AIGR Deep Research — csr-1 (Q21992): predicted DNA-binding TF activity & metabolic-process regulation

Gene: csr-1 (C. elegans, NCBITaxon:6239) | **Seed accession:** Q21992 **Hypothesis (BioReason-Pro SFT):** DNA-binding transcription factor activity (GO:0003700) and regulation of metabolic process (GO:0019222; positive regulation GO:0009893).

Executive Judgment

Verdict: REFUTED (with a critical accession-mismatch caveat).

The prediction of **DNA-binding transcription factor activity (GO:0003700)** is refuted under *both* plausible identities of the supplied identifier:

1. **The gene named `csr-1`** (WormBase locus **F20D12.1**) encodes the **CSR-1 Argonaute** — a small-RNA (22G-RNA)-guided effector with **PAZ (IPR003100)** + **PIWI (IPR003165 / Pfam PF02171)** + **RNaseH-like** domains. It is a post-transcriptional/chromatin-associated RNA slicer, **not** a sequence-specific DNA-binding transcription factor. Current UniProt accessions: **H2KZD5** (CSR-1a, 1030 aa) and **Q27GU1** (CSR-1b, 867 aa).
2. **The supplied accession Q21992 is DELETED** in UniProt (Inactive/DELETED; UniParc UPI000007665A). Its sequence maps unambiguously to a **different gene — R144.7** = **`larp-1` (La-related protein 1)**, an RNA-binding translational regulator (current reviewed entry **D5MCN2**, 1150 aa; RefSeq NP_001040868.3 "La-related protein 1"). LARP-1 binds RNA (poly-U/poly-G, RNA cap) and positively regulates translation — it is **also not** a DNA-binding transcription factor.

Neither identity supports GO:0003700. The **metabolic-process regulation** terms (GO:0019222 / GO:0009893) are so broad that they are *technically not false* for either gene (both regulate a macromolecule-metabolic process — gene silencing or translation), but they are **uninformatively broad and non-core**, and they do not rescue the mechanistic TF claim, which is the substantive part of the prediction.

Most important caveat: There is a genuine identifier problem. The seed pairs symbol `csr-1` with accession `Q21992`, but Q21992 ≠ CSR-1. A curator must first resolve which protein is intended. Under either resolution the TF prediction fails.

Sequence-Level Provenance (computed this run)

Direct pairwise 6-mer containment on retrieved sequences resolves the identity quantitatively:

Comparison	Lengths	6-mer containment	Interpretation
Q21992 (deleted seq) vs **LARP-1 (D5MCN2)**	1065 vs 1150	**0.971**	Same protein (identical N-terminus `MAEKQPMLSFAKVVSGQAED...`); Q21992 is an older LARP-1 isoform
Q21992 vs **CSR-1a Argonaute (H2KZD5)**	1065 vs 1030	**0.000**	Unrelated — no shared 6-mers
LARP-1 (D5MCN2) vs CSR-1a (H2KZD5)	1150 vs 1030	**0.000**	Unrelated proteins

This confirms the seed's `csr-1 → Q21992` pairing is an accession error: Q21992 is larp-1, sharing zero detectable sequence with the CSR-1 Argonaute.

Evidence Matrix

Citation	Evidence type	Supports/ Refutes	Claim tested	Key finding	Context	Confidence / limitations
UniProt H2KZD5, Q27GU1 (database)	Structural/ domain	**Refutes** GO:0003700	csr-1 is a DNA-binding TF	CSR-1a/b carry PAZ + PIWI + RNaseH-like; no DNA-binding/TF domain	C. elegans, protein family	High; database-level but based on curated domain models
P 19804758 (Claycomb 2009)	Mutant phenotype / localization	Refutes	csr-1 identity/ function	CSR-1 Argonaute + EGO-1/DRH-3/ EKL-1 localize to chromosomes and are required for chromosome segregation via 22G-RNAs; CSR-1 does not down-regulate target mRNA/ protein	C. elegans germline/ embryo	High; primary paper
P 24178449 (Wedeles 2013, review)	Review synthesis	Refutes	csr-1 identity	CSR-1 binds 22G-RNAs complementary to ~25% of genes; only essential Argonaute of 24	C. elegans	High for identity; review-level
P 22863779 (Avgousti 2012)	Direct assay	Qualifies (BP)	csr-1 regulates gene expression	CSR-1 directly binds histone mRNA; RNAi pathway positively regulates	C. elegans germline	High; shows post-transcriptional (not TF) regulation

Citation	Evidence type	Supports/ Refutes	Claim tested	Key finding	Context	Confidence / limitations
				histone expression		
P 25510497 (Tu 2015)	Comparative genomics	Refutes	csr-1 identity	Conserved nuclear Argonaute role in chromatin/ segregation/ germline expression across Caenorhabditis	C. elegans / C. briggsae	High
UniProt Q21992 record (database)	Database	**Competing/ refutes**	Q21992 = csr-1	Q21992 is DELETED; UniParc maps to R144.7 (larp-1), RefSeq NP_001040868 = "La-related protein 1"	Identifier check	High; direct record retrieval
UniProt D5MCN2 (database)	Structural/ domain	Refutes	Q21992-seq is a DNA-binding TF	LARP-1: La domain (IPR045180), DM15, La-HTH; only a winged-helix-**like** structural fold (IPR036388); function = RNA binding + positive regulation of translation	C. elegans	High; La proteins bind RNA, not DNA
	Mutant phenotype / localization	Refutes (for larp-1 identity)	Q21992/ larp-1 is a	LARP-1 has La motif + LARP1 domain, binds	C. elegans	High; primary paper — RNA-binding, not

Citation	Evidence type	Supports/ Refutes	Claim tested	Key finding	Context	Confidence / limitations
P 18515547 (Nykamp 2008)			DNA-binding TF	RNA in vitro, localizes to germline P bodies, and attenuates Ras-MAPK signaling by controlling pathway mRNA/ protein abundance	germline/ oogenesis	DNA-binding TF
Computed provenance (this run)	Computational	Refutes/ Competing	Q21992 = csr-1	6-mer containment: Q21992↔LARP-1 = 0.971; Q21992↔CSR-1a = 0.000	Sequence analysis	High; direct calculation
UniProt Q17370 (database)	Database	Competing	Source of TF signal	nhr-47 (Nuclear hormone receptor) surfaces on `gene:csr-1` text search — a genuine DNA-binding TF but a *distinct* gene	C. elegans	Medium; illustrates possible name/ paralog confusion

GO Decision Table (leads — require curator verification)

Predicted term	Aspect	Applies to CSR-1 Argonaute?	Applies to Q21992/LARP-1?	Recommended action
GO:0003700 DNA-binding transcription factor activity	MF	No (PAZ/PIWI RNA slicer)	No (La/LARP1 RNA-binding)	**Reject / do not add**
GO:0019222 regulation of metabolic process	BP	Only as over-broad parent of ncRNA gene silencing	Only as over-broad parent of translation regulation	**Do not add — too broad, non-core**
GO:0009893 positive regulation of metabolic process	BP	No specific support	Weak/broad (positive regulation of translation)	**Do not add — too broad, non-core**
GO:0070551 endoribonuclease cleaving siRNA-paired mRNA	MF	**Yes (IDA, WormBase)**	n/a	Correct MF for CSR-1 (already annotated)
GO:0008266 poly(U) RNA binding	MF	n/a	**Yes (IDA, WormBase)**	Correct MF for LARP-1 (already annotated)

GO Curation Implications (leads — require curator verification)

- **GO:0003700 (DNA-binding transcription factor activity, MF): DO NOT ADD / REJECT.** Refuted for both csr-1 (Argonaute) and Q21992/larp-1 (RNA-binding). No experimental or domain evidence for sequence-specific DNA binding.
- **GO:0019222 / GO:0009893 (regulation / positive regulation of metabolic process, BP): treat as NON-CORE and too broad.** Do not add on the strength of this prediction. If any regulatory BP is warranted it should be specific and mechanism-appropriate:

- For **CSR-1**: e.g., *regulatory ncRNA-mediated post-transcriptional gene silencing* (GO:0035194), *siRNA-mediated gene silencing by mRNA destabilization* (GO:0090625) — already present with experimental evidence.
- For **LARP-1** (if that is the intended protein): *positive regulation of translation* (GO:0045727) — already present.
- **Recommended MF for the actual proteins** (informative, not "protein binding"):
- CSR-1: *endoribonuclease activity, cleaving siRNA-paired mRNA* (GO:0070551); *siRNA binding / RNA binding*.
- LARP-1: *poly(U) RNA binding* (GO:0008266); RNA cap binding.

Mechanistic Scope

The predicted MF (DNA-binding TF) posits **direct, sequence-specific DNA binding driving transcription**. The actual immediate molecular activities are RNA-centric: CSR-1 is a **small-RNA-guided endoribonuclease/effector** that recognizes target mRNAs by 22G-RNA base-pairing (chromosome segregation, histone-mRNA and germline-mRNA regulation, mRNA surveillance/licensing). LARP-1 is an **RNA-binding translational regulator**. Any "regulation of metabolic process" is a *downstream, category-level* description of gene-expression regulation, not a direct DNA-binding transcriptional activity.

Conflicts and Alternatives

- **Identifier/accession mismatch (primary conflict):** Q21992 (deleted) ≠ csr-1; it is larp-1/R144.7. This is the single most important issue for the curator.
- **Winged-helix-like fold artifact:** LARP-1 carries a *winged-helix-LIKE DNA-binding structural superfamily* fold (IPR036388/IPR036390). A predictor keying on that structural signature could spuriously infer "DNA-binding" — but La/winged-helix proteins bind **RNA**, so this is a false positive.
- **Paralog/name confusion:** A text search on `gene:csr-1` also surfaces **nhr-47** (a nuclear hormone receptor and bona fide DNA-binding TF). Frequency bias toward abundant TF/NHR classes plus fuzzy name/xref linking is a plausible origin of the GO:0003700 prediction, but nhr-47 is a separate gene.
- No primary literature attributes DNA-binding transcription-factor activity to CSR-1 or LARP-1.

Knowledge Gaps

1. **Which protein does the prediction actually target?** Checked: Q21992 is deleted and maps to larp-1; the symbol says csr-1 (Argonaute). Resolution needed from the curator/model provenance. Matters because the correct-function assignment differs (Argonaute vs La-protein), though the TF verdict is the same. Resolve by confirming the sequence the model was given (compare to H2KZD5/Q27GU1 vs D5MCN2).
2. **Did BioReason ingest the stale/deleted record?** If the model used Q21992's larp-1 sequence but the symbol csr-1, that is a data-hygiene error worth flagging upstream.
3. **Basis of the TF signal.** Not directly observable here; a domain/embedding attribution (winged-helix-like fold vs nhr text co-occurrence) would confirm the false-positive mechanism.

Discriminating Tests

- **Sequence-identity check:** Align the exact sequence behind the prediction against H2KZD5/Q27GU1 (CSR-1) and D5MCN2 (LARP-1); % identity resolves identity instantly.
- **Domain scan (InterProScan/HMMER):** Presence of PIWI+PAZ ⇒ Argonaute; La+DM15 ⇒ LARP; absence of homeodomain/zinc-finger/bHLH/NHR DNA-binding domains ⇒ not a TF.
- **DNA vs RNA binding assay class:** Curated evidence for CSR-1 is IP-small-RNA-seq / mRNA binding; for a TF one would expect ChIP-seq/EMSA on DNA motifs — none exists.
- **Compare against `nhr-47` (Q17370):** confirms the true DNA-binding TF neighbor is a different gene.

Curation Leads

- **Action:** Reject the BioReason-Pro prediction of GO:0003700 for csr-1/Q21992 as an over-annotation/mis-identification. Do not add GO:0019222 or GO:0009893 (too broad / non-core).
- **Flag the accession:** Note in the review that Q21992 is a DELETED UniProt entry corresponding to larp-1 (R144.7), not the csr-1 Argonaute locus (F20D12.1; current H2KZD5 / Q27GU1). Recommend the review use a current csr-1 accession.
- **Candidate references to verify (exact snippets):**

- P 19804758
  — "the Argonaute CSR-1 … are required for proper chromosome segregation."
- P 24178449
  — "CSR-1 interacts with small RNAs known as 22G-RNAs … the only essential C. elegans Argonaute out of 24 family members."
- P 22863779
  — "CSR-1 directly binds to histone mRNA in an ego-1-dependent manner."
- **Suggested curator questions:** (a) Which sequence did the model score? (b) Is the TF call driven by the LARP winged-helix-like fold or by nhr-name co-occurrence? (c) Should the review record be re-keyed to a live csr-1 accession?
- **Suggested experiment (if identity ever in doubt):** InterProScan + reciprocal-best-hit orthology to human AGO/PIWI vs LARP1; ChIP-seq would be expected to be negative for DNA-motif occupancy.

Provenance: UniProt REST (Q21992 status, H2KZD5, Q27GU1, D5MCN2), UniParc UPI000007665A cross-references, NCBI eutils esummary (NP_001040868 = "La-related protein 1"), and PubMed primary/review literature. All retrievals executed programmatically in this run.
