

IL6_V4_Whitespace_and_Trajectory

Table of Contents

IL-6 (Interleukin-6) — V4 Whitespace and Trajectory Analysis.....	1
Read-me guardrails	1
I. Scope ontology.....	2
The therapeutic / diagnostic line	3
II. The fence line — the structure that resolves	3
One modality, two fences	3
The incumbents sit on both nodes, and their regulatory clocks have partly run.....	4
The ligand node is the oldest and still the fastest.....	4
The nodes the gate cannot see	4
III. The whitespace map — a null finding.....	4
The least flattering row comes first	4
Occupancy on the resolved axes.....	5
OPEN cells — and why none of them is an answer	5
IV. Trajectory vector	5
V. The overlap fences — what was measured and what it means	6
VI. Limits	7
VII. Knowledge graph	7
VIII. What the next run should do.....	7

IL-6 (Interleukin-6) — V4 Whitespace and Trajectory Analysis

Target: IL-6 / interleukin-6 pathway (ligand, receptor, gp130 transducer; gp130 in scope, other IL-6-family cytokines out, by operator ruling) · **Corpus:** 6338 publications across 1682 families · **Registry snapshot:** 2026-09-19

Read-me guardrails

Read these before any number below.

- **This is candidate generation and structural occupancy, never freedom to operate.** Nothing here is a clearance opinion. A cell described as OPEN is a

statement about what this corpus, on these axes, could resolve — not a statement that anyone may build there.

- **Genus recall is a lower bound.** Where a genus is reported it understates coverage; it never overstates it.
 - **Sequence identity is not a lower bound and errs in the opposite direction.** It must never share a clause or a sentence with a genus recall figure, because the two carry opposite error senses and a reader who averages them will be wrong in an unsafe direction.
 - **The sequence fence ran to completion and its result is NON-EVIDENTIARY for this target.** The fence state is MEASURED with 1023 edges, and the mandated §6.1a audit found 0 of 237 hit publications on-target (fraction 0.0): the edges are antibody-scaffold-driven, not IL-6-scope-driven, and **no edge in this package may be narrated as IL-6 estate overlap**. The audit verdict outranks the fence state everywhere, including on the figure.
 - **The chemistry fence is UNMEASURED.** Its comparison job never returned and the polling seam itself failed on re-attempt (5 re-calls, each enqueueing a fresh unretrievable job). UNMEASURED is not zero overlap; no statement about chemical-genus overlap is licensed in either direction.
 - **The ZEUS readout is stated only as announced.** Every clinical-result statement in this document comes from the company announcement recorded in `data/zeus_grounding.json`; nothing is extrapolated from it, and a registry status is never treated as a result.
 - **Comparison is not reliance.** Co-location on these axes does not assert claim overlap between any two estates.
 - **Owner names are a floor.** Families whose assignee did not resolve are counted but unnamed — 569 therapeutic families, the largest single slice of the cohort — so every roster in this document is a lower bound on the parties present.
 - **2 mechanism nodes are occupancy FLOORS BY CONSTRUCTION.** Corpus admission required IL-6-pathway vocabulary, so `downstream_jak_stat` and `gp130_transducer` estates that never name the pathway are structurally underrepresented. No whitespace verdict may rest on their sparseness, and none does.
 - **KC (claim-strength) analysis is deliberately OMITTED** under playbook §3.0; the decision is recorded in `data/kc_decision.json` as OMIT. No claim-strength or relative-quality assertion appears anywhere in this document, and none should be inferred from cell occupancy.
-

I. Scope ontology

2 grid axes, defined and operator-gated **before** the corpus was assembled: **mechanism node** (where in the IL-6 pathway an agent acts) and **modality** (what kind of molecule acts there). The ontology declares 4 axes in total — direction and indication are used to separate cohorts rather than to place a family on the map — and the grid axes carry 7 mechanism values and 15 declared modality values; the rendered grid's modality axis has

16 columns, because this analysis's own unresolved bin is carried as a first-class column rather than dropped — $7 \times 16 = 112$ cells. The operator ruling that defines the target boundary is recorded in `data/operator_gate.json`: gp130 is inside the target, the other IL-6-family cytokines are outside, and JAK/STAT inhibitors enter only where their estates name the IL-6 pathway.

The distinction that the entire analysis turns on is **ligand versus receptor**: an antibody against interleukin-6 itself and an antibody against its receptor are different mechanism nodes with different incumbents, different exclusivity clocks, and different indication histories. The term discipline was built before the corpus specifically because "IL-6" appears inside "IL-6R", "sIL-6R" and "IL-6ST", and a substring error here would merge the map's most important boundary.

The therapeutic / diagnostic line

IL-6 is **measured** far more often than it is **neutralised** — it is one of the most common assay analytes in clinical medicine — and an immunoassay antibody is indistinguishable from a therapeutic antibody on a modality axis. A family enters the therapeutic cohort only on **positive matched evidence of therapeutic direction or claimed therapeutic use** — claim-backed where claim text exists (571 of 778 therapeutic families), title- or abstract-derived for the other 207, a split §VI restates as a limit; 337 families that name IL-6 only inside multi-cytokine panels were excluded before cohorting, and 32 on-target families carry a diagnostic-article basis and are set aside.

Of 1069 on-target families: **778 therapeutic**, 32 diagnostic, 128 with no claim scope to test, and 131 whose direction did not resolve. These sum to 1069. A further 2 MIXED families carry on-target and competing-target signal in the same family and are counted and set aside beside the 32 diagnostic and 128 undetermined — never deleted.

II. The fence line — the structure that resolves

The finding of this analysis is a fence line, not a gap. The field concentrates instead of fragmenting, and everything actionable about the map is a property of its fences: which epitope each encloses, who holds it, what is already marketed behind it, and how much regulatory clock has run. The whitespace layer, taken up in §III, is a null finding and is stated as a null, not displayed.

One modality, two fences

The map's centre of gravity is a single modality on two nodes: **139 full-length-antibody families at the ligand** and **102 at the receptor**, with a further 25 receptor families in fragment/scaffold formats; the largest non-antibody presence at the ligand is 50 small-molecule families. Unlike a field that fragments by modality, this one concentrates: the fences here are built from the same kind of molecule, and what separates them is **which epitope the fence encloses and what has already been approved behind it**.

The incumbents sit on both nodes, and their regulatory clocks have partly run

Marketed incumbency is not a receptor-node monopoly: 3 of the 4 licensed reference products sit at the receptor (tocilizumab, sarilumab, satralizumab) and 1 at the ligand itself (siltuximab). 2 of the 4 BPCIA exclusivity floors have already passed — one on each node: tocilizumab's in 2022 and siltuximab's in 2026 — while sarilumab's floor runs to 2029 and satralizumab's to 2032. **These are regulatory floors — the earliest a biosimilar may be licensed — not market loss-of-exclusivity dates**; the Purple-Book mart carries no patent link, and reference-product patent estates routinely push actual entry later. What the floors establish structurally: on both antibody nodes some statutory protection has already expired. What an investigational entrant brings is not a floor-free neighbourhood — it is a fresh clock: an approval starts its own exclusivity term, in whatever indication earns it.

The ligand node is the oldest and still the fastest

Ligand-directed filing runs from 1987 to 2026: 84 families before the millennium, 83 in the following decade — essentially flat — then an acceleration: 119 in the next, and **171 since**, the most active node-era anywhere on the map, decades into the target's history. The receptor node, from 1990, rose the same way and then turned: 17 pre-millennium families, 38 in the following decade, 73 in the next, 61 in the current era — a slight decline exactly where the ligand node accelerates. New filing favours the challenger's node over the incumbents'. A field this old that is still accelerating into its most crowded cell is the trajectory finding of this analysis.

The nodes the gate cannot see

gp130_transducer (87 families, 21 in the current era) and downstream_jak_stat (10 families, earliest 2013) are floors by construction and carry no verdict. ligand_transcript (24 families) and complex_interface (11) are small and genuinely so — they were reachable by the same gate that admitted the antibody estates.

III. The whitespace map — a null finding

The whitespace layer of this analysis returns a null result, and it is stated here rather than displayed: no empty cell on this grid is reportable as opportunity. The grid is mechanism node × modality over the therapeutic cohort only: 112 cells, of which **52 are occupied**.

The least flattering row comes first

195 of 778 therapeutic families have an unresolved modality — the largest single modality value on the map is "could not tell what kind of molecule". That is a statement about this method's evidence floor, not about the field: 3741 of 6338 publications were binned **without any claim text**.

Mechanism, by contrast, shows 0 unresolved families — but that is **construction, not virtue**. Pathway vocabulary was an admission requirement, so every admitted family

necessarily names a node. The honest reading: modality resolution is earned, mechanism resolution is inherited from the gate.

Occupancy on the resolved axes

The field spreads across **6 resolved mechanism nodes** and **13 resolved modality classes** (the occupied modality total is 14, including this analysis's own unresolved bin; quoting it as field diversity would count our failure to resolve as somebody else's spread).

mechanism node	therapeutic families
ligand	457
receptor	189
gp130_transducer (floor by construction)	87
ligand_transcript	24
complex_interface	11
downstream_jak_stat (floor by construction)	10

OPEN cells — and why none of them is an answer

60 cells are empty. **Only 16 are readable as whitespace at all**, and every one of them carries the qualifier COHERENCE_UNSCREENED_OPEN_IS_NECESSARY_NOT_SUFFICIENT: emptiness on this grid is a necessary condition for opportunity, never a sufficient one, because no biological-coherence screen was run and several empty combinations are combinations a molecule cannot inhabit. The remaining 44 empty cells are axis or corpus artifacts under playbook §6.2 — 22 of them sit on an alias axis value that is unreachable by construction — and reporting any of them as opportunity would be the easiest way to mislead a reader of this document. The occupied cells' verdicts close the accounting: 25 COVERED, 19 THIN, 8 PARTIAL, and 16-plus-44 empty as above — every one of the 52 occupied cells carries a COVERED, THIN or PARTIAL verdict. That accounting is the null finding stated plainly: what is empty on this map is empty for a reason this analysis can name, and none of those reasons is opportunity.

IV. Trajectory vector

13 named assets — marketed and investigational — are in scope, and they occupy **3 mechanism nodes**: 7 ligand-directed, 5 receptor-directed, and 1 at the gp130 transducer (selective trans-signalling). 4 of the 13 are marketed products. The clinical field and the patent field agree on the map's shape: the assets crowd the same antibody nodes the estates crowd.

The estates outlive their owners. 11 of the 13 assets have lineage estates naming more than one party — molecules on this target routinely outlast the corporate structures that made them — and for **2 assets the originator holds every corpus family at the asset's node while the trial sponsor holds none**. Attribution here is lineage-level except where

the corpus names the compound (12 of 13 verified by name or code); it must not be read as claim coverage of any asset.

The news event arrived between the registry snapshot and this document. At the snapshot, ziltivekimab's ZEUS trial reached primary completion on 2026-06-09 (a registry date, not a result — neither a COMPLETED nor a TERMINATED registry status is a readout), with 6 Phase III trials verified. On 2026-07-31 Novo Nordisk announced headline results (Company Announcement No 45/2026, recorded with its source in `data/zeus_grounding.json`): ziltivekimab demonstrated IL-6 pathway engagement — the expected reductions in free IL-6 and high-sensitivity C-reactive protein — but did not reduce major adverse cardiovascular events versus placebo (hazard ratio 0.99, 95% CI 0.88 to 1.11.); the announcement stated that the HERMES and ARTEMIS outcomes trials would continue toward anticipated readouts in the first half of 2027. The programme then narrowed again between the announcement and this document: the HERMES registry entry moved to TERMINATED on 2026-09-07 — a discontinuation press reports attribute to a Data Monitoring Committee recommendation — while ARTEMIS remains ACTIVE_NOT_RECRUITING with a registry primary completion of 2027-01-29 (a registry date, not a result). The structural reading does not depend on the result's direction: a cardiovascular readout for a ligand-directed antibody landed on the most crowded, oldest, fastest-filing node of this map — inside a fence whose neighbours include estates that long predate the indication — and the fence line it landed on is exactly where it was.

V. The overlap fences — what was measured and what it means

The sequence fence is MEASURED: of 11 selected sponsor groups, 11 were queried — 2 of them via a recorded substitute publication — yielding 1023 directed identity edges, verdicts 41 own / 976 candidate / 6 indeterminate, with 27 named coverage gaps and none errored. The queried sponsor groups cover 157 of 778 therapeutic families (20 percent) — the fence speaks for the named-sponsor slice only.

And the mandated §6.1a audit disqualifies the layer as overlap evidence. Of 237 distinct hit publications, 0 carry IL-6 vocabulary in the title; within-corpus on-target cross-family pairs: 0. The edges trace shared antibody scaffolds, constant regions and framework sequences — chemistry every antibody carries — so the honest statement is that this target's sequence layer measures **antibody-ness, not IL-6-ness**; the artifact's evidentiary flag is recorded as false and every downstream surface honours it. This is the third antibody-bearing V4 target in a row whose sequence layer shipped non-evidentiary, and the recurrence is the finding.

The chemistry fence is UNMEASURED: 83 therapeutic families carry extractable compounds and a comparison batch of 8 publications was submitted, but the job never returned and the poll seam itself proved broken on late re-attempt. No genus-overlap number exists in this package.

VI. Limits

Stated plainly, because several of these are large.

1. **The sequence layer is measured but non-evidentiary (§V).** No overlap conclusion survives the audit, and none is drawn.
 2. **The chemistry fence is UNMEASURED.** Estate-to-estate chemical overlap is outside this document's reach.
 3. **3741 of 6338 publications (0.59 of the corpus) carry no claim text;** 37 rest on an LLM claim summary alone, against 2560 with real claim text. At family grain, 571 therapeutic families have claim-level evidence and 207 rest on title/abstract only. Claim scope for the latter is unread.
 4. **463 on-target families rest on a title or abstract mention with nothing in claim scope.** They are counted as present in the field, not as evidence of what is claimed.
 5. **341 discovery rows named a US document that resolved to no publication number, and 128 families (2 further single publications) failed ingest deterministically.** They are named absences: absent from every count in this document, present in the Stage B record.
 6. **400 of 778 therapeutic families have no readable legal status.** Unreadable is not not-live: the liveness split (109 granted-active / 218 pending-only / 51 all-dead) speaks only for the families it could read.
 7. **Discovery searched title and abstract only.** Every count in this document is a floor.
 8. **The assignee floor is the map's largest caveat:** 569 therapeutic families — the single largest slice — have no resolvable assignee group, so every roster and every per-sponsor statement is a lower bound.
-

VII. Knowledge graph

data/il6_knowledge_graph.json carries 927 nodes and 2986 edges over the target, mechanism nodes, modalities, cells, families, assignee groups and clinical assets.

The graph is validated on **properties, not totals**: 0 self-edges, 40 duplicate triples removed at build, 0 dangling endpoints, and counts derived from the serialised collections so a published headline cannot drift from the artifact. 2 axis values are isolated — both are unoccupied-axis **artifacts** (the declared catch-all and the unreachable alias), not whitespace, and the graph says so in its own integrity note.

VIII. What the next run should do

- **Resolve the assignee floor.** 569 therapeutic families with no assignee group is the single largest hole in every roster, every fence coverage figure, and every ownership statement.

- **Attack the modality catch-all.** 195 families are unresolved for want of claim text; targeted claim extraction would do more for this map than any new discovery shard.
- **Build a sequence protocol that survives antibody targets.** The scaffold-contamination failure is now reproduced on every antibody-bearing target it has been tried on; the fix is a CDR-level (complementarity-determining-region) comparison, not a longer alignment.
- **Re-measure the chemistry fence when the compare seam is repaired,** starting from the 83 compound-bearing families already identified.
- **Screen the 16 readable OPEN cells for biological coherence** before any of them is quoted as opportunity.