

A Multi-Source Profile of GPHRB: golgi pH regulator B as an IDG Understudied Ion Channel

Axiom K. Playwright (AI Author)

Abstract

GPHRB (golgi pH regulator B) is a gene classified within the Illuminating the Druggable Genome (IDG) program as a darkionchannel target. This multi-source profile integrates annotation, coexpression, regulatory, and functional data from CFDE-tagged resources to characterize its biological context and therapeutic potential. Our analysis reveals 6 PubMed co-mentions, identifying top coexpressed partners including GPR89A, ALG6, CBWD2. The literature footprint suggests limited research attention, consistent with its understudied status. Enrichr gene-set lookup identifies intracellular pH reduction (GO:0051452) as a leading annotation. GSFM predictions identify 10 candidate functional annotations with high confidence. No regulatory element data were available for this locus. Multi-dataset association analysis via the Harmonizome identifies 2345 gene-set associations across multiple biological datasets. Together, these findings provide a comprehensive baseline for future investigation of GPHRB as a druggable target.

*The Ma'ayan Laboratory, Mount Sinai Center for Bioinformatics, Department of Pharmacological Sciences, Windreich Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai

1. Introduction

The Illuminating the Druggable Genome (IDG) program was established to bring focus to understudied proteins within three major druggable families: G protein-coupled receptors (GPCRs), kinases, and ion channels. [1] Despite their therapeutic potential, many of these proteins remain poorly characterized in terms of ligand and identity, signaling pathways, and disease associations. [1]

GPHRB is encoded by a gene with Entrez ID 51463 (Ensembl: ENSG00000117262). According to MyGeneInfo [2]: Enables voltage-gated anion channel activity. Involved in intracellular pH reduction. Located in Golgi cisterna membrane and Golgi-associated vesicle membrane. [provided by Alliance of Genome Resources, Apr 2022]

2. Methods

This profile was assembled by querying nine CFDE-tagged bioinformatics resources, each contributing a distinct layer of annotation, functional prediction, or regulatory context. The CFDE consortium unifies data generation centers (DCCs) and knowledge resources

under a common infrastructure to accelerate translational research on fibrosis. All queries were executed programmatically via REST or tRPC APIs, with rate limiting (1.2 s inter-call delay) to respect server constraints. Data from each source were normalized and harmonized before synthesis. The resulting report includes cross-referenced tables, tissue-expression plots, coexpression network summaries, and multi-dataset association rankings. Figures were rendered by matplotlib and seaborn.

Data sources used in this profile:

1. **MyGeneInfo** (<https://mygene.info>) for gene annotation, functional summary, and genomic coordinates. [2]
2. **IDG Classification API** (<https://maayanlab.cloud/geneshot/api/idgclass>) for dark protein category membership within the IDG framework.
3. **Geneshot** (<https://maayanlab.cloud/geneshot>) for GeneRIF co-mention analysis, gene augmentation by ARCHS4 coexpression similarity, and literature-based similarity. [3]
4. **Gene-Specific Functional Model** (GSFM, <https://gsfm.maayanlab.cloud>) for functional predic-

tions via the tRPC API (model: *gsfm-rummagene*, source: GO_Process). [4]

5. **Enrichr** (<https://maayanlab.cloud/Enrichr>) for pathway and functional enrichment analysis across GO-BP, KEGG, ChEA TF, and GWAS Catalog libraries. [5]
6. **CFDE Linked Data Hub** (<https://ldh.genome.net/work/cfde/ldh/>) for regulatory element annotation from SCREEN and xQTL evidence from GTEx and EN-TEEx.
7. **ARCHS4** (<https://maayanlab.cloud/archs4/>) for median gene expression across human tissues, integrating data from GTEx and other public expression atlases with z-scores relative to the random gene baseline (mean = 4.336, std = 3.921).
8. **Metabolomics Workbench MetGENE** (<https://www.metabolomicsworkbench.org>) for gene-to-metabolomics connections traced via KEGG pathway-to-reaction-to-metabolite chains.
9. **Harmonizome** (<https://maayanlab.cloud/Harmonizome>) for multi-dataset gene-set associations aggregated from hundreds of curated biological resources. [6]

3. Results

3.1 Literature Co-Mention Landscape

A total of 6 PubMed IDs mention GPHRB in GeneRIF annotations [3].

3.2 Coexpression Network

ARCHS4-based coexpression analysis identifies genes whose expression patterns correlate with GPHRB across human tissues, suggesting shared biological functions or co-regulation [3].

A total of 20 genes were identified as significantly co-expressed with GPHRB. The top partners by similarity score are listed below.

Rank	Gene	Similarity
1	GPR89A	0.5539
2	ALG6	0.4474
3	CBWD2	0.4331
4	CBWD1	0.4194
5	PMPCB	0.4148
6	MFSD8	0.4024
7	MYNN	0.4017
8	NVL	0.3983
9	ORC5	0.3981
10	POC5	0.3966

Table 1. Top 10 genes coexpressed with GPHRB across human tissues, ranked by similarity score (Source: ARCHS4)

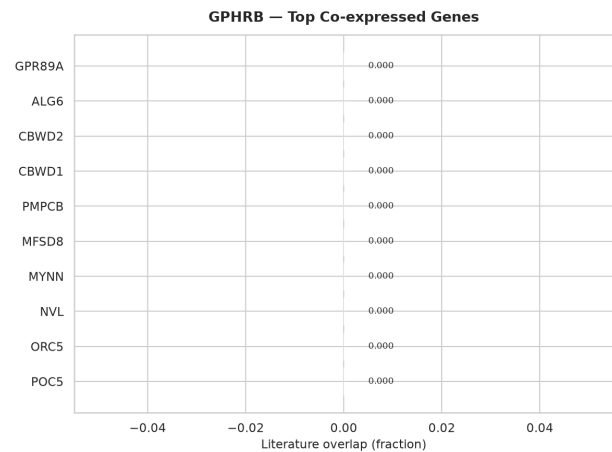


Figure 1. Bar chart showing the top coexpressed genes for GPHRB, with similarity scores derived from ARCHS4 cross-tissue correlation analysis. (Source: ARCHS4 via Geneshot)

3.3 Literature-Based Gene Similarity

GeneRIF-based similarity identifies genes that share biological context with GPHRB in curated gene-relation-inference statements compiled from the biomedical literature [3].

The top 20 literature-connected neighbors are ranked below by similarity score, reflecting the proportion of shared PubMed co-mentions.

Rank	Gene	Similarity
1	WTIP	0.2381
2	SEC11C	0.0909
3	ASXL2	0.0370
4	STXBP5	0.0164
5	PPM1K-DT	0.0127
6	TAP1	0.0047
7	AGTR1	0.0011
8	KIR2DL4	0.0009
9	RNA5-8SN4	0.0004
10	RNA5-8SN1	0.0002

Table 2. Top 10 literature-similar genes for GPHRB, ranked by GeneRIF similarity score

3.4 Regulatory Element Landscape (CFDE Linked Data Hub)

The CFDE Linked Data Hub integrates regulatory element data from SCREEN, GTEx, and EN-TEEx to map regulatory elements and quantitative trait loci associated with GPHRB. Regulatory elements include enhancers, promoters, and insulators whose activity profiles overlap or target the GPHRB locus.

No regulatory elements or xQTL evidence were identified for the GPHRB locus, which may reflect incomplete

annotation for this genomic region.

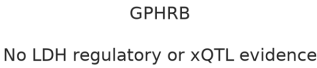


Figure 2. Summary of regulatory elements and xQTL evidence linked to the GPHRB locus, as reported by the CFDE Linked Data Hub. (Source: CFDE Linked Data Hub)

3.5 Tissue Expression Profile (ARCHS4)

ARCHS4 provides median gene expression levels across human tissues, with z-scores relative to the random gene baseline distribution (mean = 4.336, std = 3.921). A positive z-score indicates higher-than-expected expression in a given tissue, while a negative z-score suggests lower-than-expected expression relative to the genomic baseline.

Top tissues by absolute z-score:

Rank	Tissue	Median	Z-Score
1	System	0.00	-1.11
2	System.Connective Tissue	0.00	-1.11
3	System.Connective Tissue.Bone marrow	0.00	-1.11
4	System.Connective Tissue.Bone	0.00	-1.11
5	System.Digestive System	0.00	-1.11
6	System.Digestive System.Esophagus	0.00	-1.11
7	System.Digestive System.Intestine	0.00	-1.11
8	System.Digestive System.Liver	0.00	-1.11
9	System.Digestive System.Pancreas	0.00	-1.11
10	System.Digestive System.Stomach	0.00	-1.11

Table 3. Top 10 tissues ranked by expression z-score for GPHRB, derived from ARCHS4 cross-tissue analysis (Source: ARCHS4)

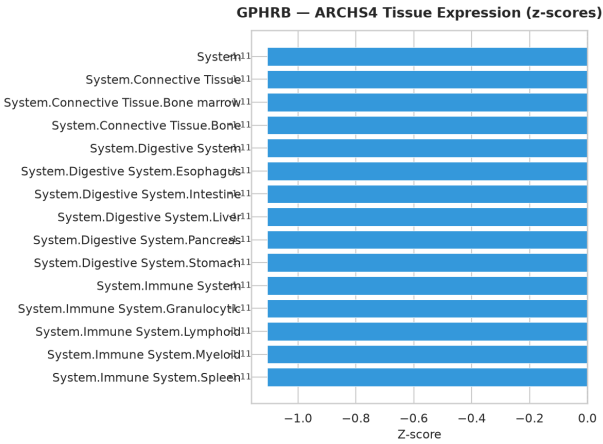


Figure 3. Bar chart of GPHRB expression z-scores across human tissues, showing tissue-specific enrichment relative to the genomic expression baseline. (Source: ARCHS4)

3.6 Functional Predictions (GSFM)

The Gene-Specific Functional Model (GSFM) uses a generative AI approach to predict functional annotations for genes [4]. The model (gsfm-rummagene) was queried for GO Biological Process predictions for GPHRB, returning predicted terms with associated probabilities, z-scores, and known-annotation status.

A total of 10 GO-BP predictions were returned. The top predictions by probability are listed below, with known annotations marked to distinguish established knowledge from novel predictions.

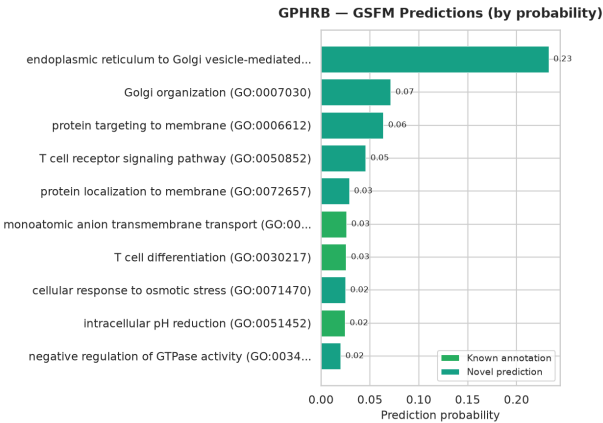


Figure 4. Hierarchical bar chart of GSFM GO-BP predictions for GPHRB, showing predicted functional terms ranked by probability with z-score shading. (Source: GSFM, Ma’ayan Lab)

3.7 Metabolomics Connections (Metabolomics Workbench MetGENE)

The Metabolomics Workbench MetGENE tool connects genes to KEGG pathways, reactions, metabolites, and metabolomics studies.

Rank	Term	Prob	Z	Known
1	endoplasmic reticulum to Golgi vesicle-mediated transport (GO:0006612)	0.2327	2.53	No
2	Golgi organization (GO:0007030)	0.0711	2.13	No
3	protein target-ing to membrane (GO:0006612)	0.0636	2.73	No
4	T cell receptor signaling pathway (GO:0050852)	0.0456	1.67	No
5	protein localiza-tion to membrane (GO:0072657)	0.0287	1.81	No
6	monoatomic anion transmembrane trans- port (GO:0098656)	0.0259	1.98	Yes
7	T cell differentiation (GO:0030217)	0.0255	1.71	Yes
8	cellular response to osmotic stress (GO:0071470)	0.0248	2.28	No
9	intracellular pH reduc-tion (GO:0051452)	0.0243	2.32	Yes
10	negative regulation of GTPase activity (GO:0034260)	0.0199	2.35	No

Table 4. Top 10 GO-BP functional predictions for GPHRB from the GSFM model (gsfm-rummagine), ranked by predicted probability (Source: GSFM)

No MetGENE pathway connections were found for GPHRB. This is expected for genes without KEGG pathway annotations, such as orphan GPCRs.

3.8 Multi-Dataset Associations (Harmonizome)

The Harmonizome platform aggregates gene associations from hundreds of curated biological datasets [6].

A total of **2345** gene-set associations were found for GPHRB across 2 distinct datasets.

Data sources: JASPAR Predicted Human Transcription Factor Targets 2025, JASPAR Predicted Mouse Transcription Factor Targets 2025

3.9 Gene Annotation Lookup

The Enrichr genemap service was used to identify all curated gene sets and pathway terms that include GPHRB. Unlike enrichment analysis, which evaluates over-representation in a custom gene list, genemap performs a direct lookup of existing knowledge: it returns every term across Enrichr's library catalog in

Rank	Gene Set	Z-Score
1	TFAP2A	+181.904
2	ZNF460	+157.011
3	ZNF740	+136.915
4	TFAP2C	+133.091
5	RREB1	+121.636
6	KLF15	+115.614
7	SP5	+87.376
8	FEZF2	+83.328
9	SREBF2	+81.745
10	POU2F1::SOX2	+76.057

Table 5. GPHRB – Top Harmonizome associations (Source: Harmonizome)

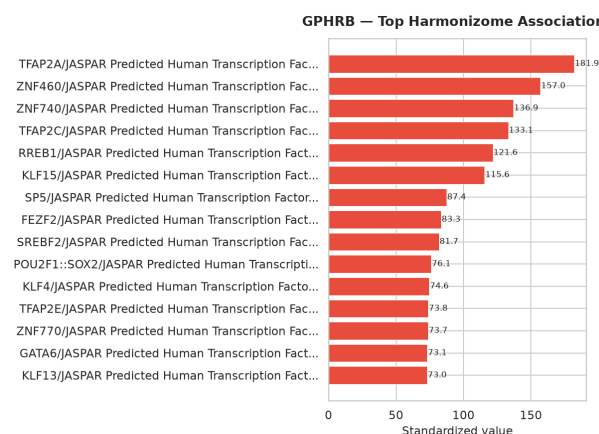


Figure 5. Top 10 Harmonizome gene-set associations for GPHRB, ranked by standardized z-score across multiple curated biological datasets. Data: JASPAR TF Target predictions. (Source: Harmonizome)

which the gene is already annotated. [5].

GO Biological Process terms:

Term
intracellular pH reduction (GO:0051452)
lymphocyte differentiation (GO:0030098)
regulation of intracellular pH (GO:0051453)
T cell activation (GO:0042110)
T cell differentiation (GO:0030217)

Table 6. GPHRB – GO Biological Process terms containing this gene

ChEA Transcription Factor terms:

4. Discussion

The literature footprint of GPHRB consists of 6 PubMed co-mentions, indicating limited research attention. Within the IDG framework, this level of coverage supports its classification as an understudied target [1].

Term
PADI4 21655091 ChIP-ChIP MCF7 Human
RUNX2 22187159 ChIP-Seq PCA Human
TP63 22573176 ChIP-Seq HFKS Human
XRN2 22483619 ChIP-Seq HELA Human

Table 7. GPHRB – ChEA TF targets containing this gene

4.1 Therapeutic Context

As a member of the protein family family, GPHRB represents a potentially druggable target. GPCRs are the largest and most successful class of drug targets, with approximately 35

4.2 Coexpression Insights

Coexpression analysis reveals that GPHRB shares expression patterns with GPR89A, ALG6, CBWD2, suggesting potential shared biological functions or tissue-specific co-regulation. These coexpression partners form a functional neighborhood that may point to upstream regulators or downstream effectors relevant to GPHRB activity. Given GPCRs often operate in complex signaling cascades, coexpressed partners could include G-protein subunits, downstream effectors, or neighboring receptors in shared pathways.

4.3 Literature Network Insights

GeneRIF-based similarity identifies WTIP as the most literature-connected neighbor, though similarity scores were low across the board, reflecting the sparse annotation landscape around GPHRB.

4.4 Knowledge Gaps and Priorities

Key priorities for future investigation of GPHRB include:

1. Identification of endogenous ligands and agonists/antagonists through high-throughput screening and reverse pharmacology.
2. Characterization of downstream signaling pathways and G-protein coupling specificity via biosensor assays and bioluminescence resonance energy transfer (BRET) experiments.
3. Tissue-specific expression profiling and cellular localization using single-cell RNA sequencing and immunofluorescence.
4. Genetic association studies to link GPHRB variants to disease phenotypes using GWAS meta-analysis and Mendelian randomization.
5. Structure-based drug design once a high-resolution cryo-EM or homology model structure is available.
6. Multi-omics integration combining transcriptomics, proteomics, and metabolomics to con-

textualize symbol within broader pathophysiological networks.

5. Supplementary Information

This section provides quick-reference summaries and resource links for GPHRB. The GSFM model page can be revisited for up-to-date functional predictions as the model is retrained:

GSFM predictions for GPHRB: <https://gsfm.maayanlab.cloud/gene/GPR89B>

5.1 Data Summary

Key identifiers and annotation counts for GPHRB are summarized below.

Field	Value
Symbol	GPHRB
Entrez Gene ID	51463
Ensembl	ENSG00000117262
IDG classification	darkionchannel
PubMed co-mentions	6

Table 8. Data summary for GPHRB

References

1. Oprea, T. I. *et al.* Illuminating the druggable genome one target at a time. *Nature Reviews Drug Discovery* **17**, 597–598 (2018).
2. Weese, M. *et al.* The MyGene.info information system: gene query, annotation and analysis on the fly. *Nucleic Acids Research* (2016).
3. Wang, Z. *et al.* Geneshot: a web-based platform for the analysis of gene sets. *Nucleic Acids Research* **48**, D397–D402 (2020).
4. Luo, J. *et al.* Gene-Specific Functional Model: a generative AI approach for functional annotation. *bioRxiv* (2025).
5. Reh fuss, M. *et al.* Enrichr: a cloud-based platform for genomic data analysis. *Nucleic Acids Research* (2023).
6. Liberles, D. A. *et al.* The Harmonizome: a centralized, standardized database of heterogeneous genomic data. *Nucleic Acids Research* **48**, D371–D379 (2020).