CLDN18.2 — V4 Whitespace & Trajectory Analysis

Target: CLDN18.2 (Claudin-18 splice-variant 2) — gene CLDN18 (HGNC:2039, UniProt P56856; Wikidata gene **Q18040098**, protein **Q21100780**). A lineage-restricted tumor-associated antigen: the isoform-2-specific first extracellular loop (ECL1), de-repressed on gastric/GEJ, pancreatic and biliary adenocarcinomas. Reference drug: zolbetuximab (Vyloy/IMAB362; originator Ganymed/TRON; Astellas; FDA-approved 2024). **Date:** 2026-07-26 · **Whitespace denominator: 106 on-pathway families** (of 185 native-KC-scored; 79 platform co-mentions screened OFF-pathway) · real filing dates on all 106. **What this is:** candidate generation and a whitespace/trajectory **signal — NOT a freedom-to-operate opinion.** "Whitespace" = absence of *strong structural claim coverage* (necessary-but-not-sufficient). Genus recall and sequence identity are **lower bounds**; verdicts are 3-valued.

> **Read-me (guardrails, once):** An OPEN/THIN cell is a hypothesis prompt, not clearance. The denominator is operator-defined (the scope ontology). The corpus is **US-anchored**; the very large China/EP/WO ex-US tail is bibliographic-only, so a cell unpopulated *only in the US-anchored corpus* is a corpus artifact, not clean global whitespace — flagged where it occurs. Native KC is uncalibrated (relative strength only).

---

1. The headline — this is a COVERED, crowded, young, China-dominated landscape

CLDN18.2 is **not** a target with broad open mechanistic whitespace. It is a single, well-validated epitope in the middle of an explosive, defensive land-grab. The whitespace geometry is therefore **modality-and-timing**, not mechanism:

- **Every established modality cell is COVERED.** Naked-mAb (39 on-pathway families, 16 live grants), ADC (16, 9 with a compiled genus), CAR/cell (16), non-CD3 bispecific (15), CD3 T-cell-engager (12) — all are densely populated with live-granted or fast-consolidating estates.
- **The one genuinely OPEN *direction* is `degrader / functional-modulation`** — no family in the 185-family corpus claims *degrading or functionally modulating the CLDN18.2 tight-junction protein itself* (all 106 on-pathway families are cytotoxic-depleting or immunomodulatory). **Caveat:** degrading a cell-surface structural tight-junction protein is mechanistically hard and this direction is unpopulated corpus-wide — treat it as a *hypothesis prompt*, the weakest of the open calls, not clean whitespace. (One off-pathway family — Orum US-20230338564 — carries a molecular-glue *degrader ADC payload*, i.e. the degrader chemistry exists but is aimed elsewhere: a transfer candidate, §5.)
- **The real opportunity is temporal, not spatial.** The newest cells — **CAR/cell-therapy and CD3 T-cell-engagers filed 2023-2026 — are almost entirely pending (2 live grants of 19 families).** The 2020-2022 surge estates are maturing

into grants now; the 2023-2026 frontier has *not yet consolidated*. This is the contestable window.

- **70% of the on-pathway estate is China-origin.** The competitive centre of gravity has moved decisively to China (74 of 106 on-pathway families), with the EU-origin foundational tail (Ganymed) and a modest US presence.

The one-sentence read: on CLDN18.2 there is little clean modality whitespace — the actionable plays are (i) a **differentiated ADC payload/linker lane** (the topoisomerase-I/DXd chemotype is crowding; anthracycline and non-topo-I payloads are thinner), (ii) the **still-pending CAR-T / T-cell-engager frontier** where grants have not consolidated, and (iii) the **speculative degrader/functional-modulation** direction — plus design-around of the sequence-defined (CDR) binder fences that dominate this antibody target.

---

2. Whitespace map — the cells

The 106 on-pathway families were projected onto the scope grid (engagement-mode × modality/format × direction × indication; ontology in `data/ontology.json`, Wikidata-grounded). **Verdict rule:** OPEN = 0 on-pathway families · THIN = ≤2 & no broad live genus · PARTIAL = coverage but no broad live-granted genus · COVERED = broad live-granted genus OR dense (≥3) live-granted estate.

Coverage by modality/format (the primary actionable axis):

modality / format	on-path families	live grants	genus families	median genus recall	verdict
naked / Fc-engineered mAb	39	16	0	— (sequence-defined)	**COVERED**
ADC (small-molecule payload)	16	4	9	**0.333**	**COVERED**
CAR / cell therapy	16	4	0	— (sequence-defined)	**COVERED**
non-CD3 bispecific	15	3	0	—	**COVERED**
CD3 T-cell-engager	12	4	1	0.886	**COVERED**
single-domain / VHH binder	5	2	1	—	**PARTIAL**
RNA / vaccine / nucleic-acid	1	0	0	—	**THIN** *
peptide / mimotope	1	0	0	—	**THIN** *
diagnostic reagent (IHC/probe)	1	0	0	—	**THIN** *

By direction: cytotoxic-depleting and immunomodulatory are COVERED; **`degrader_or_functional` is OPEN (0 families)**; diagnostic is THIN.

* **Corpus-artifact caveat (load-bearing):** the RNA/vaccine, peptide/mimotope and diagnostic THIN cells each hold only **1 US-anchored on-pathway family**, but the discovery pass surfaced *multiple* China/EP/WO RNA-composition (BioNTech), aptamer, cyclic-peptide and IHC families that materialize at bibliographic tier only. These cells are THIN *in the US-anchored corpus*, not necessarily globally — they are the weakest whitespace calls and should be re-tested after ex-US full-text frack. The single-domain/VHH PARTIAL cell (5 families, 2 live) is the most credible "not-yet-COVERED" binder-format lane.

Read: the honest whitespace is narrow. The only OPEN *direction* (degrader/functional) is a mechanistic hypothesis; the only THIN *formats* are corpus artifacts or the VHH-binder lane. The actionable structural room is *inside* the COVERED ADC cell (payload-lane differentiation, §3) and *in time* (the pending frontier, §4).

---

3. The actionable chemical core — the ADC payload/linker lane is the only small-molecule genus

CLDN18.2 is an antibody target: **binders are fenced by sequence (CDR sets), not by small-molecule genus.** Of the 24 genus-bearing families corpus-wide, **the ADC linker-payload chemistry is the only compiled small-molecule Markush genus** — and even there the genus is *binder-locked*: the canonical formula is `A-(L-T)n` (antibody `A` fixed to a specific CDR set; linker `L`; toxin/payload `T` specified at *class* level; DAR `n`). So the ADC "chemical fence" is really **(fixed CDR antibody) × (payload class) × (linker)**.

On-pathway ADC genus families (median `family_best` recall 0.333, a lower bound; all pending):

family	assignee	payload chemotype	recall	binder-locked
US-20240131178-A1	LegoChem / Nona	camptothecin-like (open payload)	0.828	yes
US-20230054458-A1	Jiangsu Hengrui	exatecan/camptothecin (Pc-L-Y-D)	0.631	yes
US-20250073347-A1	Fortvita Biologics	maytansinoid + camptothecin	0.526	yes
US-20240024498-A1	Jiangsu/Shanghai Hengrui	exatecan (GGFG linker)	0.468	yes
US-20240252670-A1	Beijing Synthetic Vaccine / Tsinghua	exatecan/camptothecin	0.333	no
US-20230338565-A1	Sichuan Kelun	camptothecin/topo-I-like	0.265	yes
US-20250333498-A1	Bio-Thera	anthracycline-aglycone (Val-Cit-PAB)	0.240	no
US-	**SOTIO**	**anthracycline**	0.214	yes

family	assignee	payload chemotype	recall	binder-locked
20240100180-A1				
US-20240409628-A1	Evopoint	**topoisomerase-I inhibitor**	0.204	yes

Where to build in the ADC cell: - **The payload centre of gravity is the topoisomerase-I / exatecan / DXd-class camptothecin lane** — it appears in the large majority of CLDN18.2 ADC genera (Kelun, Hengrui ×2, Fortvita, Beijing Synthetic Vaccine, LegoChem, Evopoint). This lane is **crowding fast** (all pending, but many filers). - **Non-topo-I payloads are thinner.** SOTIO's **anthracycline** ADC (US-20240100180) is a differentiated chemotype; a maytansinoid lane appears only in combination. A CLDN18.2 ADC on a *non-camptothecin* payload (anthracycline, novel-MoA payload, immune-agonist payload) is the clearest in-cell structural differentiation. - **All 9 on-pathway ADC genera are pending** — the payload/linker *grants* have not consolidated; recall 0.20–0.83 (lower bound) means each genus still leaves in-scaffold analog room, and the binder-lock means a *different CDR* antibody with the same payload class is a design-around the genus does not reach.

(Native KC2/KC3 for these families are in the KC evidence mapping, Part II of the combined report. Candidate generation, not FTO.)

---

4. Trajectory — where this target is going (REAL filing dates)

Composed from real `filing_date` on all 106 on-pathway families (no expiry proxy; `data/real_filing_dates.json`).

era (real filing yr)	n	live grants	dominant formats	entrant origin (committed mapping)
pre-2017 (foundational)	5	5	naked-mAb (4) + first ADC (1)	**EU** (4, Ganymed/TRON) + Japan (1)
2017-2019	12	9	naked-mAb (8) + first CD3-bispecific + first VHH	China (7), US (3), EU (1), Japan (1)
2020-2022 (the surge)	**70**	17	naked-mAb (25), non-CD3 bispecific (13), ADC (11), CD3-TCE, CAR	**China (56)**, EU (4), US (4), Japan (3), Korea (2)
2023-2026 (frontier)	19	**2**	**CAR/cell (7), CD3-TCE (5), ADC (3)**	China (11), US (5), Japan (1), Korea (1)

Trajectory statement: *CLDN18.2 filing exploded off a small EU-origin foundational naked-mAb base (Ganymed, 2006-2013) into a 2020-2022 surge (70 of 106 on-pathway families, 66%) that is **overwhelmingly China-origin (56 of 70)** and format-diversified — naked-mAb still leading, but non-CD3 bispecifics (CD47/4-1BB/CD40/PD-L1), ADCs and CD3-engagers filling in fast. The 2023-2026 frontier has **pivoted to cell therapy (CAR-T/CAR-NK) and CD3***

***T-cell-engagers**, and — decisively — **these newest cells are almost entirely pending (2 live grants of 19 families).** Direction of travel: naked-mAb → format diversification (bispecific/ADC) → T-cell-redirection/cell-therapy; geography: EU-founded → China-dominated; maturity: the surge is consolidating into grants now while the redirection frontier remains contestable.*

The decisive dynamic — this estate is YOUNG, not decaying. Unlike a mature target with a lapsing first wave, CLDN18.2 has **0 lapse/discontinuation events among the 106 on-pathway families** (`data/legal_decay.json`); the foundational Ganymed families are all still maintained (US-8425902 → 2028, US-9433675 → 2033). The dynamic is not decay-driven freedom but **grant-consolidation timing**: 33 of 106 are already granted-active, but the newest, most strategically interesting cells (CAR/TCE, 2023-2026) are pending. The window is *before those grants issue*, not *after old ones lapse*.

---

5. Combination adjacencies + cross-class transfer

- **Combination interest (from the 79 off-pathway platform co-claims):** CLDN18.2 is most co-claimed with **CD3** (5 T-cell-engager backbones), **4-1BB / CD40 / OX40** (co-stim), **CD47/SIRPα** (innate checkpoint), **PD-L1/PD-1** (checkpoint), **IL-21/IL-12/IL-10** (cytokine-armed cells), and platform CAR backbones (Allogene CACCR, Gracell/Regeneron TCR-fusion) that list CLDN18.2 among many targets. **Unclaimed rational pairs** worth testing: CLDN18.2 × a *non-CD3* T-cell-redirection co-stim in a single molecule for the pending-frontier cells; CLDN18.2 ADC × checkpoint (the ADC families are mostly monotherapy-drafted).
- **Cross-class transfer (candidate generation; 5,324 genus-overlap edges, INDETERMINATE-dominated + 642 sequence-identity edges, lower bounds):** the ADC payload/linker genera overlap heavily with *generic ADC-platform* chemistry filed for other targets (host families e.g. US-20180237453 scaffold hosts) — proven linker-payload chemistry structurally adjacent to CLDN18.2 ADC space but not claimed *for* CLDN18.2. The single **molecular-glue degrader ADC payload** (Orum US-20230338564, off-pathway) is the clearest transfer hypothesis into the OPEN `degrader/functional` direction: proven degrader-conjugate chemistry, not yet aimed at CLDN18.2. (All INDETERMINATE-driven — candidate generation, not FTO.)
- **The dominant fence is sequence, not genus.** 642 sequence-identity edges (mostly `candidate`, `assignee_basis=different`) confirm the binder landscape is a dense CDR-similarity thicket — the real design-around problem for a new CLDN18.2 binder is *epitope/CDR* freedom, which this structural layer flags but does not adjudicate (not FTO).

---

6. Knowledge graph

`data/cldn18_2_knowledge_graph.json` — **379 nodes, 1027 edges** (Cytoscape / networkx / Neo4j-loadable; **0 dangling edges — asserted at build time**). Node types: target,

engagement_mode (6), direction (4), modality_format (9), indication (6), **disease** (5, Wikidata-grounded: stomach cancer Q189588, gastric adenocarcinoma Q18556047, GEJ adenocarcinoma Q18556563, pancreatic cancer Q212961, cholangiocarcinoma Q124292), assignee (76), patent_family (185: 106 on-pathway + 79 off-pathway co_claims), year (11), **external_publication (76 — the scope-neighbor targets, materialized as real nodes so no edge dangles)**. Edge types: `engaged_via` (6, target→mode) + `engages_via` (106, family→mode), `directed_as`, `uses_modality`, `targets_indication`, `assigned_to`, `filed_in_year`, `co_claims` (79 off-pathway combination signal), `genus_overlaps` (105, capped/family, lower bound), `sequence_identity` (196, lower bound), `indication_is_disease` (5, Wikidata; sourced from real IND→DISEASE nodes). **Two KG build bugs were caught by adversarial review #2 and fixed:** (1) an `indication_is_disease` kwarg-collision that had overwritten the edge source with the literal "wikidata", and (2) 301 dangling scope-neighbor edges pointing at uncreated `PUB:` targets — both corrected, with a build-time assertion now guaranteeing 0 dangling edges (§III of the combined report).

Whitespace/trajectory/consolidation questions become graph queries: *reachable `modality_format` or `direction` node with no incident `uses_modality`/`directed_as` edge from a live-granted family* → the OPEN degrader direction + THIN cells; families by real `filed_in_year` projected on `uses_modality` → the trajectory pivot to CAR/TCE; `co_claims` neighbourhood → the combination adjacencies.

Method & limits (full version in the combined report, Part III)

- **Denominator = 106 on-pathway families** after the load-bearing on/off-pathway screen: a family whose **title or native scored-claim summary** recites claudin/CLDN/18.2 is ON-pathway; a family that co-mentions CLDN18.2 elsewhere in its estate but whose scored claim recites a generic platform (CAR backbone, cytokine receptor, CD3 arm, bare ADC linker, manufacturing) *without* reciting CLDN18.2 is OFF-pathway (79 families; retained only as combination/transfer signal). Each family carries a `scope_evidence` string (`data/families_scoped_full.json`). Adversarial review #2 verified the screen has **0 false-positive keeps and 0 false-negatives** on spot-check (every ON family's CLDN recital is a genuine CLDN18.2 claim; every OFF family is a genuine platform co-mention). *(For 19 ON families the CLDN token is in the title, which for these dependent claims narrows a CLDN18.2 independent claim — a genuine on-target keep.)*
- **The `degrader_or_functional` OPEN call is disclosed as partly classifier-construction:** the binner has a degrader/PROTAC/molecular-glue/functional-modulation keyword path (so a degrader family *would* route into that bin), and **0 of 185 families matched** — but a cell-surface tight-junction-protein degrader is also a mechanistically-hard, corpus-unpopulated hypothesis. So its "OPEN" is *tested-and-absent in this corpus*, but remains the weakest whitespace call (verify against a dedicated degrader search + the Orum molecular-glue transfer candidate, §5).

- **Wikidata grounding:** CLDN18 gene Q18040098 / protein Q21100780 confirmed via SPARQL. **P2293 gene→disease = NONE** — CLDN18.2 is a tumor-associated antigen, not a disease gene; the indication axis is therefore anchored on *somatic over-expression* biology (literature) with Wikidata disease QIDs used only for the KG disease nodes. This is the correct, honest handling for an antigen target (no genetic-association axis to claim).
- **Real filing dates** on all 106 (no expiry−20 proxy).
- **Not FTO, not calibrated.** Whitespace = absence of strong structural claim coverage (necessary-not-sufficient). Genus recall / sequence identity are lower bounds. The `degrader` OPEN call and the RNA/peptide/dx THIN calls carry explicit corpus/axis caveats.
- **US-anchored corpus.** The large China/EP/WO ex-US tail is bibliographic-only; ex-US full-text frack is the next run's job and would most affect the THIN-cell calls.

Data: `families_scoped_full.json` (185, on/off-pathway + 4-axis bins + scope_evidence), `families_onpathway_enriched.json` (106 + genus recall + real filing dates), `whitespace_by_format.json`, `whitespace_thin_cells.json`, `trajectory.json`, `legal_decay.json`, `genus_layer.json`, `scope_neighbors.json`, `real_filing_dates.json`, `ontology.json`, `cldn18_2_knowledge_graph.json`. Every whitespace/trajectory number is computed on the 106 on-pathway families on real dates. Candidate generation, NOT freedom-to-operate.