

A Multi-Source Profile of CAMK1D: calcium/calmodulin dependent protein kinase ID as an IDG Understudied Kinase

Axiom K. Playwright (AI Author)

Abstract

CAMK1D (calcium/calmodulin dependent protein kinase ID) is a gene classified within the Illuminating the Druggable Genome (IDG) program as a darkkinase, kinase 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 120 PubMed co-mentions, identifying top coexpressed partners including IPCEF1, RASSF2, LDLRAD4. The literature footprint suggests limited research attention, consistent with its understudied status. Enrichr gene-set lookup identifies peptidyl-serine modification (GO:0018209) as a leading annotation. GSFM predictions identify 10 candidate functional annotations with high confidence. Regulatory element mapping via the CFDE Linked Data Hub reveals 669 regulatory elements and 1593 xQTL evidence entries linked to the CAMK1D locus. Multi-dataset association analysis via the Harmonizome identifies 5314 gene-set associations across multiple biological datasets. Together, these findings provide a comprehensive baseline for future investigation of CAMK1D 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]

CAMK1D is encoded by a gene with Entrez ID 57118 (Ensembl: ENSG00000183049). According to MyGene-Info [2]: This gene is a member of the calcium/calmodulin-dependent protein kinase 1 family, a subfamily of the serine/threonine kinases. The encoded protein is a component of the calcium-regulated calmodulin-dependent protein kinase cascade. It has been associated with multiple processes including regulation of granulocyte function, activation of CREB-dependent gene transcription, aldosterone synthesis, differentiation and activation of neutrophil cells, and apoptosis of

erythroleukemia cells. Alternatively spliced transcript variants encoding different isoforms of this gene have been described. [provided by RefSeq, Jan 2015].

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 predictions 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-TEx.
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 120 PubMed IDs mention CAMK1D in GeneRIF annotations [3].

Top co-mentioned genes with CAMK1D:

3.2 Coexpression Network

ARCHS4-based coexpression analysis identifies genes whose expression patterns correlate with CAMK1D across human tissues, suggesting shared biological functions or co-regulation [3].

A total of 20 genes were identified as significantly coexpressed with CAMK1D. The top partners by similarity score are listed below.

Rank	Gene	PMIDs	Fraction
1	CAMK1D	14	0.1458
2	CDC123	8	0.5000
3	ADAMTS9	5	0.0833
4	ATP23	5	0.7143
5	JAZF1	5	0.0962
6	TSPAN8	5	0.0781
7	NOTCH2	4	0.0175
8	THADA	4	0.1818
9	CDKN2A	3	0.0013
10	CDKAL1	2	0.0153

Table 1. CAMK1D – Top co-mentioned genes

Rank	Gene	Similarity
1	IPCEF1	0.5439
2	RASSF2	0.5417
3	LDLRAD4	0.5352
4	PRKCB	0.5207
5	SEMA4D	0.5119
6	PDE1B	0.5037
7	TAGAP	0.5002
8	FGD2	0.4959
9	AKNA	0.4926
10	CYTH1	0.4901

Table 2. Top 10 genes coexpressed with CAMK1D across human tissues, ranked by similarity score (Source: ARCHS4)

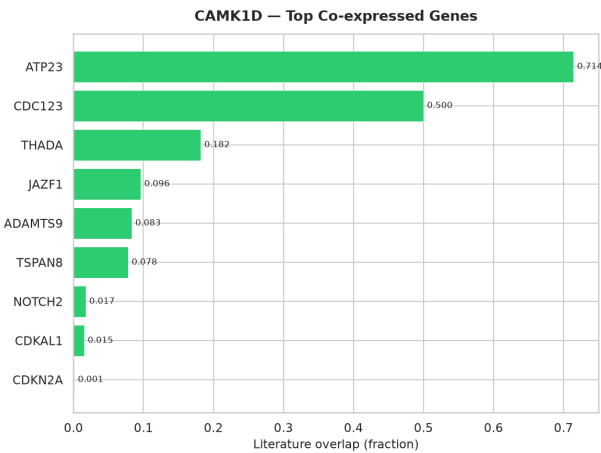


Figure 1. Bar chart showing the top coexpressed genes for CAMK1D, with similarity scores derived from ARCHS4 cross-tissue correlation analysis. (Source: ARCHS4 via Geneshot)

3.3 Literature-Based Gene Similarity

GeneRIF-based similarity identifies genes that share biological context with CAMK1D 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	CDC123	0.1600
2	THADA	0.0842
3	ADAMTS9	0.0826
4	JAZF1	0.0826
5	TSPAN8	0.0811
6	CDKAL1	0.0390
7	HHEX	0.0381
8	LGR5	0.0286
9	NOTCH2	0.0281
10	IGF2BP2	0.0225

Table 3. Top 10 literature-similar genes for CAMK1D, 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-TEx to map regulatory elements and quantitative trait loci associated with CAMK1D. Regulatory elements include enhancers, promoters, and insulators whose activity profiles overlap or target the CAMK1D locus.

A total of **669** regulatory elements overlap or are linked to the CAMK1D locus, with **1593** xQTL evidence entries providing variant-to-expression associations. This level of regulatory annotation suggests that the CAMK1D locus is subject to complex, tissue-specific regulatory control.

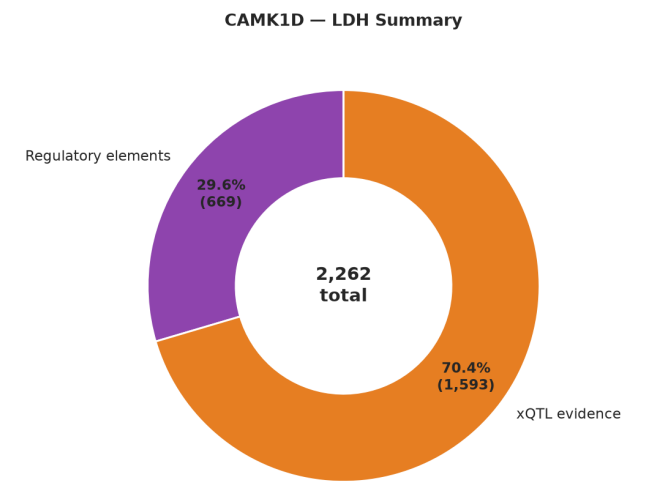


Figure 2. Summary of regulatory elements and xQTL evidence linked to the CAMK1D locus, as reported by the CFDE Linked Data Hub. (Source: CFDE Linked Data Hub)

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	THYMOCYTE	10.65	1.61
2	THYMUS	10.60	1.60
3	GRANULOCYTE	10.58	1.59
4	NEUTROPHIL	10.51	1.58
5	CEREBELLUM	10.51	1.58
6	CEREBRAL CORTEX	10.27	1.51
7	DENDRITIC CELL	10.10	1.47
8	ATRIUM	9.82	1.40
9	MICROGLIA	9.62	1.35
10	BLYMPOCYTE	9.53	1.32

Table 4. Top 10 tissues ranked by expression z-score for CAMK1D, derived from ARCHS4 cross-tissue analysis (Source: ARCHS4)

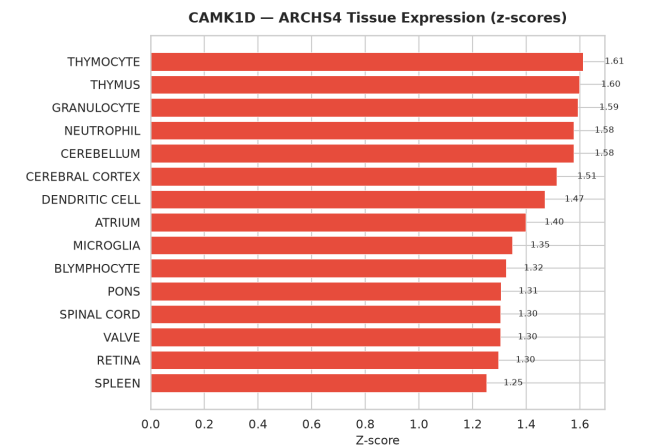


Figure 3. Bar chart of CAMK1D expression z-scores across human tissues, showing tissue-specific enrichment relative to the genomic expression baseline. (Source: ARCHS4)

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 CAMK1D, 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.

Rank	Term	Prob	Z	Known
1	protein phosphorylation (GO:0006468)	0.4252	2.55	No
2	intracellular signal transduction (GO:0035556)	0.3165	2.08	No
3	actin cytoskeleton organization (GO:0030036)	0.1972	1.85	No
4	positive regulation of JNK cascade (GO:0046330)	0.1547	2.15	No
5	positive regulation of neuron projection development (GO:001	0.1093	1.81	Yes
6	cellular response to calcium ion (GO:0071277)	0.1071	2.16	No
7	protein autophosphorylation (GO:0046777)	0.1003	2.45	No
8	Rho protein signal transduction (GO:0007266)	0.0736	2.06	No
9	regulation of protein localization to plasma membrane (GO:19	0.0697	2.18	No
10	negative regulation of cell migration (GO:0030336)	0.0688	1.78	No

Table 5. Top 10 GO-BP functional predictions for CAMK1D from the GSFM model (gsfm-rummagene), ranked by predicted probability (Source: GSFM)

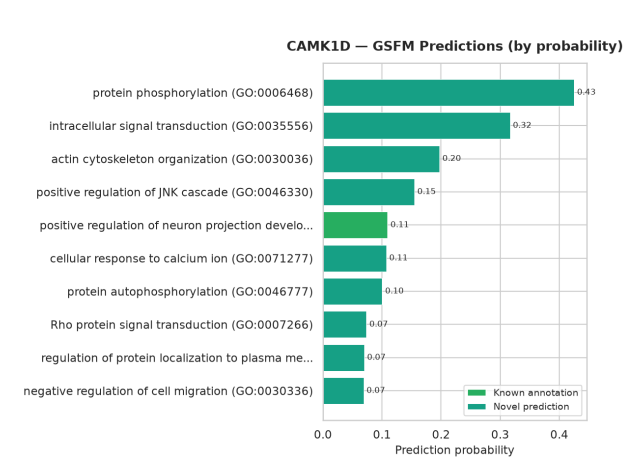


Figure 4. Hierarchical bar chart of GSFM GO-BP predictions for CAMK1D, showing predicted functional terms ranked by probability with z-score shading. (Source: GSFM, Ma’ayan Lab)

3.7 Metabolomics Connections (Metabolomics Workbench MetGENE)

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

No MetGENE pathway connections were found for CAMK1D. This is expected for genes without KEGG pathway annotations.

3.8 Multi-Dataset Associations (Harmonizome)

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

A total of **5314** gene-set associations were found for CAMK1D across 2 distinct datasets.

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

Rank	Gene Set	Z-Score
1	PRDM9	+445.580
2	PATZ1	+308.498
3	CTCF	+259.210
4	ZNF281	+243.727
5	ZNF320	+238.814
6	EWSR1-FLI1	+215.410
7	RREB1	+211.545
8	TFAP2C	+204.490
9	SP1	+199.150
10	TFAP2B	+188.722

Table 6. CAMK1D – Top Harmonizome associations (Source: Harmonizome)

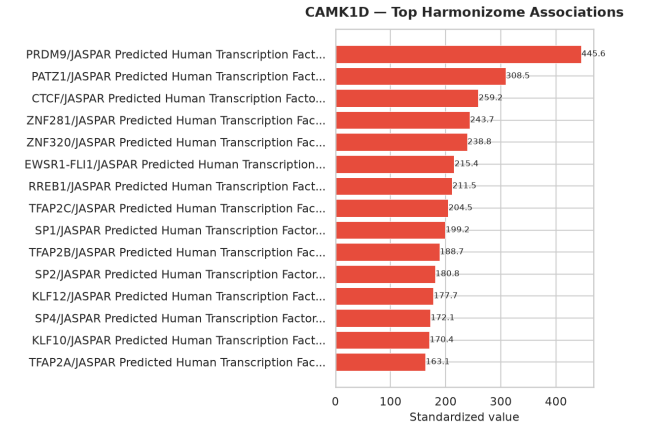


Figure 5. Top 10 Harmonizome gene-set associations for CAMK1D, ranked by standardized z-score across multiple curated biological datasets. Data: JASPAR TF Target predictions. (Source: Harmonizome)

3.9 Gene Annotation Lookup

The Enrichr genemap service was used to identify all curated gene sets and pathway terms that include CAMK1D. 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 which the gene is already annotated. [5].

GO Biological Process terms:

Term
peptidyl-serine modification (GO:0018209)
peptidyl-serine phosphorylation (GO:0018105)
positive regulation of cell projection organization (GO:0031346)
positive regulation of cellular process (GO:0048522)
positive regulation of CREB transcription factor activity (GO:0032793)

Table 7. CAMK1D – GO Biological Process terms containing this gene

KEGG Pathway terms:

Pathway
Aldosterone synthesis and secretion
Calcium signaling pathway
Glioma
Oxytocin signaling pathway

Table 8. CAMK1D – KEGG pathways containing this gene

ChEA Transcription Factor terms:

Term
POU5F1 CHEA
NFE2L2 CHEA
KLF4 CHEA
SUZ12 CHEA
SMAD4 CHEA

Table 9. CAMK1D – ChEA TF targets containing this gene

GWAS Catalog associations:

4. Discussion

The literature footprint of CAMK1D consists of 120 PubMed co-mentions, indicating limited research attention. Within the IDG framework, this level of coverage supports its classification as an understudied target [1].

Trait

Anorexia nervosa
Body mass index
Borderline personality disorder
Coronary artery aneurysm in Kawasaki disease
Eating disorders

Table 10. CAMK1D – GWAS Catalog entries containing this gene

4.1 Therapeutic Context

As a member of the protein family family, CAMK1D represents a potentially druggable target. IDG targets within this class are understudied relative to well-annotated family members, with limited characterization of biological function, expression patterns, and disease associations. The sparse annotation landscape of CAMK1D complicates therapeutic prioritization. Characterization efforts for understudied IDG targets typically involve expression profiling, signaling pathway mapping, cellular phenotypic screening, or structure-based virtual screening once a reliable structural model is available.

4.2 Coexpression Insights

Coexpression analysis reveals that CAMK1D shares expression patterns with IPCEF1, RASSF2, LDLRAD4, 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 CAMK1D activity, providing clues about its biological context and potential therapeutic relevance.

4.3 Literature Network Insights

GeneRIF-based similarity identifies CDC123 as the most literature-connected neighbor, though similarity scores were low across the board, reflecting the sparse annotation landscape around CAMK1D.

4.4 Knowledge Gaps and Priorities

Key priorities for future investigation of CAMK1D include:

1. Elucidation of biological function and molecular interactors through proteomic interaction assays and cellular phenotypic screening.
2. Characterization of downstream signaling pathways and biochemical activity via targeted assays and omics profiling.
3. Tissue-specific expression profiling and cellular localization using single-cell RNA sequencing and immunofluorescence.

4. Genetic association studies to link CAMK1D 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 contextualize CAMK1D within broader pathophysiological networks.

5. Supplementary Information

This section provides quick-reference summaries and resource links for CAMK1D. The GSFM model page can be revisited for up-to-date functional predictions as the model is retrained:

GSFM predictions for CAMK1D: <https://gsfm.maayanlab.cloud/gene/CAMK1D>

5.1 Data Summary

Key identifiers and annotation counts for CAMK1D are summarized below.

Field	Value
Symbol	CAMK1D
Entrez Gene ID	57118
Ensembl	ENSG00000183049
IDG classification	darkkinase, kinase
PubMed co-mentions	120

Table 11. Data summary for CAMK1D

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. Rehfuß, 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).