anti-CD38 — V4 Target Opportunity Report

Target: CD38 — CD38 molecule (HGNC:1667, UniProt P28907, Wikidata Q14885119)
Substrate: 494 claim-scope-binned patent families, of which 360 are on-pathway for CD38
Structure: Part I is what to act on. Part II is the structural evidence under it. Part III is how it was built and where it fails.

> *Public edition. This presents the **whitespace and trajectory** reads that accompany the article. The underlying claim-strength scoring, per-family score tables, knowledge graph and raw inputs are part of the internal analyst deliverable and are **not reproduced here**; that strength layer is relative-ordering, calibration-pending work and is deliberately not published as a ranking.*

Read-me guardrails

These bound every statement in this document. They are repeated in Part III.

1. **This is not a freedom-to-operate opinion.** No claim-construction, no infringement reading, no validity opinion. Scope-overlap and sequence-identity results are **candidate generation**: they say where to look, never that anything is clear.
2. **Every count is a lower bound.** The corpus is a floor, not a census — discovery saturated at its row cap on 8 slices, one upstream leg failed on 12 of 71 slices, and 29 families failed ingest at the newest publication series.
3. **The claim-strength layer is uncalibrated, and is not published here.** It orders families and cells **relative to one another** within a single run. It is not a valuation, not a price, and not comparable across targets or cohorts. It is part of the internal deliverable and is deliberately **not reproduced in this edition as a ranking**; where a reading below rests on it, this document says so and does not restate the score.
4. **That layer is publication-scoped, not family-scoped.** Its scored units are publications standing in for a larger member population, so it under-describes the corpus it is computed over.
5. **A "not measured" is not a zero.** Where a layer did not run, this report says so and does not convert the silence into a finding.

PART I — ACT ON THIS

1. Actionable summary

1.1 The headline finding is a negative one, and it is the finding

This run probed 35 cells that the grid called OPEN. After adjudication, **0 of them are reportable as whitespace**.

That is the deliverable. Not a hedge on the way to a recommendation — the result. Of those probed cells, 3 turned out to be occupied once someone looked, and 32 could not be measured at all. Every one of the 26 unmeasured probes **did run**; they returned nothing because a leg of the search errored or the phrasing did not match, not because the space is empty.

So there is no whitespace claim to act on. What follows is what the run *can* support: where the wall is, where the wall is thin, and where nobody has yet looked properly.

1.2 Where the wall is

There are two centres of this field, and they are **not the same cell**.

	the densest cell	the strongest cell
cell	target_cell_killing x deplete_cytotoxic	ectodomain_epitope x antagonize_block
verdict	COVERED	COVERED
on-pathway families	160	96
identified by	family count	the internal claim-strength read (not published)

The cell leading on family count is target-cell killing by cytotoxic depletion — the therapeutic mainstream. The cell that leads on the internal claim-strength read is a different one: the ectodomain epitope approached by antagonist blocking. Both are COVERED, and 4 cells in total carry that verdict, over 118 live granted on-pathway families.

Neither is a place to design into, but they are walls of different kinds. The crowded cell is the larger of the two and the epitope cell is the smaller, yet on the internal claim-strength read the ordering reverses. Crowding and claim architecture come apart here, and a strategy that treats family count as a proxy for strength would rank these two cells the wrong way round. The ordering itself is part of the withheld layer; what is publishable, and what matters for siting a programme, is that **the densest cell and the strongest cell are not the same cell**.

1.3 Where the wall is thin — with the caveat that thin is a measurement, not an opening

verdict	cells	what it means
THIN	9	at or below the family-count floor
EVIDENCE_THIN	7	occupied, with title-only evidence predominating
PARTIAL	6	occupied without meeting the COVERED bar
OCCUPIED_SECONDARY	20	reached as a secondary axis value

EVIDENCE_THIN is the honest one to act on: those cells are occupied by families whose binning rests on **title text alone** — 114 of 360 on-pathway families (31.7 per cent) were binned without claim text. A cell that is thin *because nobody read the claims* is a research task, not a filing opportunity.

1.4 Design-around: what the structural layer actually licenses

The sequence-overlay layer is the one place this run produced a positive, checkable, actionable result. Across 2793 family pairs carrying a sequence edge, **47 pairs combine high identity with claim anchoring on the same alignment**.

population	pairs	why it is the wrong number to quote alone
any sequence edge	2793	includes platform and constant-region similarity
claim-anchored, any edge	211	anchored, but the anchoring sequence may be the weak one
identity at or above the comparability threshold	1793	identity no claim recites
both, but on different sequences	163	the weaker reading, and the easy one to quote by mistake
both, on the same alignment	**47**	**the evidentiary subset**

Every pair in the bolded row is enumerated by identifier in the internal deliverable. **That list is the design-around worksheet**: each entry is two families whose claimed sequences are close enough that a binder designed away from one lands near the other. It is candidate generation — the pairs to put in front of counsel, not a clearance.

The screen behind it is deliberately conservative in one direction and lossy in another: 84.8 per cent of possible sequence pairs were discarded by the k-mer prefilter **without being aligned**. The count is a floor.

anti-CD38 — structural overlay of CLAIMED antibody sequences

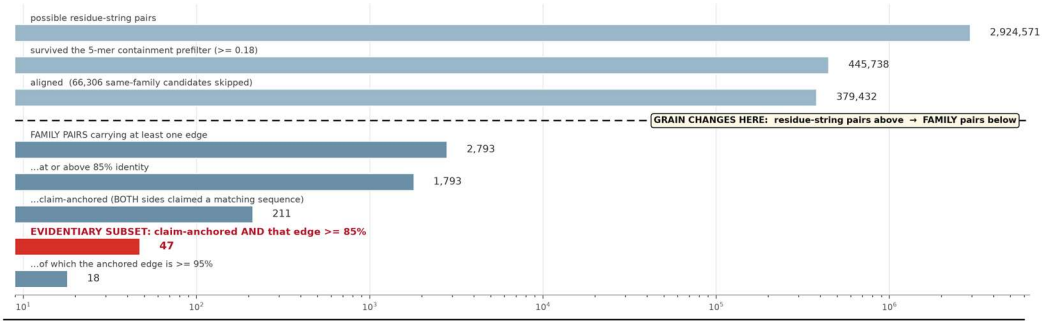
Off-platform local alignment over 3,318 protein sequences from 140 families. The figure draws ONLY the artefact's evidentiary subset: the 47 family pairs where both sides CLAIMED the matching sequence and that same match is at or above 85% identity.

14 sponsors, 24 families, 37 cells — and the sponsor split is only TRUSTED on 11 of the 47 pairs.

The larger reading — 163 pairs — is NOT drawn: it lets the identity and the claim-anchoring come from two different sequence pairs. On 23 of these 47 pairs the two already diverge.

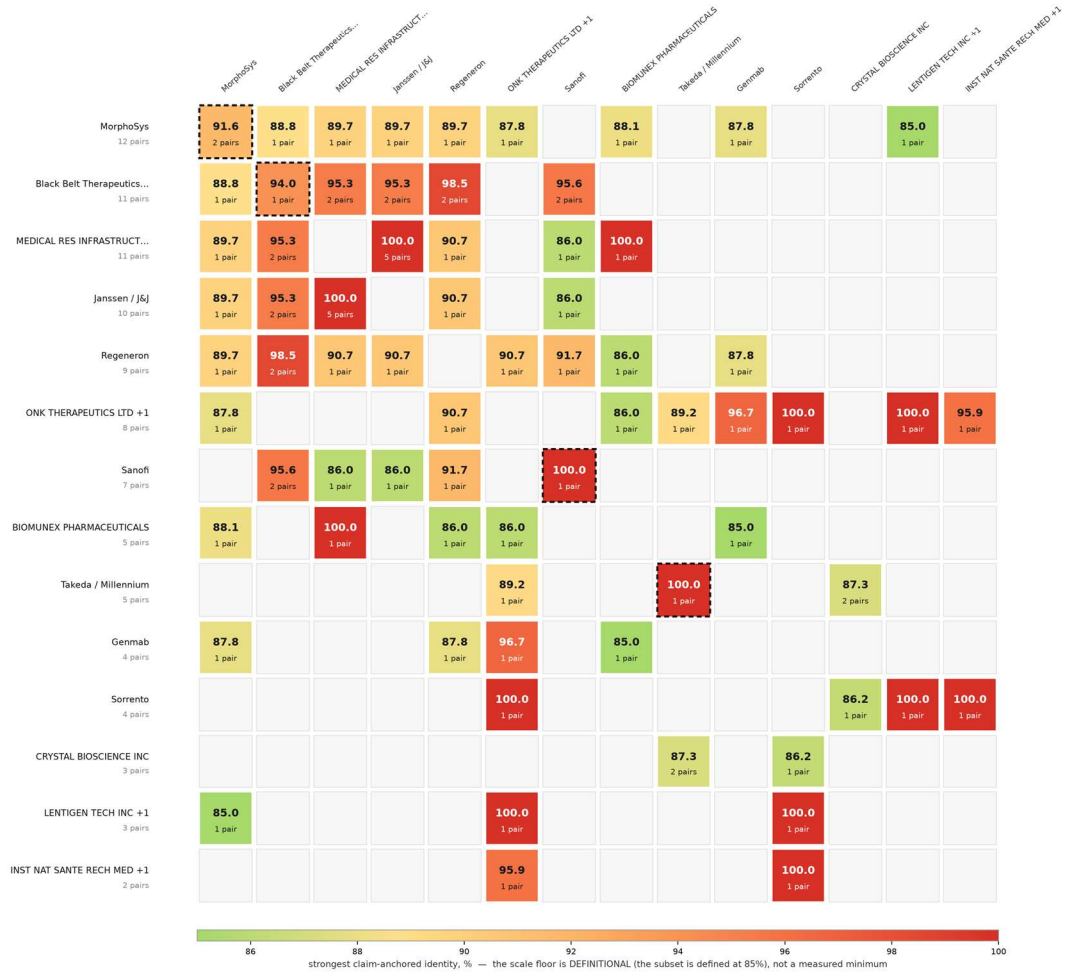
A — SURVIVORSHIP: the headline is a LOWER BOUND, not a count

2,478,833 pairs (84.8%) were discarded by the prefilter WITHOUT BEING ALIGNED. A discarded pair is NOT TESTED — it is not proven dissimilar.



B — SPONSOR × SPONSOR over the 47 evidentiary pairs (cell = strongest CLAIM-ANCHORED identity, %)

37 of 105 possible sponsor cells are filled (33 cross-sponsor, 4 same-sponsor, dashed on the diagonal). The blanks are UNTESTED-or-absent, never cleared.



strongest claim-anchored identity, % — the scale floor is DEFINITIONAL (the subset is defined at 85%), not a measured minimum

SPONSOR IS NOT AN OWNERSHIP RECORD. Only 11 of the 47 pairs carry sponsor_partition_trusted — on the other 36 the same-vs-cross reading is a NAME match, not a resolved corporate identity.

Biopython PairwiseAligner, local mode, BLOSUM62, gap open -11 / extend -1. Constant regions (507 sequences) and the shared CD38 antigen (107) are removed before pairing; VARIABLE_DOMAIN (1447 sequences) is RETAINED and still confounds — 14 of these 47 pairs are that class, where shared germline framework can align as high as shared scope.

A pair is claim-anchored only if both sides carry claimed == true. Identity = matched columns / aligned columns; alignments under 50 columns are discarded. `claimed` is TRISTATE — null means NOT DETERMINED, so pre-grant families can never be claim-anchored and are absent from this figure by construction, not by finding.

Sponsor labels are the FIRST co-assignee, capped at 24 characters, with '+n' marking further co-assignees; the full strings are in structural_overlay.sponsor_cells. No whitespace verdict in this run takes a sequence edge as an input.

CANDIDATE GENERATION, NOT a freedom-to-operate determination. Sequence identity OVER-calls and is a LOWER BOUND on concern: a high cell is a question for counsel, and a blank cell is never "free to operate".

anti-CD38 structural overlay of claimed antibody sequences: the survivorship funnel from all possible residue-string pairs down to the evidentiary subset, and the sponsor matrix over that subset.

*Figure — the survivorship behind the number above, and who the surviving pairs sit between. The funnel's left-hand stages are residue-string pairs and its right-hand stages are family pairs; the grain changes once and the figure marks where. The matrix plots the identity on an edge **both** sides claimed, which is not the same quantity as a pair's headline identity.*

1.5 The chemical-core lane is not available on this run

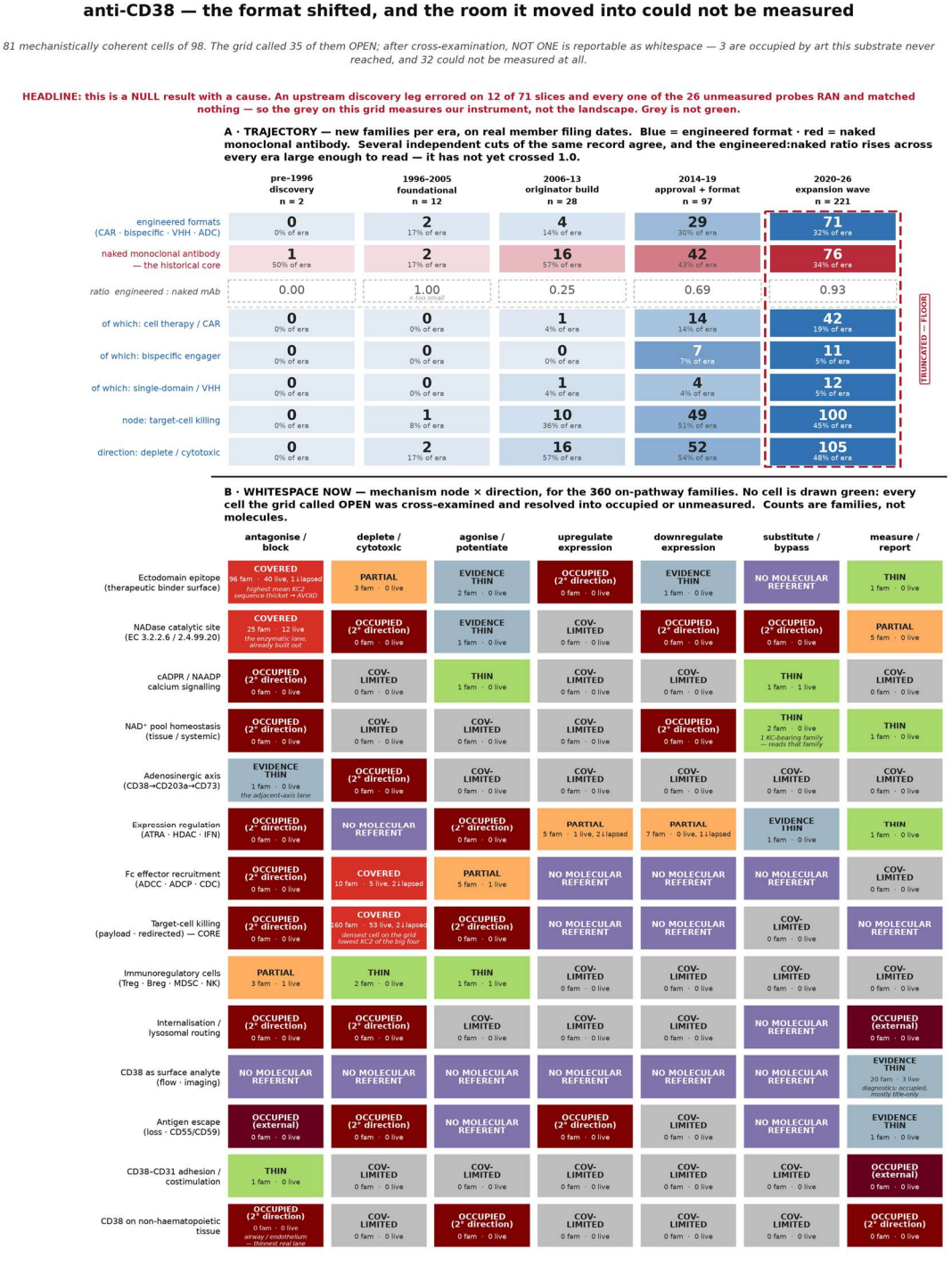
The genus/Markush layer did not compile. Its verdict is `NULL_NOT_TESTED`, and 0 of 360 probed on-pathway families returned a numeric recall. There is therefore **no actionable chemical core in this report** — no genus tiers, no where-to-build-in-chemical-space. This is a layer that did not run, and it is reported as absent rather than as an empty result.

For an antibody-dominant target this loss is narrower than it would be for a small-molecule target, but it is real: it removes the ability to say whether a small-molecule CD38 modulator would sit inside anyone's claimed genus.

1.6 Trajectory vector, and why its right-hand edge is untrustworthy

360 families carry a usable date across 5 populated eras. The direction of travel is away from naked antagonist antibodies toward engineered and multi-specific formats.

The newest era is the most under-counted in the corpus, because the families that failed ingest failed disproportionately at the newest publication series. Any downturn at the right-hand edge of the trajectory is a property of coverage, not of filing behaviour. Do not read a slowdown there.



anti-CD38 whitespace and trajectory: filings by era, and the mechanism-node by direction grid coloured by cell verdict.

Figure — the trajectory by era (upper panel) and the verdict grid (lower panel). No cell on this grid is whitespace; the greys and purples are cells where the evidence is limited or the mask fired, not cells that were found empty.

PART II — THE STRUCTURAL EVIDENCE

2. Scope ontology

The ontology was committed **before** the card store it bins, which is what makes the axis values a prior commitment rather than a description of what happened to be found. The axes carry 14 mechanism nodes, 7 directions, 12 modality values and 18 indication values.

Two axis limitations bear on every cell count below. A value the binner used that the pre-committed ontology does not declare. Families carrying it are ABSENT from the node x modality grid (which iterates the ontology) and sit on NEITHER side of the format-migration ratio. It is reported rather than folded into a neighbouring value, because folding it would be a post-hoc ontology edit and the ontology's whole evidential value is that it was fixed before the corpus was read.. And the indication axis is largely inert: 14 of 18 declared indication values carry any family at all.

3. Lapse timeline and legal state

state	on-pathway families
live granted	118
pending	165
lapsed	9
expired	1
withdrawn	24
unknown	43

No cell qualifies as LAPSED_OPEN. The rule requires that a cell's legal state be fully known before its lapses can be read as an opening, and 43 on-pathway families carry an unknown state. Reporting a lapse-driven opening while that many families are legally unclassified would be reporting the gap in the legal data as an opportunity.

21 on-pathway families carry an adverse event. The territorial detector never fired on this corpus, and its attainability was checked rather than assumed.

4. Combination and transfer — the sequence adjacency layer

188 families were probed for sequence adjacency and 2429 edges came back. **Every edge points outside this corpus**: 0 edges fall between two families that are both in it.

That is a real measurement with a real reading — the anti-CD38 families in this corpus are sequence-adjacent to work that is *not* anti-CD38 — and it is the transfer surface. It is also why this layer cannot be used to rank families within the corpus against each other.

5. Knowledge graph

A typed knowledge graph underlies this analysis — mechanism nodes, claim directions, modalities, indications, patent families, assignees and filing years, joined by the structural relations the report reads. It is part of the internal deliverable and is **not published**.

One property of it is worth stating because it bears on how any count in this document should be read: the graph was built with self-loops and duplicate rows removed, and the edge types this run declared and found empty are carried as an explicit zero list rather than omitted. An edge count is not an integrity check — a graph can report a large number and be largely a node pointing at itself — which is why the integrity properties, not the totals, are what this edition reports.

PART III — METHODOLOGY, RECLASSIFICATIONS & LIMITS

6. How the substrate was built

stage	in	out
discovery slices run	71	2481 rows
distinct after dedup	2481	825
admitted by the relevance gate	825	821
submitted for ingest	821	552
carded	497	494
binned	494	494

The funnel's request classes sum to 552 and account for every submission.

3 families sit in an unknown state and are named individually rather than folded into a bucket.

7. The on/off-pathway reclassification

134 of 494 binned families (27.1 per cent) were reclassified off-pathway. The largest class recites CD38 in a multi-marker panel without directing a claim at it; the next is directed at a different target entirely.

This reclassification removes 27.1 per cent of the corpus, which is why every downstream denominator in Part II is the on-pathway set and says so.

8. Both adversarial audits returned REJECT

review	scope	must-fix	should-fix	verdict
review one	the claim-strength layer	8	8	REJECT
review two	whitespace + structural layer	24	19	REJECT
total		**32**	**27**	

All 59 findings are carried in `findings_ledger.json` with the reviewer's own identifier, quote and locator, and each closes against an **evidence command that is re-run on every render**. A closure that stops being true stops the render. 0 must-fix findings are open.

The four defects worth naming, because each one had produced a number that looked fine:

- **The analytic ceiling was self-falsifying.** It enumerated part of the vertex space and published a maximum that the run's own observed value exceeded. The corrected ceiling is 0.5723 against an observed 0.4561.
- **The strength layer's scope was misdescribed as family-level.** Corrected to publication-scoped, with the member population it stands in for now measured.
- **The structural overlay's headline was low by a large factor** — claim anchoring had been read off one dedup representative and off each pair's single best edge, where family-correct and any-edge gives 211 claim-anchored pairs.
- **The coherence mask dropped families without saying so.** 3 on-pathway families landed in no coherent cell and are now named.

9. The verifier, and its measured catch rate

Every number in this document is computed and injected; the renderer refuses to emit a document containing a hand-typed quantity. Of 3048 computed figures, 374 are bound into the three documents.

A green gate is not evidence until its catch rate is measured. The mutation harness seeds corruptions transcribed from the reviewers' own findings and re-runs the gates: **12 of 12 caught**, with 3 clean controls that must not fire. Layers other than the renderer are reported as not tested rather than as passing.

The verifier that checks the finished documents runs 13 checks. It reads digits **out of the rendered prose** and asks whether each traces to a computed figure, rather than recomputing a value the way the document did — the second method confirms a document instead of testing it. Its structural checks assert conservation, containment and null-vs-zero **properties** rather than totals, because a graph can report a correct edge count while being largely a node pointing at itself.

Every check also ships a mutation of its own input and must reject it. A check that passes its own mutation is reported as vacuous and fails the run; on this run 0 checks are vacuous, for a falsifiability rate of 100 per cent.

The verifier earned its place on this run. The design-around section's evidentiary subset was originally bound to the weaker overlap figure — 163 pairs, which lets the identity and

the claim anchoring come from two different sequences — instead of the same-alignment figure, 47. Both are computed, both are real, and the gates that enforce derivation cannot tell them apart; only a check that knows which one is the evidentiary subset can.

10. Limits — what this run cannot tell you

1. **No chemical core.** The genus layer did not compile.
2. **No whitespace.** 0 open cells survived cross-examination.
3. **Discovery has a structural hole at the pre-grant edge.** 269 discovered rows had no resolvable entry point, and they sit in the pre-grant and pending statuses a forward-looking analysis wants most.
4. **12 degraded slices returned nothing on every leg**, and those slices carried the late-stage programme code names.
5. **The internal strength layer covers only part of the corpus** — 68.4 per cent of carded families — so every reading that rests on it is a reading of that subset.
6. **One binning shard is not reproducible** — there is no committed script that produced it, and it is disclosed rather than quietly re-run.
7. **Provenance identifiers are unavailable on this deployment**, so the producer census behind the internal strength layer is a served-block re-read rather than a resolved provenance chain.

11. What the next run should do

1. **Fix the upstream leg before anything else.** 12 of 71 slices lost a leg, and that failure feeds the unmeasured cells that made this run's whitespace answer empty.
2. **Re-probe the 32 unmeasured cells with alternate phrasing**. Their probes ran and matched nothing; the phrasing is the suspect, not the space.
3. **Retry the leading-edge ingest failures**, which is where the trajectory under-counts.
4. **Read claims for the title-only families**. That single action would move the EVIDENCE_THIN cells from a data defect to a finding either way.
5. **Re-attempt the genus compile.** It is the missing actionable lane, and its absence is a service state rather than a property of the target.

Method, in brief — and what is not published here

How the substrate was built is unabridged above: 71 term-sliced discovery calls across four jurisdictions, a relevance gate, a 552-submission ingest, a per-family claim-scope read that placed every family on four pre-committed axes, a per-family legal read, and an off-platform sequence-overlay layer. **What each family claims was read by a model from claim text where claim text existed** — and where it did not, from the title alone, which is why the title-only fraction is stated in Part I rather than buried.

What is deliberately withheld, per the series' publication rules: the claim-strength scoring layer and every per-family, per-cell and per-assignee table derived from it — it is uncalibrated and relative-ordering only, so publishing it as a ranking would assert a

precision that does not exist and would put named-party quality judgements on the record; the knowledge graph and the scope ontology; the scored inputs and the scripts; and the internal audit records and findings ledger. The reading is published; the engine is not.

What that means for a reader. Every number that remains — family counts, cell verdicts, live-grant and lapse counts, the sequence-overlay pair counts, the filing trajectory — is structural, and reproducible from the public record with patience. None of it is a valuation and none of it is freedom-to-operate. The guardrails at the head of this document apply to every sentence in it without exception, and the two that matter most bear repeating: **this is not a freedom-to-operate opinion**, and **every count is a lower bound** — the corpus is a floor, not a census.

The headline finding is unchanged by the trim. This run probed 35 cells the grid called OPEN and **0 of them are reportable as whitespace**. That result rests on structural evidence alone; nothing withheld here supports it, and nothing withheld here would soften it.

Figure: the whitespace/trajectory grid and the structural-overlay exhibit accompany this edition.