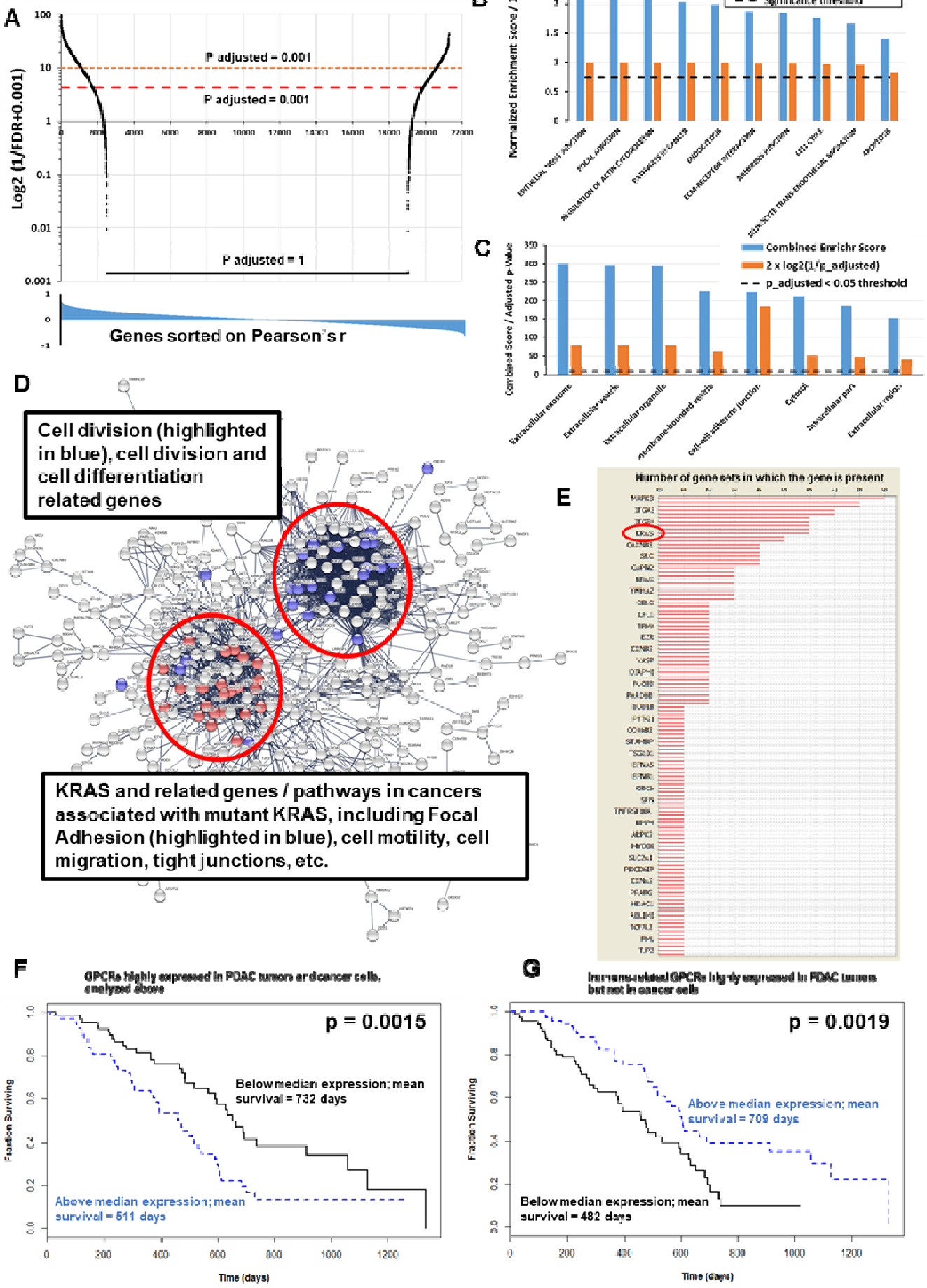
Figure 1. Differential Expression (DE) analysis of gene and GPCR expression in solid tumors compared to normal tissue: PDAC tumors and normal pancreas as an example

(**A**) MDS plot, showing clustering of PDAC tumor (TCGA) and normal pancreas (GTEx) samples.
(**B**) Normalized expression in CPM of *GPRC5A* in PDAC and normal pancreas.
(**C**) The 30 highest expressed GPCRs in PDAC and their corresponding expression in normal pancreatic tissue.
(**D**) The 21 GPCRs overexpressed >10-fold in PDAC compared to normal pancreas.
(**E**) Grade 1 (G1) and Grade 3 (G3) PDAC tumors have similar GPCR expression changes compared to normal pancreas.
(**F**) Increased expression of *GPRC5A*, the highest expressed GPCR in PDAC, in nearly all PDAC samples compared to normal tissue.
(**G**) Frequency of 2-fold increase and percent of TCGA-PDAC samples with higher maximal expression compared to normal pancreas of the indicated GPCRs with comparison of the frequency of mutations of *KRAS* and *TP53*, the most frequent somatic, non-silent mutations in PDAC tumors in TCGA.
(**H**) Expression is similar for different pathological stages/types (T) of the highest expressed GPCRs in PDAC.
(**I**) Changes in GPCR expression in PDAC alter the GPCR repertoire that couple to different G proteins.

Figure 2. GPCRs expression correlates with cancer-related pathways and predicts survival in PDAC

(**A**) Distribution of Pearson's r (blue) for expression of all genes in PDAC correlated with mRNA expression of a subset of 5 overexpressed GPCRs (ADGRF1, ADGRF4, GPRC5A, HRH1 and LPAR5), with corresponding Bonferroni adjusted p-values calculated on p-values associated with each gene's Pearson r value.
(**B**) KEGG [12] gene sets with positive enrichment among genes most significantly positively or negatively correlated with expression of the identified subset of overexpressed GPCRs, based on GSEA pre-ranked analysis [13].
(**C**) Enrichr [14] analysis of the genes positively correlated in (**A**), with FDR < 0.001, based on their enrichment in cellular compartments, from the Jensen compartment database [15].
(**D**) Network analysis via STRING [16] of the genes positively correlated in (**A**), with FDR < 0.001.
(**E**) Leading-edge analysis of GSEA results from (**B**), showing genes which are common to multiple enriched KEGG genes sets.
(**F, G**) Kaplan-Meier curves for the impact of combined, normalized expression of subsets of GPCRs on PDAC patients. Total number of patients = 141; 59 patients were censored due to inadequate follow-up. p values were calculated using the Peto & Peto modification of the Gehan-Wilcoxon test. (**F**) Impact of highly expressed GPCRs on median survival: 652 days (if below median expression) and 470 days (if above median expression) and for (**G**) Impact of highly expressed immune-related GPCRs on median survival: 460 days (below median expression) and 603 days (above median expression).

FIGURE 2


A
P adjusted = 0.001
P adjusted = 0.001
P adjusted = 1
Log2 (1/FDR+0.001)
Genes sorted on Pearson's r
B
Normalized Enrichment Score
1 - FDR
Significance threshold
Normalized Enrichment Score / 1 - FDR
EPITHELIAL TIGHT JUNCTION
FOCAL ADHESION
REGULATION OF ACTIN CYTOSKELETON
PATHWAYS IN CANCER
ENDOCYTOSIS
ECM-RECEPTOR INTERACTION
ADHERENS JUNCTION
CELL CYCLE
LEUKOCYTE TRANSENDOTHELIAL MIGRATION
APOPTOSIS
C
Combined Enrichr Score
2 x log2(1/p_adjusted)
p_adjusted < 0.05 threshold
Combined Score / Adjusted p-Value
Extracellular exosome
Extracellular vesicle
Extracellular organelle
Membrane-bounded vesicle
Cell-cell adherens junction
Cytosol
Intracellular part
Extracellular region
D
Cell division (highlighted in blue), cell division and cell differentiation related genes
KRAS and related genes / pathways in cancers associated with mutant KRAS, including Focal Adhesion (highlighted in blue), cell motility, cell migration, tight junctions, etc.
E
Number of gene sets in which the gene is present
MAPK3
ITGA3
ITGB4
KRAS
CACNB3
SRC
CAPN2
RRAS
YWHAZ
CBLC
CFL1
TPM4
EZR
CCNB2
VASP
DIAPH1
PLCB3
PARD6B
BUB1B
PTTG1
COX6B2
STAMBP
TSG101
EFNA5
EFNB1
ORC6
SFN
TNFRSF10A
BMP4
ARPC2
MYD88
SLC2A1
PDCD4P
CCNA2
PPARG
HDAC1
ABLIM3
TCF7L2
PML
TJP2
F
GPCRs highly expressed in PDAC tumors and cancer cells, analyzed above
p = 0.0015
Below median expression; mean survival = 732 days
Above median expression; mean survival = 511 days
Fraction Surviving
Time (days)
G
Immune-related GPCRs highly expressed in PDAC tumors but not in cancer cells
p = 0.0019
Above median expression; mean survival = 709 days
Below median expression; mean survival = 482 days
Fraction Surviving
Time (days)

Results

Differential expression (DE) of GPCRs in solid tumors compared to normal tissues

We focused on GPCRs with both substantial DE and magnitude of expression in solid tumors, i.e., : 1) > 2-fold increase/decrease in DE in tumors compared to normal tissue, 2) FDR < 0.05 and 3) median expression in tumors > 1 TPM. We used the latter threshold for median expression in order to identify GPCRs that may be useful as therapeutic targets, for which higher expression is preferable. For DE analysis, we divided the 20 TCGA tumor types into 45 tumor subtypes (**Table 1**), based on histological classification of tumors in TCGA metadata. We found that different tumor subtypes within the same TCGA tumor classification have distinct GPCR expression, e.g., subtypes of breast cancer (BRCA), thyroid cancer (THCA) and esophageal cancer (ESCA) (**Figs S1A-F**).

Figure 1A shows DE for PDAC tumors (as an example) compared to normal pancreas. An MDS plot (**Fig 1A**) reveals clusters for tumors and normal tissue, implying distinct transcriptomic profiles. The more diffuse cluster of PDAC samples likely reflects their heterogeneity. Smear and volcano plots (**Figs S2A, B**) reveal many genes (>5000) with high, statistically significant DE (FDR <<0.05). **Figures S2C-E** show examples of genes (other than GPCRs) with high overexpression that prior studies implicated as having a role in PDAC. Multiple other tumor types also show expression of genes relevant to the malignant phenotype and cluster separately from their respective normal tissues, and have a large number of genes with DE, thus supporting the validity of our analysis.

Many GPCRs show DE in tumors, including those from each GPCR class: A (rhodopsin-like), B (secretin-like), frizzled and adhesion GPCRs. The highest expressed GPCRs in PDAC tumors (as an example, this is generalizable to other tumor types) are generally overexpressed compared to normal tissue and include orphan receptors (e.g., *GPRC5A* and *ADGRF4*/*GPR115*) and GPCRs with known agonists (e.g., *GPR68*) (**Figs 1B, C, D**). *GPRC5A*, the most highly expressed GPCR in PDAC, is 50-fold higher expressed; 95% of PDAC samples have >8-fold higher median *GPRC5A* expression than in normal pancreas (**Fig 1E**). Within a tumor type, a large majority of individual tumors express such overexpressed GPCRs at far higher levels than corresponding normal tissue (**Fig 1B**, e.g., GPRC5A); a subset of GPCRs are expressed in >90% of PDAC tumors at abundances greater than in any normal pancreas sample (**Fig 1F**).

Similar to the example of PDAC above, numerous GPCRs are highly, consistently overexpressed in other tumor types. For instance, in SKCM; increased expression of *GPR56*, *GPR143* and *EDNRB*, the most highly overexpressed GPCRs compared to normal skin, occurs in >90% of melanoma samples (**Figs S2F-H**). Such highly overexpressed GPCRs are expressed in the vast majority (typically >90%) of samples within a tumor subtype. Of note, overexpression of certain GPCRs tends to be more prevalent within specific tumor types/subtypes than are common mutations. For example, *KRAS* and *TP53* are the most frequently mutated genes in PDAC (>70% and >60%, of samples respectively) but increased expression of multiple GPCRs occurs with greater frequency in these tumors (**Fig 1F**) [17].

As a protein-coding family of genes, GPCRs are disproportionately enriched among overexpressed genes in solid tumors, when compared to all protein-coding genes. Evidence for this was obtained as follows: In each tumor type indicated (**Fig 5H-I**), the ratio of number of coding genes with increased expression (above a prescribed threshold) over the total number of differentially expressed genes present was computed for a) GPCRs only; b) all coding genes. Fischer's exact test was used to verify if overrepresentation of GPCRs among genes with increased expression is significant ($p < 0.05$). Data shown are for coding genes with >4 fold increased expression in solid tumors, i.e., highlighting genes with drastic increases in expression. We found that many tumor types have 2-fold or greater overrepresentation of GPCRs among coding genes with large increases in expression.

GPCR expression and DE in tumors compared to normal tissues

We compiled a list of GPCRs overexpressed in solid tumors with fold-changes and FDR along with expression in TPM (for median expression and within-group comparisons of different genes) and CPM (for inter-group comparisons of the same gene). The analysis revealed that 35 of 45 tumor types/subtypes show increased expression of >30 GPCRs; 203 GPCRs are overexpressed in at least one type of cancer (**Supplementary Table 4; Supplement 3**), including 47 orphan GPCRs and >15 GPCRs that couple to each of the major G protein classes. Increased expression of 130 GPCRs occurs in ≥4 tumor subtypes (**Supplement 3**). A subset of GPCRs is overexpressed in many tumors, e.g., *FPR3* in 38 of the 45 tumor categories. **Supplementary Table 4** lists other examples along with GPCRs that have reduced expression compared to normal tissue. Importantly, of the 203 GPCRs with increased expression in one or more tumors, 77 are targets for approved drugs. These include ADORA2B, *CCR5* and *F2R*, which are overexpressed in 27, 27 and 26 tumor subtypes, respectively. **Table S4** and **Supplement 3** provide further details regarding such druggable GPCRs.

GPCR expression is associated with cancer-related pathways and with survival: PDAC as an example

A subset of GPCRs in PDAC is highly overexpressed and prominently expressed in the tumors and in PDAC cells (*vide infra*). Combining expression (normalized to median expression in PDAC) of 5 of the most highly DE GPCRs (ADGRF1, ADGRF4, GPRC5A, HRH1 and LPAR5) yields a composite 'marker' whose expression positively correlates with a subset of ~1200 genes with high statistical significance (Bonferroni adjusted p-values < 0.001) (**Fig 2A**).

We conducted further analyses related to PDAC, including with GSEA [13] of the sets of negatively and positively associated genes, pre-ranked/weighted by their FDRs and found an enrichment of a number of KEGG [12] pathways relevant to cancer (**Fig 2B**). The set of positively associated genes shows similar associations with cancer-related pathways when analyzed via GO [18, 19] and Enrichr [14]. Enrichr also identifies, based on the Jensen compartment database [15], an enrichment of vesicle and exosome-related gene products among the set of positively correlated genes (**Fig 2C**). Network-analysis of the genes positively associated with the composite GPCR marker via STRING [16] provides an intuitive picture of this gene set (**Fig 2D**). Two 'clusters' of genes and pathways are evident; those associated with KRAS (including KRAS itself) and related processes (e.g., focal adhesion pathways) and a second cluster associated with regulation of cell cycle, cell division and differentiation. Expression of highly overexpressed GPCRs is positively correlated with one another and with expression of KRAS, implicating this GPCR subset as a PDAC signature. Leading edge analysis of the GSEA results confirmed that KRAS and other oncogenes are common elements in multiple enriched gene sets associated with this GPCR signature (**Fig 2E**). Survival analysis indicates that GPCR expression has prognostic relevance: patients with greater than the median expression of the five GPCRs had a ~200 day shorter survival compared to those with less than the median expression (**Fig 2F; Fig S3A)**.

We identified a subset of highly overexpressed chemokine receptors (CCR6, CCR7, CXCR3 and CXCR4) not expressed in PDAC cancer cells but likely associated with immune cell activation. Genes that correlated with expression of these GPCRs are involved with immune-associated., especially T-cell and B-cell related, pathways (**Fig S3B**). Combined expression of these GPCRs is a positive predictor of survival (**Fig 2G**). The observation that GPCR expression may be a marker for survival is not unique to PDAC (**Fig 2F, G; Fig 5 E-G**). For the examples shown, individual GPCRs have an association with survival but combinations of such GPCRs are even better predictors of survival. GPCR expression may thus serve as a prognostic indicator in multiple tumor types.

The finding that GPCRs with high expression and DE in tumors show an association with tumorigenic pathways appears generalizable. For example, expression of GPCRs highly expressed/overexpressed in adenocarcinomas (**Fig 5D**) is positively correlated with expression of genes from pathways similar to ones shown in **Figure 2D** for PDAC; i.e., focal adhesion and cell motility as well as pathways related to cell cycle and division. By contrast, GPCR expression of solid tumor types from different branches of the cancer/GPCR phylogenetic tree (**Figure 5D)**

shows an association with different pathways from those observed in adenocarcinomas. For example, GPR143, EDNRB and ADGRG1 are highly expressed and differentially expressed in SKCM and are adverse indicators of survival (**Figs S2F-H and Fig 5E**). Pathways enriched among genes that correlated with expression of the GPCRs are ones implicated in metastatic SKCM (**Figs S4 A-C**), such as transferrin transport [20], melanosome organization [21] and insulin receptor signaling [22]. In general, highly expressed GPCRs in solid tumors show a positive correlation between GPCR expression and expression of tumorigenic pathways, implicating these GPCRs as potentially functional oncogenes.

Impact of mutations and stage/grade of tumors on GPCR expression?

GPCR expression and DE is largely independent of tumor stage and grade. **Figure 1G** shows the similarity in GPCR DE (compared to normal tissue) for Grades 1 and 3 PDAC tumors. Median expression of GPCRs was also similar in PDAC tumors with different pathological T (**Fig 1H**). Similarly, Stage I and Stage IIIA BRCA IDC HR+ (Hormone Receptor positive) tumors have comparable GPCR expression and DE (**Fig 3D**).

GPCR expression appears largely independent of driver mutations, such as BRCA HR+ IDC tumors with either *PI3KA* or *P53* mutations (**Fig 3A-C**); both groups have similar GPCR expression and compared to normal breast tissue, DE of the same GPCRs. Similar results occur for LUAD and STAD that have or lack P53 mutations. Mechanisms that lead to increased GPCR expression in solid tumors may thus not depend on specific driver mutations. Presence of highly overexpressed GPCRs may be a more ubiquitous feature of tumors than the presence of specific driver mutations, as exemplified by PDAC (**Fig 1F**) and generalizable to most other other tumor types where numerous GPCRs show DE (**Table 1**)**.**

GPCR expression is similar in metastatic and primary tumors

Among TCGA data, the SKCM gene expression dataset has the most replicates of metastases. Primary and metastatic SKCM show similar expression and DE of GPCRs (e.g., *GPR143, EDNRB* and other highly expressed GPCRs) even though major differences occur in overall gene expression between primary and metastatic SKCM (**Fig 3E, F; S2F-H)**. We found similar results for GPCR expression with primary and metastatic BRCA and THCA tumors and for recurrent and primary ovarian tumors (**Figs S5A-F**), though the number of replicates for each is small (<10).

GPCRs highly expressed in tumors are highly expressed in cancer cells

We assessed RNA-seq data for GPCR expression in cancer cell lines from the EBI portal generated via the iRAP analysis pipeline [23] for cell lines in CCLE [24] and from Genentech [25]. The use of a different analysis pipeline than that used for TCGA data does not allow for direct statistical comparisons of the datasets but semi-empirically confirms that the majority of GPCRs in TCGA tumors are present in cancer cells (and vice versa). We also mined RNA-seq for primary melanoma cells [26] and PDAC cells [27] from the NCBI GEO database. The data from these sources (**Methods, Section 1**) allow an approximate comparison with data for tumors. **Supplement 3** shows GPCR expression in cancer cell lines.

As an example, GPCRs with highest median expression in TCGA PDAC tumors are generally highly expressed in PAAD cell lines and patient-derived PDAC cells [27] (**Fig 3G**). A few exceptions exist, perhaps from effects of cell culture or expression by non-cancer cells in tumors. Even so, highly expressed GPCRs in PAAD cells are highly expressed in PDAC tumors (**Fig 3H**), findings that also occur in other tumors, such as SKCM (**Fig 3I, J**). Thus, most highly expressed GPCRs in tumors are also highly expressed in cancer cells and vice versa.

Most overexpressed GPCRs are rarely mutated

The most frequently mutated GPCRs in solid tumors are rarely overexpressed (**Figs 4A-B, 6H**) and conversely, highly overexpressed GPCRs in solid tumors are rarely mutated (**Fig 4A-B, Supplementary Table 5**). In SKCM,

which has the highest mutation burden among TCGA tumor types, the most highly overexpressed GPCRs (*GPR143*, *EDNRB* and *GPR56*) are mutated in <2% of SKCM tumors whereas frequently mutated GPCRs (e.g., *GPR98,* mutated in nearly 40% of tumors) typically have low expression. The most frequently overexpressed GPCRs across all tumors (e.g., *FPR3*; **Table S5**) are mutated in <1% of all tumors surveyed, compared to frequently mutated GPCRs, e.g., *GPR98, GPR112*, which are mutated in >5% of all TCGA tumors surveyed. Thus, the frequency of GPCR mutations and likelihood of overexpression do not correlate (**Fig 4B**). The majority of GPCRs overexpressed in >20 tumor types/subtypes are mutated in < 50 samples out of > 5000 TCGA samples surveyed. Further, as discussed in following sections on GPCR mutation, mutations to these GPCRs are predicted to have no functional impact and are not enriched significantly for mutations at specific sites; thus, overexpressed GPCRs in tumors are not expected to be altered in their function by mutations.

FIGURE 3

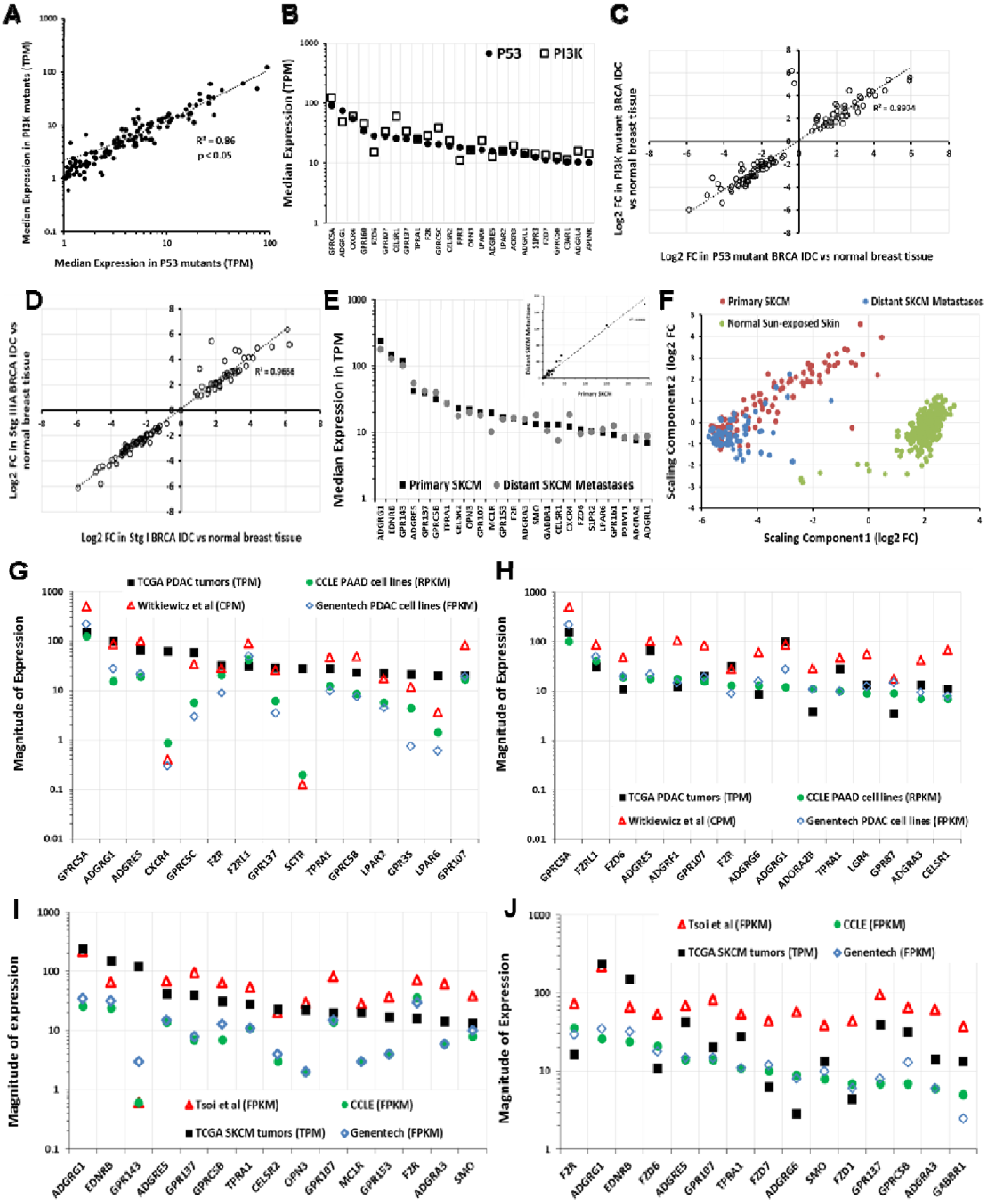

A
Median Expression in PI3K mutants (TPM)
Median Expression in P53 mutants (TPM)
R² = 0.86
p < 0.05
B
Median Expression (TPM)
P53
PI3K
C
Log2 FC in PI3K mutant BRCA IDC vs normal breast tissue
Log2 FC in P53 mutant BRCA IDC vs normal breast tissue
R² = 0.8934
D
Log2 FC in Stg IIIA BRCA IDC vs normal breast tissue
Log2 FC in Stg I BRCA IDC vs normal breast tissue
R² = 0.9666
E
Median Expression in TPM
Primary SKCM
Distant SKCM Metastases
F
Primary SKCM
Distant SKCM Metastases
Normal Sun-exposed Skin
Scaling Component 2 (log2 FC)
Scaling Component 1 (log2 FC)
G
TCGA PDAC tumors (TPM)
CCLE PAAD cell lines (RPKM)
Witkiewicz et al (CPM)
Genentech PDAC cell lines (FPKM)
Magnitude of Expression
GPRC5A
ADGRG1
ADGRE5
CXCR4
GPRC5C
F2R
F2RL1
GPR137
SCTR
TPRA1
GPRC5B
LPAR2
GPR35
LPAR6
GPR107
H
Magnitude of Expression
TCGA PDAC tumors (TPM)
CCLE PAAD cell lines (RPKM)
Witkiewicz et al (CPM)
Genentech PDAC cell lines (FPKM)
GPRC5A
F2RL1
FZD6
ADGRE5
ADGRF1
GPR107
F2R
ADGRG6
ADGRG1
ADORA2B
TPRA1
LGR4
GPR87
ADGRA3
CELSR1
I
Magnitude of expression
Tsoi et al (FPKM)
CCLE (FPKM)
TCGA SKCM tumors (TPM)
Genentech (FPKM)
ADGRG1
EDNRB
GPR143
ADGRE5
GPR137
GPRC5B
TPRA1
CELSR2
OPN3
GPR107
MC1R
GPR153
F2R
ADGRA3
SMO
J
Magnitude of Expression
Tsoi et al (FPKM)
CCLE (FPKM)
TCGA SKCM tumors (TPM)
Genentech (FPKM)
F2R
ADGRG1
EDNRB
FZD6
ADGRE5
GPR107
TPRA1
FZD7
ADGRG6
SMO
FZD1
GPR137
GPRC5B
ADGRA3
GABBR1

Figure 3. GPCR expression and presence of driver mutations and the similarity in GPCR expression of primary tumors, metastases and cancer cells derived from the tumors.

(**A**) Correlation of median expression of GPCRs in P53 mutant and PI3K mutant HR-positive BRCA IDC tumors

(**B**) Median expression of the 25 highest expressed GPCRs in P53-mutated tumors compared to expression in PI3K mutant HR-positive BRCA IDC tumors

(**C**) Fold-changes of GPCRs in P53 mutant and PI3K mutation HR-positive BRCA IDC tumors compared to normal breast tissue.

(**D**) Fold-changes of GPCRs in Stage 1 and Stage 3 ER-positive BRCA IDC tumors over normal breast tissue.

(**E**) Expression of the 25 highest expressed GPCRs and (inset) correlation of median GPCR expression between primary and metastatic SKCM.

(**F**) Gene expression of primary and distant metastatic SKCM tumors cluster differently and have large numbers of DE genes (**Supplement 3**)

(**G**) Median expression of highest expressed GPCRs in PDAC tumors compared to cancer cells (CCLE, n = 33; Genentech, n = 16; Witkiewicz et al., 2016, n = 72).

(**H**) Median expression of highest expressed GPCRs in CCLE PDAC cell lines compared to other cell lines (as in **Fig 3G**), primary cells and PDAC tumors.

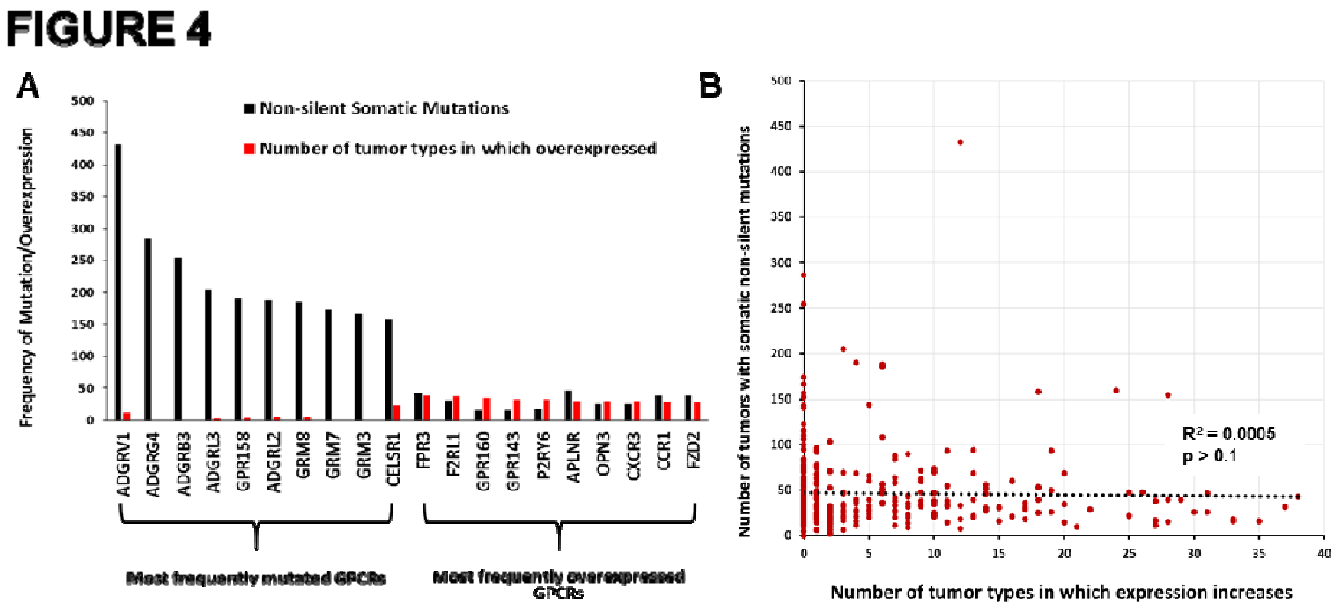
(**I**) Median expression of highest expressed GPCRs in SKCM tumors compared to cancer cells (CCLE, n = 45; Genentech, n = 44; Müller et al., 2014, n = 29).

(**J**) Median expression of highest expressed GPCRs in CCLE SKCM cell lines compared to other cell lines (as in **Fig 6I**), primary cells and SKCM tumors.

Figure 4. Frequent overexpression of GPCRs in tumors does not correlate with frequency of mutation

(**A**). The number of tumors with GPCR mutations and the number of tumor types/subtypes in which the same GPCR is overexpressed for the 10 most frequently mutated GPCRs in TCGA tumors surveyed and the 10 most frequently overexpressed GPCRs.

(**B**). The frequency of overexpression compared to the frequency of mutations.



GPCR expression in normal tissues and solid tumors

In this resource, we include expression data for all annotated non-olfactory GPCRs in solid tumors and corresponding normal tissue. **Supplement 2** shows GPCRs annotated in the IUPHAR / BPS Guide to Pharmacology GPCR database [2], their signal transduction via G protein heterotrimers and if they are orphan GPCRs. Tissues and tumors typically express >150 GPCRs (at detection thresholds >0.1 TPM) that couple to the major types of G proteins (Gs, Gi/o, Gq/11, G12/13), most frequently Gi/Go and Gq/G11 (**Figs S6A-B**). We calculated the abundance of GPCRs that couple to each G protein by summing median GPCR expression (TPM), thereby yielding an expression 'repertoire' for each signaling mechanism (**Figs S6C-D**). Gs-coupled GPCRs typically account for the smallest GPCR expression repertoire for which such coupling is known. **Supplement 3** provides GPCR expression and G protein linkage data for normal tissues and solid tumors. Summing expression (TPM) of all GPCRs provides an estimate of the proportion of GPCRs among total mRNA (**Figs S6E-F**).

The GPCR expression repertoire varies among tissues in terms of total expression, number and identities of GPCRs. Tumors typically have a different GPCR repertoire than normal tissue, with increased or decreased expression of many GPCRs (**Fig 1C**, **Figs S6E-F)**. Total GPCR expression and the number of GPCRs above expression thresholds increases in certain tumors (e.g., PDAC) but decreases in others (e.g., LIHC and SKCM) compared to normal tissue. Tumors also differ from normal tissue with respect to the abundance of GPCRs that couple to different G proteins (**Fig 1I, Figs S6C-D)**, suggesting changes in signaling. For example, Gs-coupled GPCR expression decreases in many tumors (e.g., PDAC, **Fig 1I**), implying decreases in cAMP signaling.

GPCRs as potential therapeutic targets in cancer

Among the >200 GPCRs overexpressed in at least one of the 45 tumor subtypes, 77 are targets for drugs approved by the FDA and/or EMA. (**Supplement 3**). Among these GPCR drug targets, >50% are overexpressed in 4 or more tumor subtypes (**Fig 5A**) and 15 GPCRs are increased in expression in 10 or more tumor subtypes. These results highlight the potential of GPCRs as targets in cancers and importantly, for the possible repurposing of drugs approved for other indications. Among the 77 GPCRs with increased expression in tumors, nearly two thirds link to either Gs- or Gi-coupled signaling and thus are predicted to regulate cAMP formation (**Fig 5B**).

Of the solid tumor types we analyzed, lung, colon, pancreatic, breast, and prostate cancers account for the largest annual number of deaths in the U.S. (https://www.cancer.gov/types/common-cancers). **Figure 5C** shows currently druggable GPCRs with increased expression in subtypes of those tumor types. Approved drugs target at least 10 GPCRs that have increased expression in those tumors. Hierarchical clustering of GPCR expression in different tumor types (based on their median expression of all GPCRs) reveals that GPCR expression distinguishes tumor types into groups that are consistent with other molecular/physiological traits (**Fig 5D**). For example, squamous cell carcinomas and adenocarcinomas form separate clusters, implying that GPCR expression appears to characterize categories of tumors and that certain GPCRs may be targets across tumor classes/families.

Further, data indicating the relevance of GPCRs as potential prognostic indicators in multiple tumors (**Figs 2F-G and 5E-G**) as well as the observation that GPCRs as a gene family appear enriched in tumors (**Fig 5H-I**) highlight the disease relevance and potential as targets represented by highly expressed/overexpressed GPCRs.

Functionality of overexpressed GPCRs

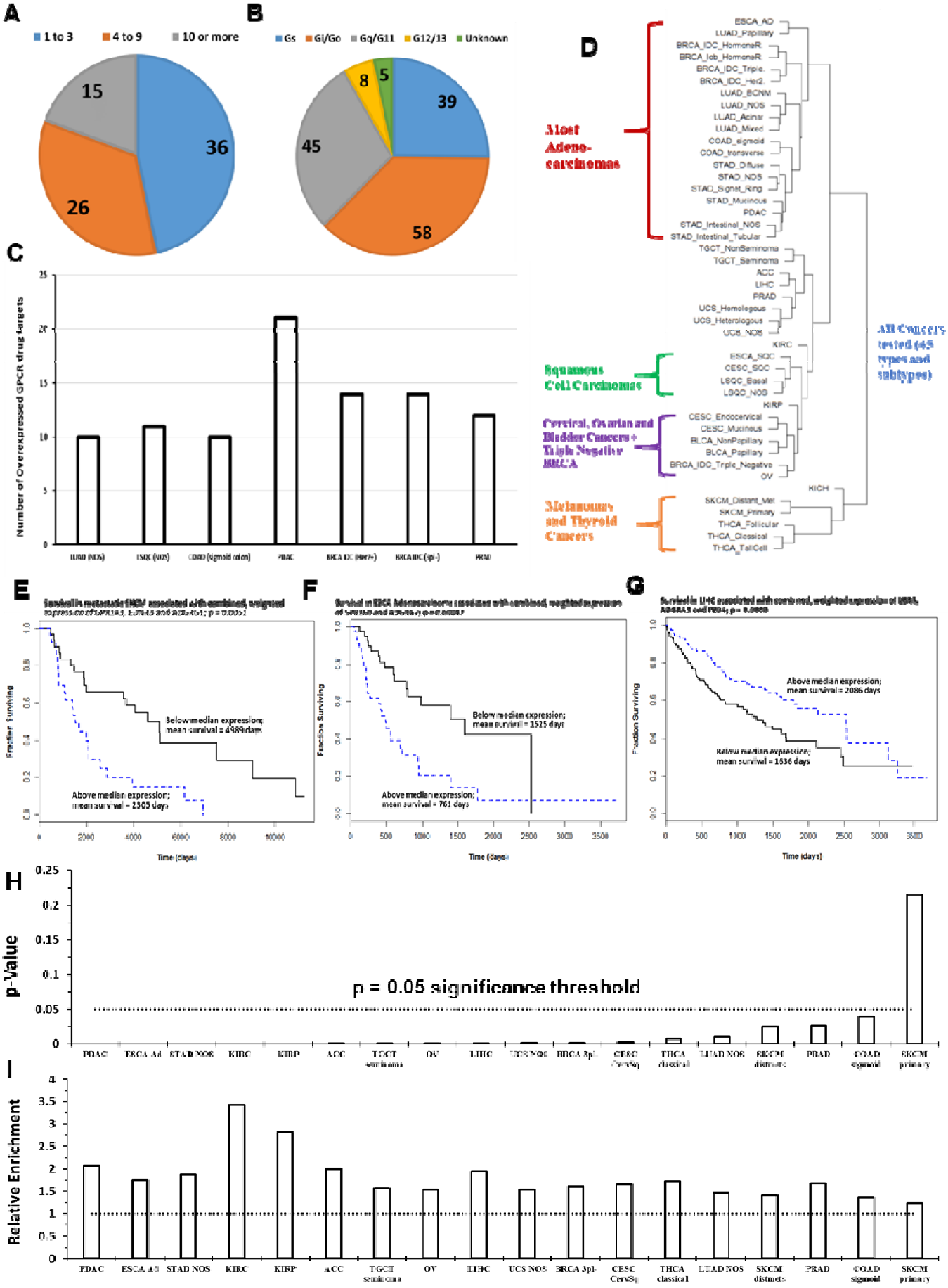
Evidence for functional roles in cancer cells of GPCRs that are highly expressed and overexpressed in solid tumors and cancer cells include findings for *PAR1/F2R* in breast cancer [28], gastric cancer [29], colon cancer [30] and melanoma [31] cells and for *PAR2/F2RL1 [31]* in melanoma, breast [32], and colon cancer cells [33]. Higher *PAR2* expression in ovarian cancer predicts poorer prognosis [34]. *EDNRB,* which is highly overexpressed in SKCM, promotes migration and transformation of melanocytes and melanoma cells and inhibition of *EDNRB* is pro-apoptotic [35, 36]. *GPR143* promotes migration [37] and chemotherapeutic resistance [38] of melanoma cells.

GPR160 and *GPRC5A*, two frequently overexpressed GPCRs, are orphan receptors that influence the malignant phenotype. Knockdown of *GPR160* in prostate cancer cells increases apoptosis and growth arrest [39]. It has been suggested that *GPRC5A* is an oncogene that promotes proliferation, migration and colony formation of PDAC cells [40-42]. GPCRs with increased expression may thus be functional in cancer cells and activated by endogenous agonists or have constitutive activity that regulates signaling via heterotrimeric G proteins and/or β-arrestin [4]. At least certain of the many overexpressed GPCRs may thus serve as phenotypic drivers ("functional oncogenes").

Incorporating similar omics analysis to that presented in this study, our laboratory has recently shown [42] that *GPR68* (a proton sensing GPCR) is highly overexpressed in PDAC tumors, in particular, in pancreatic cancer associated fibroblasts (CAFs). We validated these data at the protein level as well and discovered that *GPR68* mediates a symbiotic cross-talk between CAFs and PDAC cells and contributes to the tumor phenotype. Such findings provide an example as to how the omics data in this study can reveal overexpressed GPCRs that may be relevant to cancer cells themselves and to other cells in the tumor microenvironment.

Figure 5. Solid tumor-expressed GPCRs that are targets for approved drugs
(**A**) The number of GPCRs that are targets for approved drugs and have increased expression in 1-3, 4-9 or ≥10 tumor subtypes.
(**B**) The linkage to G proteins of the 77 GPCRs targeted by approved drugs and with increased expression in at least 1 tumor subtype. Note: multiple GPCRs couple to more than one G protein.
(**C**) The number of GPCRs targeted by approved drugs that show increased expression in lung, colon, pancreatic, breast, and prostate cancers, the leading causes of cancer deaths in the U.S.
(**D**) Hierarchical clustering, based on GPCR expression, of all 45 types and subtypes of tumors studied.
(**E-G**) Kaplan-Meier survival curves in the indicated tumor types, for weighted, combined expression of the GPCRs as indicated—too hard to read, will need to be enlarged and/or more text needed here
(**H**) Overrepresentation of GPCRs among genes with >4-fold elevated expression (FDR < 0.05) for the indicated tumor types/subtypes with p-value calculated via Fischer's exact test
(**I**) The magnitude of overrepresentation (relative enrichment) of GPCRs corresponding to the p-value in (**H**)

FIGURE 5


A
1 to 3
4 to 9
10 or more
36
26
15
B
Gs
Gi/Go
Gq/G11
G12/13
Unknown
39
58
45
8
5
C
Number of Overexpressed GPCR drug targets
D
Most Adeno-carcinomas
Squamous Cell Carcinomas
Cervical, Ovarian and Bladder Cancers + Triple Negative BRCA
Melanomas and Thyroid Cancers
All Cancers tested (45 types and subtypes)
E
F
G
Below median expression; mean survival = 4989 days
Above median expression; mean survival = 2305 days
Below median expression; mean survival = 1525 days
Above median expression; mean survival = 761 days
Above median expression; mean survival = 2086 days
Below median expression; mean survival = 1536 days
Fraction Surviving
Time (days)
H
p-Value
p = 0.05 significance threshold
PDAC
ESCA Ad
STAD NOS
KIRC
KIRP
ACC
TGCT seminoma
OV
LIHC
UCS NOS
BRCA 3pl
CESC CervSq
THCA classical
LUAD NOS
SKCM distmets
PRAD
COAD sigmoid
SKCM primary
I
Relative Enrichment

Somatic mutations of GPCRs in solid tumors

Analysis of 5103 TCGA samples in 20 tumor types (**Table S1**; 21 tumor types if ESCA is split into Esophageal adenocarcinoma and squamous cell carcinomas) revealed many GPCRs with frequent non-silent mutations (**Fig 6A, S7A**), including a more frequently mutated subset (**Fig 6A, inset**), *GPR98*/*ADGRV1,* the most frequently mutated GPCR, occurs in >8% of TCGA samples. Tumor types with high mutation burdens have a high frequency of GPCR mutations (**Fig 6B**). Melanoma (SKCM) has the highest frequency: ~40% SKCM tumors have *GPR98* mutations (**Fig 6C, H**). Approximately 65% of tumors have ≥1 non-silent GPCR mutation. Certain GPCRs are mutated in >10% of specific tumor types (**Fig 6C**). Nmut, the number of genes with somatic non-silent mutations per tumor genome, and the number of mutated GPCRs scale linearly in individual tumors (**Figs 1E, S7E-G**). Frequently mutated GPCRs (e.g., *GPR98*, *GPR112*, *BAI3*) are more likely to be mutated as Nmut increases (**Fig 6F,** SKCM as an example). The relationship between Nmut and likelihood of GPR98 mutation is similar in SKCM and other cancers (**Fig 6G**); this is also observed for other frequently mutated GPCRs. Hence, the likelihood of a GPCR being mutated appears to depend on the accumulation of genome damage and to be independent of the mechanisms for the mutations. The relationship between GPCR mutation rates and Nmut is identical in BLCA, LUAD and SKCM, although the factors driving DNA damage and oncogenesis are likely different. Mutations of certain GPCRs, such as GPR98, may thus serve as a bellwether for genome-wide DNA damage.

Missense mutations and in-frame deletions are the most frequent non-silent mutations in GPCR genes (**Fig S7C-D, Table S3**). Mutations in frequently mutated GPCRs occur at many sites (**Fig S8A**), which is in contrast with the smaller number of such sites in common oncogenes, e.g., *KRAS [10].* (**Fig S8B**). Certain GPCR genes (e.g., *GPR98*) may be in genomic regions vulnerable to dysregulation of DNA damage and repair and belong to a subset of mutated genes; *GPR98* mutations frequently occur alongside other frequently mutated genes such as *TTN* and *MUC16* (**Fig S9A-G**). *GPR98* is among the 25 most frequently mutated genes in all tumor types surveyed; its mutational frequency is similar to that of genes (e.g., *BAGE2* [**Fig S7B]**) that are mutational hotspots [43]. As the GPCR with the largest gene length (~19,000 bp), *GPR98* has more mutational events. Compared with other very long genes, i.e., genes > 15,000 bp, *GPR98* belongs to a subset of ~10 genes with high mutational frequencies (**Fig 6I**), implicating *GPR98* as a hotspot for mutations. *GPR98* has a ~4-fold increased density of mutational events (normalized for gene length) compared to the average among these very long genes.

Most mutated GPCR genes have low levels of mRNA expression (**Fig 6H**, primary SKCM as an example) so may not be functionally relevant but certain GPCR genes (e.g., *CELSR1* and *LPHN2*/*ADGRL2*) are frequently mutated and moderately/highly expressed. Because several such GPCRs are orphan receptors (without known physiologic agonists or roles), it is unclear if they impact on cell function. As cell-surface receptors, frequently mutated, well expressed GPCRs may represent neo-antigens. For SKCM, which has the most GPCR mutations among tumors types surveyed, DE analysis of primary melanomas and distant metastases that have or lack GPCR mutations (e.g., *GPR98* and *LPHN2*) revealed little evidence that these mutations alter the tumor transcriptome, implying that such GPCR mutations are likely passenger, rather than driver, mutations (**Fig 6D**; **Fig S8C, D**). Similar findings were noted for other tumors (e.g., BLCA) which have frequent GPCR mutations. As a further approach, we evaluated GPCR mutations, predicting the likelihood of functional consequences and site-specific enrichment of the mutations via MutSIGCV 2.0 (*gdac.broadinstitute.org*). The majority of GPCRs frequently mutated (**Fig 6J**, SKCM as example) show non-silent mutations that are non-significant in terms of enrichment (compared to the background mutation rate of silent mutations over these same regions) for individual mutation sites. These mutations are not predicted to be functional by MutSIGCV 2.0, consistent with the idea that the frequent GPCR mutations are likely passenger and not driver mutations.

FIGURE 6

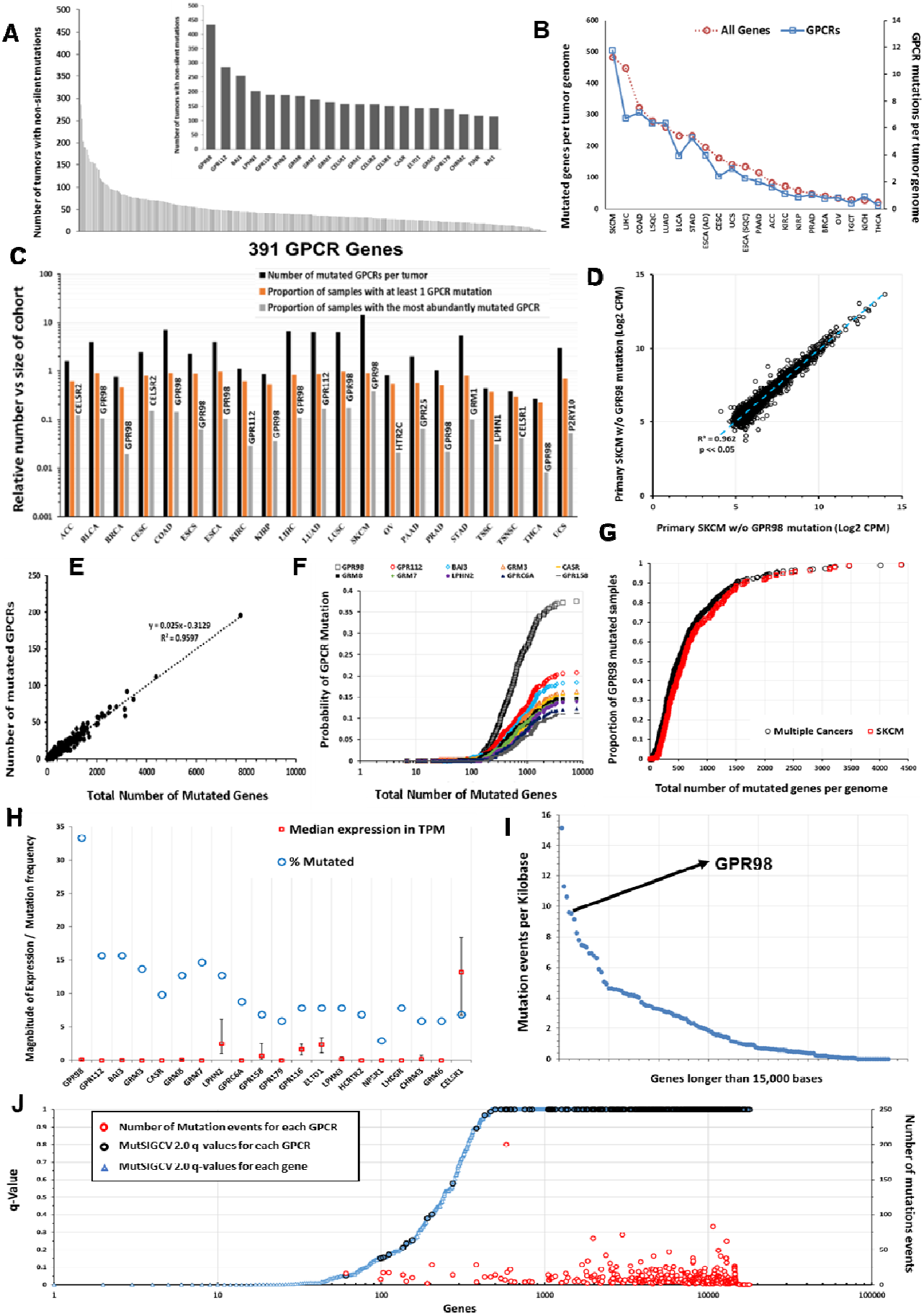

A
Number of tumors with non-silent mutations
391 GPCR Genes
B
All Genes
GPCRs
Mutated genes per tumor genome
GPCR mutations per tumor genome
C
Number of mutated GPCRs per tumor
Proportion of samples with at least 1 GPCR mutation
Proportion of samples with the most abundantly mutated GPCR
Relative number vs size of cohort
D
Primary SKCM w/o GPR98 mutation (Log2 CPM)
Primary SKCM w/o GPR98 mutation (Log2 CPM)
R² = 0.962
p << 0.05
E
Number of mutated GPCRs
y = 0.025x - 0.3129
R² = 0.9597
Total Number of Mutated Genes
F
Probability of GPCR Mutation
Total Number of Mutated Genes
G
Proportion of GPR98 mutated samples
Multiple Cancers
SKCM
Total number of mutated genes per genome
H
Median expression in TPM
% Mutated
Magnitude of Expression / Mutation frequency
I
GPR98
Mutation events per Kilobase
Genes longer than 15,000 bases
J
Number of Mutation events for each GPCR
MutSIGCV 2.0 q values for each GPCR
MutSIGCV 2.0 q-values for each gene
q-Value
Number of mutations events
Genes

Figure 6: Somatic non-silent mutations of GPCRs

(**A**) Frequency of GPCR mutations in the TCGA cohort (n = 5103). Inset: 20 most frequently mutated GPCRs.

(**B**) The average number of all genes (red) and GPCRs (blue) with somatic, non-silent mutations per tumor genome for the TCGA tumor types surveyed

(**C**) The number of mutated GPCRs per tumor for several types of solid tumors in TCGA (black), the proportion of samples in each tumor type with at least one mutated GPCR (orange) and proportion with non-silent mutations for the most commonly mutated GPCR (gray) for each tumor type

(**D**) Expression (CPM) of the 5000 most abundant genes in SKCM correlate closely in primary SKCM tumors that have or lack *GPR98* mutations(**E**) GPCR mutation frequency is linearly related to Nmut in SKCM

(**F**) Probability of GPCR mutation as Nmut increases in SKCM for the 10 most frequently mutated GPCRs

(**G**) Normalized probability of GPR98 mutation as Nmut increases in SKCM, and the same for several cancers with high mutational burden and combined for BLCA, LUAD, LSQC, COAD, and SKCM.

(**H**) The 20 most frequently mutated GPCRs in primary SKCM, with frequency of mutation and median (and upper and lower quartile) expression in TPM

(**I**) The number of mutation events per 1000 bp in SKCM for genes > 15,000 bp in length; GPR98 is highlighted

(**J**) MutSIGCV 2.0 scores in SKCM obtained from https://gdac.broadinstitute.org/, showing the q-values for the significance of mutation scores for each annotated coding gene (blue); GPCR (black) and for those GPCRs, the number of mutation events (red) among the SKCM cohort

Figure 7. Copy number variations (CNVs) of GPCRs in solid tumors

(**A**) The number of solid tumors with CNV for each GPCR across all TCGA samples plotted in descending order of frequency for high-level amplification and homozygous deletions) (see text for definition of high- and low-level amplification)

(**B**) The same as in (**A**) for low-level amplification and single-copy deletions

(**C**) In SKCM tumors (n = 367), the distribution of high-level amplification of GPCRs (n = 390 genes) compared to that of all protein coding genes (n = 24,776)

(**D**) The total number of homozygous deletions, single copy deletions, low-level and high-level amplifications for all GPCRs combined in the 7545 TCGA tumors surveyed for CNV

(**E**) The most frequently amplified GPCR for each TCGA tumor type and the proportion of samples with High and Low-level amplification for that GPCR

(**F**) Proportion of samples of various tumors types with High and Low-level amplification of *GPR160*, the most frequently amplified GPCR overall)

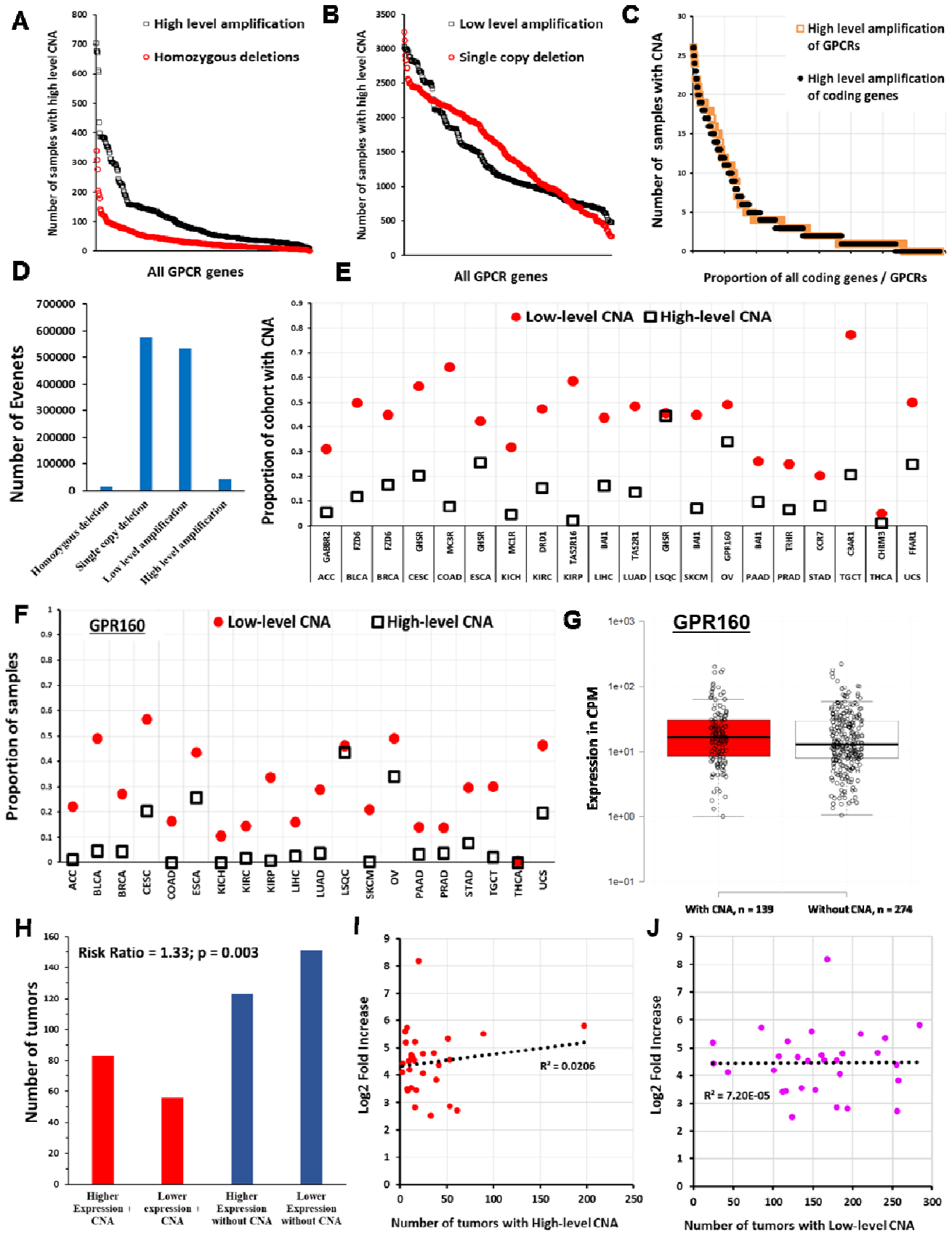
(**G**) Ovarian cancer samples with and without High-level amplification of GPR160, with the median for each group indicated. The difference between groups was not statistically significant)

(**H**) The risk ratio of elevated GPR160 expression (above the median value for OV) when GPR160 also shows high-level amplification; amplification of GPR160 increases the likelihood that GPR160 expression is elevated

(**I**) For the 30 GPCRs in OV with the highest fold-increase in expression relative to normal ovarian tissue (with FDR < 0.05 and median expression in OV > 1TPM), the corresponding number (of 579) OV tumors with high-level amplification of those same GPCRs

(**J**) The same as (**I**), but comparing fold-increase against low-level amplification

FIGURE 7


A
High level amplification
Homozygous deletions
Number of samples with high level CNA
All GPCR genes
B
Low level amplification
Single copy deletion
Number of samples with high level CNA
All GPCR genes
C
High level amplification of GPCRs
High level amplification of coding genes
Number of samples with CNA
Proportion of all coding genes / GPCRs
D
Number of Evenets
Homozygous deletion
Single copy deletion
Low level amplification
High level amplification
E
Low-level CNA
High-level CNA
Proportion of cohort with CNA
GABBR2 FZD6 FZD6 GHSR MC3R GHSR MC1R DRD1 TAS2R16 IIA1 TAS2R1 GHSR IIA1 GPR160 IIA1 TRHR CCR7 C3AR1 CHRM3 FFAR1
ACC BLCA BRCA CESC COAD ESCA KICH KIRC KIRP LIHC LUAD LSQC SKCM OV PAAD PRAD STAD TGCT THCA UCS
F
GPR160
Low-level CNA
High-level CNA
Proportion of samples
ACC BLCA BRCA CESC COAD ESCA KICH KIRC KIRP LIHC LUAD LSQC SKCM OV PAAD PRAD STAD TGCT THCA UCS
G
GPR160
Expression in CPM
With CNA, n = 139
Without CNA, n = 274
H
Risk Ratio = 1.33; p = 0.003
Number of tumors
Higher Expression + CNA
Lower expression + CNA
Higher Expression without CNA
Lower Expression without CNA
I
Log2 Fold Increase
R² = 0.0206
Number of tumors with High-level CNA
J
Log2 Fold Increase
R² = 7.20E-05
Number of tumors with Low-level CNA

Copy-number variation (CNV) of GPCRs in solid tumors

CNV of certain GPCRs occurs frequently in TCGA solid tumor samples (**Fig 7A-B, Table S1**), with some GPCRs (e.g., *GPR160*) amplified in >5% of all TCGA samples surveyed. CNV data were obtained as GISTIC 2.0 [44] thresholded data, wherein values of -2, -1, 0, 1 and 2 respectively denote homozygous deletion, single copy deletion, diploid copy number, low level amplification (i.e. increase of 0.1 to 0.9 of copy number, expressed as a log2 ratio) and high-level amplification (amplification of >0.9 of the log2 ratio, i.e. >1.7 extra copies in a diploid cell) [45]. The distribution of amplification events among GPCRs parallels that of all genes (**Fig 7C,** SKCM as an example) but a subset of GPCRs is disproportionally amplified (**Fig 7A, and Supplement 3**). Most GPCRs have infrequent amplification (in 2% of tumors or less). Amplification does not predict high expression or overexpression of GPCRs (**Fig 7I, J**; OV as an example): frequently amplified GPCRs in tumors often have limited mRNA expression in those tumors while most highly expressed, overexpressed GPCRs are not amplified.

Single-copy/heterozygous deletions of GPCRs are widespread, whereas homozygous deletions are extremely rare (**Fig 7A, D**). Generally, GPCR genes with single-copy deletions are not significantly expressed in tumors or normal tissues, implying that such deletions lack functional consequences. However, exceptions exist. *PTH1R*, which is frequently deleted in KIRC (~77% samples have single-copy deletions) has ~10-fold reduced expression compared to normal kidney tissue. Similarly, *ADRA1A* is frequently deleted and has reduced expression in hepatocellular and prostate adenocarcinomas.

Figure 7E shows the identity/frequency of amplification of the most frequently amplified GPCR in each tumor type. Several cancers (e.g., OV and LSQC) have a high level of amplification of specific GPCRs in >25% and low-level amplification in >40% of samples. **Figure 7F** shows the CNA frequency of *GPR160*, the most frequently amplified GPCR overall, among all tumors surveyed. Except for *GPR160* and *FZD6*, most GPCRs with frequent amplification are rarely overexpressed (**Supplement 2**). CNA alone thus does not generally predict increased mRNA expression in tumors compared to normal tissue and highly expressed GPCRs in tumors are typically not amplified. We tested for correlation between high-level amplification and increased mRNA expression. **Figure 7G** shows GPR160 expression in OV tumors with and without high-level amplification; CNA is not a prerequisite for high mRNA expression and the small difference in median expression is not statistically significant, based on DE analysis via edgeR. However, tumors with amplification of GPR160 show a higher likelihood (~33%, **Fig 7H**) of expressing GPR160 at levels above the median for OV. We did not observe statistically significant risk ratios relating GPCR expression with amplification for other frequently amplified GPCRs. We conclude that CNA and GPCR mRNA expression are generally poorly correlated; hence, examination of amplified GPCR genes does not predict which GPCRs are highly and/or differentially expressed in a tumor (**Fig 7I, J**).

Supplements 2 and 3 provide, respectively, the frequency of GPCR CNV and changes in expression of GPCRs in the tumors surveyed. The widespread CNV of certain GPCRs suggests that they (and/or neighboring genes that vary along with these GPCRs) contribute to the malignant phenotype and may be markers for malignancy.

Discussion and Conclusions

This resource identifies mutations, CNV and alterations in mRNA expression of GPCRs in a range of solid tumors and reveals a broad landscape of changes, suggesting a role for this gene superfamily in such tumors. The results illuminate potential functional roles and possible therapeutic utility of GPCRs in a large number of cancers.

Mutations of certain GPCRs have been implicated in cancer [9], but a comprehensive analysis of GPCR amplification, expression and DE has been lacking. The largest public datasets of normal (GTEx) [1] and cancer tissues (TCGA) provide RNA-seq data analyzed/normalized differently, making it difficult to compare datasets. For many tumors types, few replicates of 'normal' TCGA tissue are available and samples from matched non-tumor tissue from patients may not be representative of normal tissue (**Supp. Methods, Section 1**). The TOIL project enabled comparison of TCGA and GTEx data with analysis of both RNA-seq datasets via the same pipeline.

Our analysis identified frequently mutated GPCRs (e.g., *GPR98/ADGRV1* and *GPR112/ADGRG4*) in multiple cancers, especially melanoma (SKCM). GPCR mutations appear to reflect accumulation of DNA damage and mutations across the genome and may be tumor markers for this process. The low expression of the most frequently mutated GPCRs suggests that they are not likely driver mutations, supported by our finding that presence/absence of GPCR mutations appears to have little impact on the tumor transcriptome, Analysis via MUTSIGCV 2.0 supports this idea. Certain adhesion GPCRs have been proposed as potential targets based on having frequent mutations in cancer [9]. Our analyses suggest that these are passenger mutations, but may nonetheless shed light on oncogenesis, as discussed above.

CNV, in particular amplification and single-copy deletion of GPCRs, is widespread in solid tumors. CNV of GPCRs generally does not predict DE and seems to be neither necessary nor sufficient for DE of a GPCR. Especially with regards to amplification, GPCR mRNA expression appears independent of CNV, with *GPR160* as an exception. CNV is not stochastically distributed among the GPCR family, Certain GPCRs are more frequently amplified (e.g., *GPR160*) or deleted (e.g., *PTH1R* in KIRC). Amplified and deleted GPCRs may have potential as biomarkers [46].

Numerous solid tumors have increased mRNA expression of large numbers of mostly non-mutated GPCRs: 72 GPCRs are overexpressed in >10 tumor subtypes, implying that common mechanisms may regulate GPCR expression in such tumors. Highly overexpressed GPCRs are potential candidates as drug targets. Of note, 77 such overexpressed GPCRs are targets of approved drugs that have the potential to be repurposed to treat tumors. The similarity of GPCR expression in primary tumors and metastases supports such therapeutic potential.

Known driver mutations do not appear to influence GPCR expression in tumors but we excluded rare mutations. In order to ensure large numbers of replicates and high statistical significance, we analyzed tumors with high-frequency mutations (e.g., *P53* or *KRAS*) [10]. Highly expressed GPCRs are widely expressed among replicates of specific tumor types, and are more prevalent in tumors than are common driver mutations. Transcriptional regulation of GPCRs is poorly understood. Studies of GPCRs with DE in tumors may shed light on such mechanisms and perhaps also have relevance for other disease settings with altered GPCR expression. The overrepresentation of GPCRs among protein-coding genes with increased expression in solid tumors supports the hypothesis that the elevated expression of specific GPCRs is a hitherto underappreciated feature of solid tumors,

The data reveal that clusters of GPCRs may be prognostic indicators for survival and provide a molecular signature of the malignant phenotype. GPCRs whose expression adversely or positively predicts survival are candidates for antagonists or agonists, respectively as novel cancer drugs. Hierarchical clustering of tumor types based on GPCR expression identifies groups of tumors consistent with other molecular/phenotypic features of these tumors. Thus, the tumor GPCRome appears to be predictive of the broader molecular landscape of tumors.

Do GPCR mRNA data predict protein expression? Direct quantification of GPCR proteins is challenging, due to their generally low abundance and paucity of well-validated antibodies. However, mRNA expression of GPCRs,

especially highly expressed GPCRs, generally predicts the presence of functionally active GPCRs in human and animal cells [47-51]. In contrast to earlier ideas, recent evidence supports the view that mRNA expression broadly predicts protein expression [52-56] (**Supplemental Note 1**). As an example, GPRC5a protein and mRNA abundance are concordant (**Supplemental Note 2; Fig S10**). GPCR detection via mass-spectrometry has been challenging; proteomics data (e.g., [57] indicate that at present, few GPCRs are detectable by such methods, likely due to the low abundance of GPCR proteins. As noted in **Results**, functional evidence is available for numerous GPCRs with DE in solid tumors. As cell-surface proteins enriched in tumors and cancer cells, certain GPCRs may represent novel tumor-associated proteins that might be targeted for diagnosis and/or treatment.

GPCR mutations, CNV and DE thus occur at a high frequency in solid tumors. Therefore, this receptor super-family may have unappreciated functional roles in such tumors, especially since GPCR expression appears to be largely independent of tumor grade/stage and mutations. Our results imply that new insights may derive from further studies of GPCRs regarding mechanisms of gene expression and phenotype in solid tumors and perhaps other cancers. Of particular, and perhaps rapid, translational importance is the potential of GPCR-targeted drugs, including FDA/EMA-approved drugs that might be repurposed as therapeutics for a variety of solid tumors.

Acknowledgements

This work was supported by the American Society of Pharmacology and Experimental Therapeutics (ASPET) David Lehr Award and Padres Pedal the Cause #PTC2017.

Results are in part based upon data from the Cancer Genome Atlas (TCGA) Research Network: http://cancergenome.nih.gov/.

Author Contributions

KS and PAI conceived the project. KS designed and implemented the analysis approach and co-wrote the manuscript. PAI directed the project, interpreted results and co-wrote the manuscript. KM implemented analysis methods and maintained datasets/databases and the project website. RC co-wrote and edited the manuscript. HC interpreted results, designed analysis approach and co-wrote the manuscript.

Supplemental Information

Supplement 1: Supplementary Figures 1-12, Supplementary Tables 1-5 and Supplementary Notes 1-3.
Supplement 2: Data file with GPCRs annotated by GtoPdb plus GPCR mutations and CNV
Supplement 3: Data file with GPCR expression in normal tissue, tumors and cancer cells and GPCRs with DE
Additional downloadable content: *insellab.github.io*

Table 1: Tumors surveyed for Differential Expression (DE) analysis. TCGA cancer type and sub-classification, if applicable, for solid tumors with distinct histological classification are shown, along with the number of replicates and GPCRs with increased or decreased expression for each type of tumor.

	Cancer Type	Histology/Subtype	Replicates	# GPCRs ↑	# GPCRs ↓
1	Adrenocortical Cancer (ACC)	Adrenocortical carcinoma - Usual Type	73	22	4
2	Bladder Cancer (BLCA)	Papillary bladder cancer (BLCA_Pap)	130	26	12
		Non-papillary bladder cancer (BLCA_NonPap)	267	36	14
3	Breast Cancer (BRCA)	Infiltrating ductal carcinoma (IDC), Her2 positive (BRCA_IDC_Her2+)	48	46	25
		IDC, Hormone Receptor positive (BRCA_IDC_HR+)	431	49	29
		IDC, Triple positive (BRCA_IDC_3pl+)	54	51	25
		IDC, Triple negative (BRCA_IDC_3pl-)	109	51	24
		Invasive Lobular Carcinoma (ILC), Hormone R positive (BRCA_Lob_HR+)	57	50	31
4	Cervical Cancer (CESC)	Cervical Squamous Cell Carcinoma (CESC_CervSq)	252	40	18
		Endocervical Adenocarcinoma of the Usual Type (CESC_ECAD)	21	38	23
		Mucinous Adenocarcinoma of Endocervical Type (CESC_Muc)	17	38	24
5	Colon Cancer (COAD)	Colon Adenocarcinoma in the sigmoid colon (COAD_Sig)	71	41	28
		Colon Adenocarcinoma in the transverse colon (COAD_Trans)	22	31	23
6	Esophageal Cancer (ESCA)	Esophagus Adenocarcinoma (ESCA AD)	89	59	17
		Esophagus Squamous Cell Carcinoma (ESCA SQC)	92	43	9
7	Kidney Papillary Cell Carcinoma (KIRP)	-	288	28	14
8	Kidney Clear Cell Carcinoma (KIRC)	-	523	65	11
9	Kidney Chromophobe (KICH)	-	66	29	10
10	Liver Cancer (LIHC)	Liver Hepatocellular Carcinoma (LIHC)	360	11	7
11	Lung adenocarcinoma (LUAD)	Lung Papillary Adenocarcinoma (LUAD_Pap)	23	22	33
		Lung Bronchioloalveolar Carcinoma Non-Mucinous (LUAD_BCNM)	19	34	36
		Lung Adenocarcinoma- Not Otherwise Specified (LUAD_NOS)	308	33	33
		Lung Adenocarcinoma – Mixed (LUAD_Mixed)	105	29	31
		Lung Acinar Adenocarcinoma (LUAD_Acinar)	18	27	36
12	Lung squamous cell carcinoma (LSQC)	Lung Squamous Cell Carcinoma- Not Otherwise Specified (LSQC_NOS)	468	34	31
		Lung Basaloid Squamous Cell Carcinoma (LSQC_Basal)	14	38	39
13	Skin Cutaneous Melanoma (SKCM)	Primary melanomas (SKCM_Primary)	100	34	18
		Distant metastases (SKCM_DMet)	68	41	11
14	Ovarian Cancer (OV)	Ovarian Serous Cystadenocarcinoma (OV)	418	57	11
15	Pancreatic Cancer (PAAD)	Pancreatic Ductal Adenocarcinoma (PDAC)	147	68	11
16	Prostate Cancer (PRAD)	Prostate Adenocarcinoma Acinar Type (PRAD)	475	27	25
17	Stomach Cancer (STAD)	Stomach, Adenocarcinoma, Diffuse Type (STAD_Diff)	68	53	12
		Stomach, Adenocarcinoma, Not Otherwise Specified (STAD_NOS)	154	48	14
		Stomach, Intestinal Adenocarcinoma, Mucinous Type (STAD_Muc)	19	55	15
		Stomach, Intestinal Adenocarcinoma, Not Otherwise Specified (STAD_IntNOS)	73	40	16
		Stomach, Intestinal Adenocarcinoma, Tubular Type (STAD_IntTub)	76	39	12
		Stomach Adenocarcinoma, Signet Ring Type (STAD_Sig)	12	37	10
18	Testicular Cancer (TGCT)	Seminoma (TGCT_Sem)	72	73	16
		Non-seminoma (TGCT_NonSem)	65	76	9
19	Thyroid Cancer (THCA)	Thyroid Papillary Carcinoma - Classical/usual (THCA_Usual)	358	33	15
		Thyroid Papillary Carcinoma - Follicular (>= 99% follicular patterned) (THCA_fol)	101	17	6
		Thyroid Papillary Carcinoma - Tall Cell (>= 50% tall cell features) (THCA_TC)	36	33	17
20	Uterine Carcinosarcoma (UCS)	Uterine Carcinosarcoma/Malignant Mixed Mullerian Tumor (MMMT): (UCS_NOS)	24	43	19
		Uterine Carcinosarcoma/ MMMT: Heterologous Type (UCS_Het)	20	44	22
		Uterine Carcinosarcoma/MMMT: Homologous Type (UCS_Homo)	13	43	20

Methods

Software used and availability of data

RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
R v 3.3.2	The Comprehensive R Archive Network	https://cran.r-project.org/
EdgeR	Bioconductor	http://bioconductor.org/packages/edgeR
Deposited Data		
Numbers of GPCR mutations in each tumor type	This paper	Supplement 2
Occurrence of CNV for GPCRs in each tumor type	This paper	Supplement 2
GPCR expression and G protein linkages in tumors and healthy tissue	This paper	Supplement 1
GPCRs with DE in tumors	This paper	Supplement 1
GPCR expression in cancer cells	This paper	Supplement 1
GPCRs annotated by IUPHAR	IUPHAR/BJP	Supplement 2
Analyzed TCGA mutation data	TCGA/UCSC Xena	*https://insellab.github.io/gpcr_mutations*
Analyzed TCGA CNV data	TCGA/UCSC Xena	*https://insellab.github.io/gpcr_cnv*
Analyzed TCGA RNAseq data	TCGA/UCSC Xena	*https://insellab.github.io/gpcr_tcga_exp*
Analyzed GTEX RNAseq data	GTEX/UCSC Xena	*https://insellab.github.io/gpcr_gtex_exp*
Analyzed RNAseq data from cancer cells	Various cited sources	*https://insellab.github.io/gpcr_cells_exp*

1. Contact for Resource Sharing

A website has been created for sharing all data at *https://insellab.github.io/*. Links to data files will be posted on this website following peer review. All data provided will be open access following peer review; information about how to cite data generated from this study is available at *https://insellab.github.io/*.

Contact for access to data:
Krishna Sriram
Department of Pharmacology, UC San Diego.
ksriram@ucsd.edu
Tel: 858-534-2298

2. Details of Methods

2.1. Differential Expression (DE) Analysis:

Gene expression for the GTEX and TCGA datasets, assayed via RNAseq, was downloaded from the UCSC Xena Portal (xena.ucsc.edu). For DE analysis, RSEM expected counts were obtained, which were computed via the TOIL pipeline, described in [11] and available at the Xena Portal by the authors of the TOIL project (*https://xenabrowser.net/datapages/?host=https://toil.xenahubs.net*).

The analyzed data from the TOIL project were generated as follows. Merged FASTQ files were adapter-trimmed via CUTADAPT, followed by alignment via STAR [58]. Gene expression was then quantified using RSEM [59]. The HG38 reference genome, with Gencode V23 annotations, was used in the TOIL analysis. For this study, both RSEM estimated counts (for DE analysis) and RSEM TPMs (for evaluating magnitudes of expression) were used.

Files were accessed from (*https://xenabrowser.net/datapages/?dataset=tcga_gene_expected_count&host=https://toil.xenahubs.net*) for TCGA expected counts (version 2016-09-01) and (*https://xenabrowser.net/datapages/?dataset=gtex_gene_expected_count&host=https://toil.xenahubs.net*) for GTEX expected counts (version 2016-05-19).

Expression in TPMs for GPCRs was queried via *https://xenabrowser.net/heatmap/* for both TCGA and GTEX.

Following the download of gene expression data, corresponding files on sample phenotype were obtained from the relevant links hosted at *https://xenabrowser.net/datapages/?host=https://tcga.xenahubs.net* for TCGA and at *https://xenabrowser.net/datapages/?cohort=GTEX* for GTEX samples, respectively. Samples were then grouped for DE analysis based on attributes such as tissue type and tumor type.

We used a table of estimated counts from tumor and normal tissue as input for analysis in edgeR [60], yielding normalized abundance in CPM (using TMM normalization) and DE, showing magnitude of fold change and statistical significance estimated by FDR (False Discovery Rate). We estimated fold-changes of gene expression in tumors compared to normal tissue via an exact test. Genes that changed with FDR <0.05 were considered statistically significant, however we focused attention on GPCRs that besides low FDRs are also expressed at >1 TPM in tumors, as high expressed GPCRs are likely of greater interest. The 20 TCGA tumor types were divided into 45 subtypes/categories (**Table 1**) based on histological classification. Different tumor subtypes show distinct GPCR expression, e.g., Classical vs. Follicular Thyroid Cancer (THCA), Triple negative vs. Her2 Breast Cancer Cancer, Infiltrating Ductal Carcinoma (BRCA IDC), and Esophageal (ESCA) squamous cell carcinoma vs. Adenocarcinoma (**Fig S1A-F**), hence requiring this subdivision into tumor subtypes.

In addition to the standard TMM approach in edgeR, we tested upper-quartile normalization before conducting DE analysis in edgeR. The two methods yielded nearly identical results (**Fig S11C, D).** Log2 fold-changes for all genes and GPCRs show that expression changes calculated by both methods are closely correlated, with nearly identical magnitude. We also evaluated DE via EBseq [61]. EBseq and edgeR yielded very similar results (**Fig S11A, B**), in particular for GPCRs, implying that assumptions implicit in the DE analysis via edgeR/TMM normalization do not skew or bias the results. As an empirical test, we verified that GPCRs that show large fold-changes between tumor and normal samples also show large differences in normalized gene expression in TPM, e.g., *EDNRB*, *GPR143* and *ADGRG1* in SKCM (**Fig S2F-H**).

2.2. Database of normalized GPCR expression in tumors and normal tissue

For GPCR expression in tissue, expression as TPM (Transcripts Per Million, a normalization of gene abundance that corrects for effective length of genes) and CPM (Counts Per Million, number of times a gene is encountered per million reads, hence normalization for library size without length normalization) are provided in **Supplement 3**.

Data in TPM are provided for assessing the relative abundance of members of a gene family, such as GPCRs, *within* an individual sample or a set of biological replicates while data in CPM are provided for comparing a specific gene *between* multiple groups of dissimilar samples, where length normalization is problematic because dissimilar data sets may be normalized differently. We provide GPCR expression in both formats to facilitate different approaches for analysis.

This resource thus enables an estimate of GPCR abundance in more rigorous terms than comparing FPKM/TPM values for a particular gene across different tissue types. Further, this approach provides normalized gene expression estimates (in TPM and CPM) using the same units and analysis methods for both normal and cancer tissue, thus allowing for direct comparison.

Gene abundances in CPM were calculated via EdgeR. In several cases the same normal tissue dataset was used to compare multiple tumors (e.g., GTEX Kidney data was used for comparison with KIRP, KICH and KIRC). In these cases, the calculated CPM values for normal tissue from each analysis were (as expected) very similar but not equal, as EdgeR TMM normalization yields slightly different normalization factors in each case. The CPM values presented for these tissues are thus average values (e.g., for CPMs in normal Kidney tissue, the values provided are the average of the data obtained from comparisons of normal kidney with KIRP, KICH and KIRC respectively). The normal tissue types where this was performed are breast, lung and kidney.

2.3. GPCR mutation and copy number analysis

For each cancer type, tables of somatic, non-silent mutations (gene-level) and somatic mutations (SNPs and small INDELs) and Gene-level GISTIC2 thresholded copy number variation were obtained using *https://xenabrowser.net/datapages/?host=https://tcga.xenahubs.net* and links within.

For mutation data, we used results obtained via the Broad Automated Pipeline, where available. In other cases, we used data from the Baylor College of Medicine sequencing center. The source of the mutation data is indicated on the respective downloadable files and **Table S1**. In all cases, the HG19 reference genome was used for calling mutations. Mutation data for genes coding for GPCRs were extracted as part of the present study; all GPCR mutation data are available for download as supplemental material.

TCGA copy number estimates were obtained using Affymetrix SNP 6.0 arrays. The data were analyzed via GISTIC 2.0 [44] to obtain gene-level estimates of copy number variation. The resulting 'thresholded' GISTIC 2.0 data yields values of -2,-1, 0, 1, 2, indicating homozygous/2 copy deletion, heterozygous/single copy deletions, no change, low level amplification, and high level amplification, respectively. Copy number variation for GPCR genes in each tumor type was extracted and is available as downloadable material.

2.4. Which genes are included in this analysis?

We evaluated all GPCRs annotated by IUPHAR/*British Journal of Pharmacology* [2], accessible via (*http://www.guidetopharmacology.org/GRAC/ReceptorFamiliesForward?type=GPCR*). This list primarily focuses on endoGPCRs (GPCRs natively expressed in peripheral tissue, possess endogenous ligands and receptors primarily used as drug targets). The IUPHAR list includes taste and vision GPCRs but not olfactory receptors. We excluded several GPCRs annotated by IUPHAR but which are thought to be pseudogenes. The list of GPCRs in this analysis is provided in **Supplement 2**.

For these annotated GPCRs, we included information about their linkages to G proteins and their status as orphans or not. For non-orphans, an example of an endogenous ligand is provided. These data are almost entirely based on information available at the IUPHAR website cited above. In a few cases where such information is not provided by IUPHAR, we have used other literature sources for annotation.

2.5 GPCR expression in cancer cells mined from other sources

GPCR expression in a range of cancer cell lines was queried via the EBI Expression Atlas (*https://www.ebi.ac.uk/gxa/home*) for cell line profiles part of CCLE [24] as well as by Genentech [25], profiled via RNA-seq. These data were analyzed as part of the EBI Expression Atlas via the iRAP bioinformatic analysis

pipeline, described in detail in [23] wherein gene expression was computed in FPKM (Fragments Per Kilobase of exon, per Million reads), a length-normalized expression abundance estimate analogous to TPM units, used for TOIL TCGA data. Precisely, statistically relevant comparisons between gene abundances in TPM and FPKM are not feasible; however, empirical comparisons between such data sets are possible. In general, genes with high abundances in TPM or FPKM will be highly expressed relative to other genes within sets of samples; hence, our comparison of CCLE and other cell-based data vs. TOIL TCGA data serves as an empirical confirmation of the fact that GPCRs highly expressed in TCGA tumors are also present in cancer cells.

Normalized gene expression in cancer cells from other sources [26, 27, 62] were obtained via NCBI GEO, wherein analyzed RNA-seq data with quantification of gene expression were provided. As with the data from EBI above, such data allowed for empirical comparisons vs. TCGA TOIL data to confirm the presence of GPCRs in cancer cells, which were also detected in tumors.

3. Quantification and Statistical Analysis

DE analysis was performed in the R software environment via EdgeR [60], as discussed above in section 2.1. We used the following criteria to evaluate GPCRs with significant DE:

FDR < 0.05. In the majority of cases, genes with a high fold change also showed FDRs << 0.05.
Magnitude of fold change > 2 fold (increase or decrease).
Magnitude of expression > 1 TPM median expression in tumors, as calculated by RSEM in the TOIL pipeline, discussed in section 2 above. We focused on genes with significant DE and high expression because our primary goal was to identify GPCRs that may be drug targets and/or biomarkers.

For compilation and distribution of data we assembled data files primarily in Microsoft Excel, with files stored in .xlsb format.

Plots of normalized expression in tumors and normal tissue, whether in TPM or CPM, show median expression for respective cohorts, along with upper and lower quartiles, as indicated in figure legends where applicable.

The numbers of replicates in each sample group/category of normal tissue and tumors are provided in **Tables S1 and S2 and Table 1**, respectively. Tumor types with 10 or more replicates were considered for DE analysis. A small number of samples (<< 1% of TCGA samples studied) were excluded because they were duplicated in downloaded databases from TOIL and showed discrepancies between expression in these downloaded data vs expression data queried via the visualization tool on the TOIL website. These discrepant samples are provided on a downloadable list at *insellab.github.io*.

DE results presented in this text are from comparisons between TCGA and GTEX samples, but we also include in our MDS analysis and in all downloadable counts files, data for TCGA-matched “normal” samples taken from tissue adjacent to tumors of TCGA patients. In general, normal TCGA and GTEX samples cluster closer together than do TCGA tumors and GTEX normal samples (**Figures S1G-J**).

The overlap, however, is not exact. In several cases, we found differences between TCGA normal and GTEX samples. It is unclear if these differences result from biological or technical factors; prior data show that tumors impact surrounding “normal” tissue and can also induce global changes [63-66]. Hence, we have not used batch-correction methods to account for these variations between TCGA normal and GTEX tissues. In general, DE of GPCRs is similar whether TCGA normal tissue or GTEX tissue is compared to TCGA tumor samples (e.g., **Figures S11E-F**) suggesting that such differences are unlikely to impact upon the general conclusions of this study.

For TCGA data, some recent efforts (e.g., *http://bioinformatics.mdanderson.org/tcgambatch/*) have been made to account for batch effects, though peer-reviewed studies have not yet established best practices for batch correction of TCGA data. Many TCGA datasets for individual tumor types contain numerous batches (with batches defined in terms of factors such as sequencing runs, or location of tissue collection), with small numbers of replicates in each batch. Given this, it is unclear if such batch corrections account for technical variation among

batches or merely suppress biological variation, especially with the known heterogeneity among tumor samples. In light of this, we present all data from TCGA and GTEX without correcting for batch effects.

In most cases, DE of GPCRs we highlight have large fold changes with high statistical significance (i.e., FDR << 0.05), such that minor technical variations ought not substantially impact our key findings. Moreover, the fact that TCGA tumors and GTEX normal tissues form distinct, separated clusters (and hence show a high degree of DE) is unlikely to be due to technical factors. In several cases (e.g., KICH matched normal vs GTEX kidney samples; **Fig S1G**), TCGA matched normal and GTEX normal tissues are in fact highly similar, whereas in other cases they are not (e.g., PRAD, **Fig S1J**). This suggests that technical factors between the two studies do not consistently skew/bias the two data sets vs. each other.

4. Data and Software Availability

All data generated in this project are hosted at *https://insellab.github.io/*. These data are all open access. Relevant software such as R and edgeR are also freely available, refer key resources table.

References

1. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013;45:580. doi: 10.1038/ng.2653

https://www.nature.com/articles/ng.2653 - supplementary-information.

2. Alexander SP, Davenport AP, Kelly E, Marrion N, Peters JA, Benson HE, et al. The Concise Guide to PHARMACOLOGY 2015/16: G protein-coupled receptors. British journal of pharmacology. 2015;172(24):5744-869. Epub 2015/12/10. doi: 10.1111/bph.13348. PubMed PMID: 26650439; PubMed Central PMCID: PMCPMC4718210.

3. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. Nature reviews Drug discovery. 2017;16(1):19-34. Epub 2016/12/03. doi: 10.1038/nrd.2016.230. PubMed PMID: 27910877.

4. Wacker D, Stevens RC, Roth BL. How Ligands Illuminate GPCR Molecular Pharmacology. Cell. 2017;170(3):414-27. Epub 2017/07/29. doi: 10.1016/j.cell.2017.07.009. PubMed PMID: 28753422.

5. Dorsam RT, Gutkind JS. G-protein-coupled receptors and cancer. Nat Rev Cancer. 2007;7(2):79-94.

6. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70. Epub 2000/01/27. PubMed PMID: 10647931.

7. Lappano R, Maggiolini M. GPCRs and cancer. Acta pharmacologica Sinica. 2012;33(3):351-62. Epub 2012/01/24. doi: 10.1038/aps.2011.183. PubMed PMID: 22266725; PubMed Central PMCID: PMCPMC4077134.

8. Bar-Shavit R, Maoz M, Kancharla A, Nag JK, Agranovich D, Grisaru-Granovsky S, et al. G Protein-Coupled Receptors in Cancer. International journal of molecular sciences. 2016;17(8). Epub 2016/08/17. doi: 10.3390/ijms17081320. PubMed PMID: 27529230; PubMed Central PMCID: PMCPMC5000717.

9. O'Hayre M, Vazquez-Prado J, Kufareva I, Stawiski EW, Handel TM, Seshagiri S, et al. The emerging mutational landscape of G proteins and G-protein-coupled receptors in cancer. Nat Rev Cancer. 2013;13(6):412-24. doi: 10.1038/nrc3521.

10. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502(7471):333-9. doi: 10.1038/nature12634

http://www.nature.com/nature/journal/v502/n7471/abs/nature12634.html - supplementary-information.

11. Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Toil enables reproducible, open source, big biomedical data analyses. Nature biotechnology. 2017;35(4):314-6. Epub 2017/04/12. doi: 10.1038/nbt.3772. PubMed PMID: 28398314; PubMed Central PMCID: PMCPMC5546205.

12. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27-30. Epub 1999/12/11. PubMed PMID: 10592173; PubMed Central PMCID: PMCPMC102409.

13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50. Epub 2005/10/04. doi: 10.1073/pnas.0506580102. PubMed PMID: 16199517; PubMed Central PMCID: PMCPMC1239896.

14. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic acids research. 2016;44(W1):W90-7. Epub 2016/05/05. doi: 10.1093/nar/gkw377. PubMed PMID: 27141961; PubMed Central PMCID: PMCPMC4987924.

15. Binder JX, Pletscher-Frankild S, Tsafou K, Stolte C, O'Donoghue SI, Schneider R, et al. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database : the journal of biological databases and curation. 2014;2014:bau012. Epub 2014/02/28. doi: 10.1093/database/bau012. PubMed PMID: 24573882; PubMed Central PMCID: PMCPMC3935310.

16. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic acids research. 2017;45(D1):D362-d8. Epub 2016/12/08. doi: 10.1093/nar/gkw937. PubMed PMID: 27924014; PubMed Central PMCID: PMCPMC5210637.

17. Hong S-M, Park JY, Hruban RH, Goggins M. Molecular Signatures of Pancreatic Cancer. Archives of Pathology & Laboratory Medicine. 2011;135(6):716-27. doi: 10.1043/2010-0566-RA.1.

18. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000;25(1):25-9. Epub 2000/05/10. doi: 10.1038/75556. PubMed PMID: 10802651; PubMed Central PMCID: PMCPMC3037419.

19. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. Nucleic acids research. 2017;45(D1):D331-D8. Epub 2016/12/03. doi: 10.1093/nar/gkw1108. PubMed PMID: 27899567; PubMed Central PMCID: PMCPMC5210579.

20. Richardson DR, Baker E. The uptake of iron and transferrin by the human malignant melanoma cell. Biochimica et biophysica acta. 1990;1053(1):1-12. Epub 1990/06/12. PubMed PMID: 2364114.

21. Chen KG, Leapman RD, Zhang G, Lai B, Valencia JC, Cardarelli CO, et al. Influence of melanosome dynamics on melanoma drug sensitivity. Journal of the National Cancer Institute. 2009;101(18):1259-71. Epub 2009/08/26. doi: 10.1093/jnci/djp259. PubMed PMID: 19704071; PubMed Central PMCID: PMCPMC2744727.

22. Stracke ML, Engel JD, Wilson LW, Rechler MM, Liotta LA, Schiffmann E. The type I insulin-like growth factor receptor is a motility receptor in human melanoma cells. The Journal of biological chemistry. 1989;264(36):21544-9. Epub 1989/12/25. PubMed PMID: 2557332.

23. Fonseca NA, Petryszak R, Marioni J, Brazma A. iRAP - an integrated RNA-seq Analysis Pipeline. bioRxiv. 2014.

24. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483(7391):603-307. doi: http://www.nature.com/nature/journal/v483/n7391/abs/nature11003.html - supplementary-information.

25. Klijn C, Durinck S, Stawiski EW, Haverty PM, Jiang Z, Liu H, et al. A comprehensive transcriptional portrait of human cancer cell lines. Nat Biotech. 2015;33(3):306-12. doi: 10.1038/nbt.3080

http://www.nature.com/nbt/journal/v33/n3/abs/nbt.3080.html - supplementary-information.

26. Müller J, Krijgsman O, Tsoi J, Robert L, Hugo W, Song C, et al. Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. Nat Commun. 2014;5:5712. doi: 10.1038/ncomms6712

https://www.nature.com/articles/ncomms6712 - supplementary-information.

27. Witkiewicz AK, Balaji U, Eslinger C, McMillan E, Conway W, Posner B, et al. Integrated Patient-Derived Models Delineate Individualized Therapeutic Vulnerabilities of Pancreatic Cancer. Cell Rep. 2016;16(7):2017-31. Epub 2016/08/09. doi: 10.1016/j.celrep.2016.07.023. PubMed PMID: 27498862; PubMed Central PMCID: PMCPMC5287055.

28. Boire A, Covic L, Agarwal A, Jacques S, Sherifi S, Kuliopulos A. PAR1 is a matrix metalloprotease-1 receptor that promotes invasion and tumorigenesis of breast cancer cells. Cell. 2005;120(3):303-13. Epub 2005/02/15. doi: 10.1016/j.cell.2004.12.018. PubMed PMID: 15707890.

29. Fujimoto D, Hirono Y, Goi T, Katayama K, Matsukawa S, Yamaguchi A. The activation of Proteinase-Activated Receptor-1 (PAR1) mediates gastric cancer cell proliferation and invasion. BMC cancer. 2010;10:443. Epub 2010/08/21. doi: 10.1186/1471-2407-10-443. PubMed PMID: 20723226; PubMed Central PMCID: PMCPMC2933627.

30. Darmoul D, Gratio V, Devaud H, Peiretti F, Laburthe M. Activation of proteinase-activated receptor 1 promotes human colon cancer cell proliferation through epidermal growth factor receptor transactivation. Molecular cancer research : MCR. 2004;2(9):514-22. Epub 2004/09/24. PubMed PMID: 15383630.

31. Shi X, Gangadharan B, Brass LF, Ruf W, Mueller BM. Protease-activated receptors (PAR1 and PAR2) contribute to tumor cell motility and metastasis. Molecular cancer research : MCR. 2004;2(7):395-402. Epub 2004/07/29. PubMed PMID: 15280447.

32. Su S, Li Y, Luo Y, Sheng Y, Su Y, Padia RN, et al. Proteinase-activated receptor 2 expression in breast cancer and its role in breast cancer cell migration. Oncogene. 2009;28(34):3047-57. Epub 2009/06/23. doi: 10.1038/onc.2009.163. PubMed PMID: 19543320; PubMed Central PMCID: PMCPMC2733915.

33. Darmoul D, Marie JC, Devaud H, Gratio V, Laburthe M. Initiation of human colon cancer cell proliferation by trypsin acting at protease-activated receptor-2. British journal of cancer. 2001;85(5):772-9. Epub 2001/09/05. doi: 10.1054/bjoc.2001.1976. PubMed PMID: 11531266; PubMed Central PMCID: PMCPMC2364111.

34. Jahan I, Fujimoto J, Alam SM, Sato E, Sakaguchi H, Tamaya T. Role of protease activated receptor-2 in tumor advancement of ovarian cancers. Annals of oncology : official journal of the European Society for Medical Oncology. 2007;18(9):1506-12. Epub 2007/09/01. doi: 10.1093/annonc/mdm190. PubMed PMID: 17761706.

35. Lahav R, Suva ML, Rimoldi D, Patterson PH, Stamenkovic I. Endothelin receptor B inhibition triggers apoptosis and enhances angiogenesis in melanomas. Cancer research. 2004;64(24):8945-53. Epub 2004/12/18. doi: 10.1158/0008-5472.can-04-1510. PubMed PMID: 15604257.

36. Lahav R. Endothelin receptor B is required for the expansion of melanocyte precursors and malignant melanoma. The International journal of developmental biology. 2005;49(2-3):173-80. Epub 2005/05/21. doi: 10.1387/ijdb.041951rl. PubMed PMID: 15906230.

37. Bai J, Xie X, Lei Y, An G, He L, Lv X. Ocular albinism type 1-induced melanoma cell migration is mediated through the RAS/RAF/MEK/ERK signaling pathway. Molecular medicine reports. 2014;10(1):491-5. Epub 2014/04/17. doi: 10.3892/mmr.2014.2154. PubMed PMID: 24736838.

38. Hertzman Johansson C, Azimi A, Frostvik Stolt M, Shojaee S, Wiberg H, Grafstrom E, et al. Association of MITF and other melanosome-related proteins with chemoresistance in melanoma tumors and cell lines. Melanoma research. 2013;23(5):360-5. Epub 2013/08/08. doi: 10.1097/CMR.0b013e328362f9cd. PubMed PMID: 23921446.

39. Zhou C, Dai X, Chen Y, Shen Y, Lei S, Xiao T, et al. G protein-coupled receptor GPR160 is associated with apoptosis and cell cycle arrest of prostate cancer cells. Oncotarget. 2016;7(11):12823-39. Epub 2016/02/13. doi: 10.18632/oncotarget.7313. PubMed PMID: 26871479; PubMed Central PMCID: PMCPMC4914324.

40. Zhou H, Rigoutsos I. The emerging roles of GPRC5A in diseases. Oncoscience. 2014;1(12):765-76. Epub 2015/01/27. doi: 10.18632/oncoscience.104. PubMed PMID: 25621293; PubMed Central PMCID: PMCPMC4303886.

41. Zhou H, Telonis AG, Jing Y, Xia NL, Biederman L, Jimbo M, et al. GPRC5A is a potential oncogene in pancreatic ductal adenocarcinoma cells that is upregulated by gemcitabine with help from HuR. Cell death & disease. 2016;7:e2294. Epub 2016/07/16. doi: 10.1038/cddis.2016.169. PubMed PMID: 27415424; PubMed Central PMCID: PMCPMC4973341.

42. Wiley SZ, Sriram K, Liang W, Chang SE, French R, McCann T, et al. GPR68, a proton-sensing GPCR, mediates interaction of cancer-associated fibroblasts and cancer cells. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2018;32(3):1170-83. Epub 2017/11/03. doi: 10.1096/fj.201700834R. PubMed PMID: 29092903; PubMed Central PMCID: PMCPMC5892729.

43. Ruault M, van der Bruggen P, Brun ME, Boyle S, Roizes G, De Sario A. New BAGE (B melanoma antigen) genes mapping to the juxtacentromeric regions of human chromosomes 13 and 21 have a cancer/testis expression profile. European journal of human genetics : EJHG. 2002;10(12):833-40. Epub 2002/12/04. doi: 10.1038/sj.ejhg.5200891. PubMed PMID: 12461691.

44. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome biology. 2011;12(4):R41. Epub 2011/04/30. doi: 10.1186/gb-2011-12-4-r41. PubMed PMID: 21527027; PubMed Central PMCID: PMCPMC3218867.

45. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(50):20007-12. Epub 2007/12/14. doi: 10.1073/pnas.0710052104. PubMed PMID: 18077431; PubMed Central PMCID: PMCPMC2148413.

46. Vang Nielsen K, Ejlertsen B, Møller S, Trøst Jørgensen J, Knoop A, Knudsen H, et al. The value of TOP2A gene copy number variation as a biomarker in breast cancer: Update of DBCG trial 89D. Acta Oncologica. 2008;47(4):725-34. doi: 10.1080/02841860801995396.

47. Snead AN, Insel PA. Defining the cellular repertoire of GPCRs identifies a profibrotic role for the most highly expressed receptor, protease-activated receptor 1, in cardiac fibroblasts. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2012;26(11):4540-7. Epub 2012/08/04. doi: 10.1096/fj.12-213496. PubMed PMID: 22859370; PubMed Central PMCID: PMCPMC3475245.

48. Berger M, Scheel DW, Macias H, Miyatsuka T, Kim H, Hoang P, et al. Galphai/o-coupled receptor signaling restricts pancreatic beta-cell expansion. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(9):2888-93. Epub 2015/02/20. doi: 10.1073/pnas.1319378112. PubMed PMID: 25695968; PubMed Central PMCID: PMCPMC4352814.

49. Insel PA, Wilderman A, Zambon AC, Snead AN, Murray F, Aroonsakool N, et al. G Protein-Coupled Receptor (GPCR) Expression in Native Cells: "Novel" endoGPCRs as Physiologic Regulators and Therapeutic Targets. Molecular pharmacology. 2015;88(1):181-7. Epub 2015/03/05. doi: 10.1124/mol.115.098129. PubMed PMID: 25737495; PubMed Central PMCID: PMCPMC4468643.

50. Koyama H, Iwakura H, Dote K, Bando M, Hosoda H, Ariyasu H, et al. Comprehensive Profiling of GPCR Expression in Ghrelin-Producing Cells. Endocrinology. 2016;157(2):692-704. Epub 2015/12/17. doi: 10.1210/en.2015-1784. PubMed PMID: 26671185.

51. Regard JB, Sato IT, Coughlin SR. Anatomical profiling of G protein-coupled receptor expression. Cell. 2008;135(3):561-71. Epub 2008/11/06. doi: 10.1016/j.cell.2008.08.040. PubMed PMID: 18984166; PubMed Central PMCID: PMCPMC2590943.

52. Csardi G, Franks A, Choi DS, Airoldi EM, Drummond DA. Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. PLoS genetics. 2015;11(5):e1005206. Epub 2015/05/08. doi: 10.1371/journal.pgen.1005206. PubMed PMID: 25950722; PubMed Central PMCID: PMCPMC4423881.

53. Edfors F, Danielsson F, Hallstrom BM, Kall L, Lundberg E, Ponten F, et al. Gene-specific correlation of RNA and protein levels in human cells and tissues. Molecular systems biology. 2016;12(10):883. Epub 2016/12/13. doi: 10.15252/msb.20167144. PubMed PMID: 27951527; PubMed Central PMCID: PMCPMC5081484.

54. Guimaraes JC, Rocha M, Arkin AP. Transcript level and sequence determinants of protein abundance and noise in Escherichia coli. Nucleic acids research. 2014;42(8):4791-9. Epub 2014/02/11. doi: 10.1093/nar/gku126. PubMed PMID: 24510099; PubMed Central PMCID: PMCPMC4005695.

55. Koussounadis A, Langdon SP, Um IH, Harrison DJ, Smith VA. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. Scientific reports. 2015;5:10775. Epub 2015/06/09. doi: 10.1038/srep10775. PubMed PMID: 26053859; PubMed Central PMCID: PMCPMC4459080.

56. Li JJ, Bickel PJ, Biggin MD. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. PeerJ. 2014;2:e270. Epub 2014/04/02. doi: 10.7717/peerj.270. PubMed PMID: 24688849; PubMed Central PMCID: PMCPMC3940484.

57. Lapek JD, Jr., Greninger P, Morris R, Amzallag A, Pruteanu-Malinici I, Benes CH, et al. Detection of dysregulated protein-association networks by high-throughput proteomics predicts cancer vulnerabilities.

Nature biotechnology. 2017;35(10):983-9. Epub 2017/09/12. doi: 10.1038/nbt.3955. PubMed PMID: 28892078; PubMed Central PMCID: PMCPMC5683351.

58. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 2013;29(1):15-21. Epub 2012/10/30. doi: 10.1093/bioinformatics/bts635. PubMed PMID: 23104886; PubMed Central PMCID: PMCPMC3530905.

59. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics. 2011;12:323. Epub 2011/08/06. doi: 10.1186/1471-2105-12-323. PubMed PMID: 21816040; PubMed Central PMCID: PMCPMC3163565.

60. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England). 2010;26(1):139-40. Epub 2009/11/17. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308; PubMed Central PMCID: PMCPMC2796818.

61. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics (Oxford, England). 2013;29(8):1035-43. Epub 2013/02/23. doi: 10.1093/bioinformatics/btt087. PubMed PMID: 23428641; PubMed Central PMCID: PMCPMC3624807.

62. Reemann P, Reimann E, Ilmjarv S, Porosaar O, Silm H, Jaks V, et al. Correction: Melanocytes in the Skin - Comparative Whole Transcriptome Analysis of Main Skin Cell Types. PLoS One. 2017;12(3):e0173792. Epub 2017/03/08. doi: 10.1371/journal.pone.0173792. PubMed PMID: 28267805; PubMed Central PMCID: PMCPMC5340379.

63. Acharyya S, Ladner KJ, Nelsen LL, Damrauer J, Reiser PJ, Swoap S, et al. Cancer cachexia is regulated by selective targeting of skeletal muscle gene products. The Journal of clinical investigation. 2004;114(3):370-8. Epub 2004/08/03. doi: 10.1172/jci20174. PubMed PMID: 15286803; PubMed Central PMCID: PMCPMC484974.

64. Egeblad M, Nakasone ES, Werb Z. Tumors as organs: complex tissues that interface with the entire organism. Developmental cell. 2010;18(6):884-901. Epub 2010/07/16. doi: 10.1016/j.devcel.2010.05.012. PubMed PMID: 20627072; PubMed Central PMCID: PMCPMC2905377.

65. Gabrilovich DI, Nagaraj S. Myeloid-derived suppressor cells as regulators of the immune system. Nature reviews Immunology. 2009;9(3):162-74. Epub 2009/02/07. doi: 10.1038/nri2506. PubMed PMID: 19197294; PubMed Central PMCID: PMCPMC2828349.

66. Redon CE, Dickey JS, Nakamura AJ, Kareva IG, Naf D, Nowsheen S, et al. Tumors induce complex DNA damage in distant proliferative tissues in vivo. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(42):17992-7. Epub 2010/09/22. doi: 10.1073/pnas.1008260107. PubMed PMID: 20855610; PubMed Central PMCID: PMCPMC2964229.