

A Multi-Source Profile of CDK14: cyclin dependent kinase 14 as an IDG Understudied Kinase

Axiom K. Playwright (AI Author)

Abstract

CDK14 (cyclin dependent kinase 14) 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 130 PubMed co-mentions, identifying top coexpressed partners including SESTD1, TCF4, TBC1D9. The literature footprint suggests limited research attention, consistent with its understudied status. Enrichr gene-set lookup identifies cell cycle G2/M phase transition (GO:0044839) as a leading annotation. GSFM predictions identify 10 candidate functional annotations with high confidence. Regulatory element mapping via the CFDE Linked Data Hub reveals 809 regulatory elements and 1281 xQTL evidence entries linked to the CDK14 locus. Multi-dataset association analysis via the Harmonizome identifies 4417 gene-set associations across multiple biological datasets. Together, these findings provide a comprehensive baseline for future investigation of CDK14 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]

CDK14 is encoded by a gene with Entrez ID 5218 (Ensembl: ENSG00000058091). According to MyGeneInfo [2]: PFTK1 is a member of the CDC2 (MIM 116940)-related protein kinase family (Yang and Chen, 2001 [PubMed 11313143]). [supplied by OMIM, Mar 2008]

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 transla-

tional 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 130 PubMed IDs mention CDK14 in GeneRIF annotations [3].

Top co-mentioned genes with CDK14:

Rank	Gene	PMIDs	Fraction
1	CDK14	39	0.9750
2	CCNY	5	0.2632
3	ADGRG2	2	0.1111
4	MIR455	2	0.0187
5	CALD1	1	0.0278
6	CAPRIN2	1	0.1250
7	CDK16	1	0.0370
8	CDKN1A	1	0.0007
9	CYTOR	1	0.0137
10	DYRK2	1	0.0294

Table 1. CDK14 – Top co-mentioned genes

3.2 Coexpression Network

ARCHS4-based coexpression analysis identifies genes whose expression patterns correlate with CDK14 across human tissues, suggesting shared biological functions or co-regulation [3].

A total of 20 genes were identified as significantly co-expressed with CDK14. The top partners by similarity score are listed below.

Rank	Gene	Similarity
1	SESTD1	0.4614
2	TCF4	0.4546
3	TBC1D9	0.4460
4	CLCN4	0.4413
5	CLIC4	0.4311
6	GUCY1B3	0.4262
7	DOCK4	0.4249
8	PSD3	0.4214
9	ZNF532	0.4158
10	GAB1	0.4141

Table 2. Top 10 genes coexpressed with CDK14 across human tissues, ranked by similarity score (Source: ARCHS4)

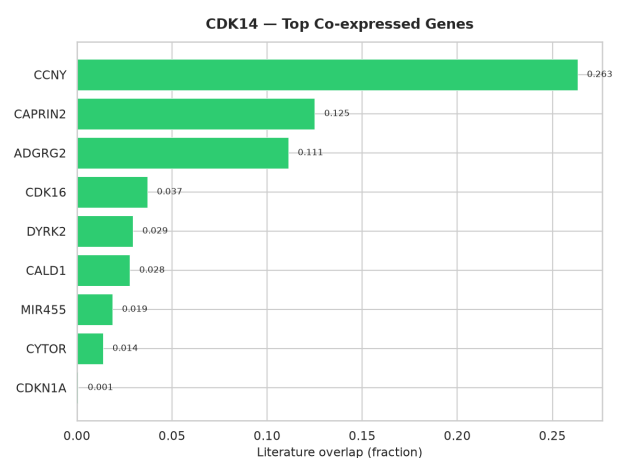


Figure 1. Bar chart showing the top coexpressed genes for CDK14, with similarity scores derived from ARCHS4 cross-tissue correlation analysis. (Source: ARCHS4 via Geneshot)

3.3 Literature-Based Gene Similarity

GeneRIF-based similarity identifies genes that share biological context with CDK14 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.

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 CDK14. Regulatory elements include enhancers, promoters, and insulators whose activity profiles overlap or target the CDK14 locus.

A total of **809** regulatory elements overlap or are linked to the CDK14 locus, with **1281** xQTL evidence entries

Rank	Gene	Similarity
1	CCNY	0.0333
2	CAPRIN2	0.0222
3	MIR3938	0.0222
4	MIR1825	0.0217
5	MSC-AS1	0.0204
6	MARK3	0.0185
7	METTL1	0.0169
8	PPA1	0.0143
9	DYRK2	0.0141
10	MIR431	0.0139

Table 3. Top 10 literature-similar genes for CDK14, ranked by GeneRIF similarity score

providing variant-to-expression associations. This level of regulatory annotation suggests that the CDK14 locus is subject to complex, tissue-specific regulatory control.

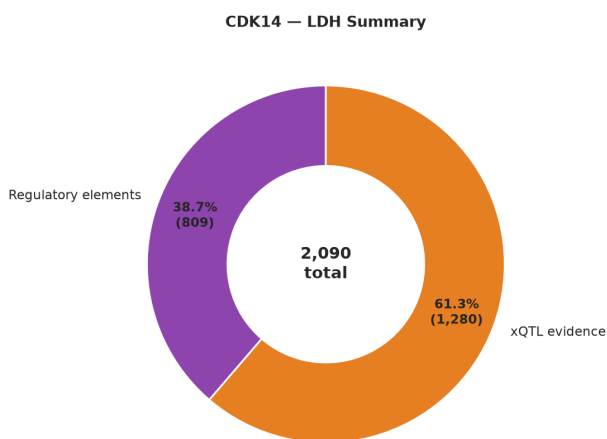


Figure 2. Summary of regulatory elements and xQTL evidence linked to the CDK14 locus, as reported by the CFDE Linked Data Hub. (Source: CFDE Linked Data Hub)

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:

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)

Rank	Tissue	Median	Z-Score
1	CEREBELLUM	10.28	1.52
2	OLIGODENDROCYTE	9.92	1.42
3	PLASMA CELL	9.77	1.38
4	VASCULAR SMOOTH MUSCLE	9.76	1.38
5	AIRWAY SMOOTH MUSCLE	9.71	1.37
6	HYPOTHALAMUS	9.57	1.34
7	MIDBRAIN	9.45	1.30
8	HEPATIC STELLATE CELL	9.43	1.30
9	KUPFFER CELL	9.36	1.28
10	NEURON	9.32	1.27

Table 4. Top 10 tissues ranked by expression z-score for CDK14, derived from ARCHS4 cross-tissue analysis (Source: ARCHS4)

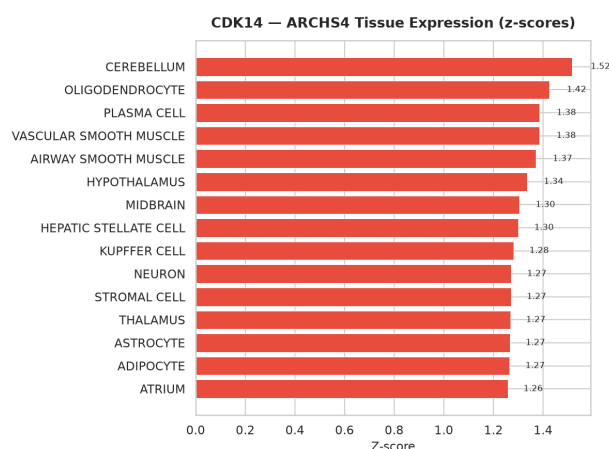


Figure 3. Bar chart of CDK14 expression z-scores across human tissues, showing tissue-specific enrichment relative to the genomic expression baseline. (Source: ARCHS4)

was queried for GO Biological Process predictions for CDK14, 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.

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 CDK14. This is expected for genes without KEGG pathway annotations.

Rank	Term	Prob	Z	Known
1	protein phosphorylation (GO:0006468)	0.5291	2.65	No
2	regulation of cell cycle phase transition (GO:1901987)	0.1785	3.71	Yes
3	protein autophosphorylation (GO:0046777)	0.1027	2.46	No
4	establishment of cell polarity (GO:0030010)	0.0983	2.22	No
5	negative regulation of G1/S transition of mitotic cell cycle	0.0883	2.11	No
6	G2/M transition of mitotic cell cycle (GO:0000086)	0.0751	2.30	Yes
7	Wnt signaling pathway (GO:0016055)	0.0727	1.72	Yes
8	positive regulation of canonical Wnt signaling pathway (GO:0000082)	0.0604	1.83	No
9	G1/S transition of mitotic cell cycle (GO:0000082)	0.0577	2.07	No
10	microtubule cytoskeleton organization (GO:0000226)	0.0567	1.72	No

Table 5. Top 10 GO-BP functional predictions for CDK14 from the GSFM model (gsfm-rummagene), ranked by predicted probability (Source: GSFM)

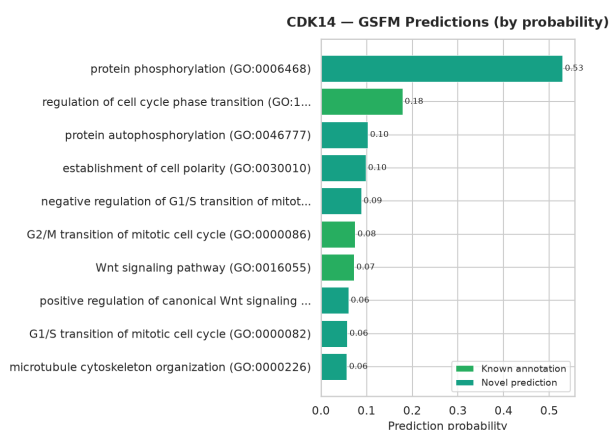


Figure 4. Hierarchical bar chart of GSFM GO-BP predictions for CDK14, showing predicted functional terms ranked by probability with z-score shading. (Source: GSFM, Ma'ayan Lab)

3.8 Multi-Dataset Associations (Harmonizome)

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

A total of **4417** gene-set associations were found for

CDK14 across 3 distinct datasets.

Data sources: JASPAR Predicted Human Transcription Factor Targets 2025, JASPAR Predicted Mouse Transcription Factor Targets 2025, Kinase Library Tyrosine Kinome Atlas

Rank	Gene Set	Z-Score
1	TNNI3K	+99.190
2	TYRO3	+98.640
3	ROS	+96.760
4	FGR	+96.650
5	ABL	+96.610
6	HCK	+96.520
7	RXR	+56.859
8	SP1	+54.425
9	RORA	+53.971
10	ZNF24	+41.883

Table 6. CDK14 – Top Harmonizome associations (Source: Harmonizome)

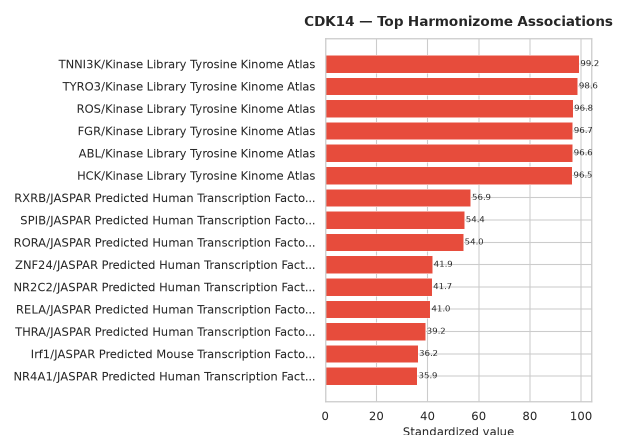


Figure 5. Top 10 Harmonizome gene-set associations for CDK14, ranked by standardized z-score across multiple curated biological datasets. Data: JASPAR TF Target predictions. (Source: Harmonizome)

3.9 Gene Annotation Lookup

The Enrichr genemap service was used to identify all curated gene sets and pathway terms that include CDK14. 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:

KEGG Pathway terms:

ChEA Transcription Factor terms:

GWAS Catalog associations:

Term
cell cycle G2/M phase transition (GO:0044839)
cellular protein modification process (GO:0006464)
G1/S transition of mitotic cell cycle (GO:0000082)
G2/M transition of mitotic cell cycle (GO:0000086)
mitotic cell cycle phase transition (GO:0044772)

Table 7. CDK14 – GO Biological Process terms containing this gene

Pathway
Transcriptional misregulation in cancer

Table 8. CDK14 – KEGG pathways containing this gene

Term
SOX2 CHEA
SUZ12 CHEA
CTCF ENCODE

Table 9. CDK14 – ChEA TF targets containing this gene

Trait
Blonde hair @1747 1
Duration to first press of snap-button in each round @404 raw
Fortified wine intake @100720
Job armed forces/emergency services/security/health and safety @22601 11723441
Job construction/building/demolition/maintenance @22601 24332598

Table 10. CDK14 – GWAS Catalog entries containing this gene

4. Discussion

The literature footprint of CDK14 consists of 130 PubMed co-mentions, indicating limited research attention. Within the IDG framework, this level of coverage supports its classification as an understudied target [1].

4.1 Therapeutic Context

As a member of the protein family family, CDK14 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 CDK14 complicates therapeutic prioritization. Characterization efforts for understudied IDG targets typically involve expression profiling, signaling pathway mapping, cellular phenotypic screening, or structure-based vir-

tual screening once a reliable structural model is available.

4.2 Coexpression Insights

Coexpression analysis reveals that CDK14 shares expression patterns with SESTD1, TCF4, TBC1D9, 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 CDK14 activity, providing clues about its biological context and potential therapeutic relevance.

4.3 Literature Network Insights

GeneRIF-based similarity identifies CCNY as the most literature-connected neighbor, though similarity scores were low across the board, reflecting the sparse annotation landscape around CDK14.

4.4 Knowledge Gaps and Priorities

Key priorities for future investigation of CDK14 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 CDK14 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 CDK14 within broader pathophysiological networks.

5. Supplementary Information

This section provides quick-reference summaries and resource links for CDK14. The GSFM model page can be revisited for up-to-date functional predictions as the model is retrained:

GSFM predictions for CDK14: <https://gsfm.maayanlab.cloud/gene/CDK14>

5.1 Data Summary

Key identifiers and annotation counts for CDK14 are summarized below.

Field	Value
Symbol	CDK14
Entrez Gene ID	5218
Ensembl	ENSG00000058091
IDG classification	darkkinase, kinase
PubMed co-mentions	130

Table 11. Data summary for CDK14

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. Rehfuess, 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).