

## Menin–MLL — V4 Structural-Overlap Report

One foundational genus read against the field — and the field turns out to be half the anchor's own estate

**Target:** menin–MLL protein–protein interaction — small-molecule inhibitors of the menin (MEN1) / MLL (KMT2A) interaction, a validated dependency in KMT2A-rearranged and NPM1-mutant acute leukemias. **Anchor:** Janssen's spiro-bicyclic thienopyrimidine genus, US-11220517-B2. **Revision date:** 2026-08-11. **Supersedes:** the 2026-07-26 edition. **Basis:** the July anchor-vs-field overlap measurements, re-attributed and re-partitioned against the 408-family August corpus.

**Read-me (the guardrails, once).** This is **candidate generation and a structural-overlap signal — NOT a freedom-to-operate opinion**. Verdicts are three-valued: IN, OUT, INDETERMINATE. INDETERMINATE is not a probability and not a clearance; it means the boundary did not resolve by this method and is where counsel earns its fee. Structural overlap and genus recall are directional **lower bounds**. **OUT is never “free to operate.”** Nothing here is an infringement read on any named party or asset.

**What changed in revision 2, and what did not.** The **chemistry has not been re-measured** — every IN/OUT/INDETERMINATE verdict is the same July genus-overlap measurement. What changed is **attribution**: the July edition drew 9 of its 19 field genera as “unattributed” and 1 as “China-origin”, and could not tell a competitive boundary from the anchor's own continuation estate. All 19 are now resolved to named owners at family level. That single change moves the report's headline.

---

## PART I — ACT ON THIS

### 1. The correction that matters

The July edition reported that the anchor genus “resolves cleanly OUT against a handful of rival genera, **IN against one**, and INDETERMINATE against the large majority,” over **19 confirmed field genera** and **36 directed comparisons**.

Three of those framings do not survive attribution:

| July edition                             | revision 2                                                                                                   |
|------------------------------------------|--------------------------------------------------------------------------------------------------------------|
| 19 field “genera”                        | <b>11 distinct INPADOC families</b> — 19 publications, family-deduplicated                                   |
| sponsors: 9 unattributed, 1 China-origin | <b>all named:</b> Janssen ×4 families (3 counterparty + the anchor), Ventyx ×3, Medshine/HelioEast, Scripps, |

| July edition                                                           | revision 2                                                              |
|------------------------------------------------------------------------|-------------------------------------------------------------------------|
|                                                                        | Origenis, Kyowa Kirin                                                   |
| “one rival genus reads as plausibly within the anchor’s scope” (IN ×1) | that IN is <b>Janssen’s own family</b> . <b>Cross-sponsor IN = 0</b>    |
| one tally over 36 directed cells                                       | <b>two tallies</b> , because half the field is the anchor’s own sponsor |

**3 of the 10 counterparty families are Janssen** — the anchor’s own sponsor — and the anchor’s own family makes 4 of 11 in total. Of the 22 materialised directed edges, **12 are Janssen-vs-Janssen**.

Re-partitioned:

| relationship                                               | OUT      | IN       | INDETERMINATE | edges |
|------------------------------------------------------------|----------|----------|---------------|-------|
| <b>Cross-sponsor — a real competitive boundary</b>         | <b>3</b> | <b>0</b> | <b>7</b>      | 10    |
| Same sponsor (Janssen) — <i>not</i> a competitive boundary | 2        | 1        | 9             | 12    |

## 2. What this buys the reader

**The competitive field is smaller and more uniformly unresolved than it looked.**

Against genuine third parties the anchor resolves cleanly **OUT in 3 directed readings** and lands **INDETERMINATE in 7** — and it reads **IN against nothing**. The July framing invited a reader to believe a rival’s chemistry had been found inside a Janssen fence. On this attribution, none has.

**Practical consequences:**

1. **Do not price the single IN as competitive risk.** It is Janssen reading against Janssen — a portfolio-shape fact, not a boundary anyone must clear.
2. **The named third parties are the map worth having.** Ventyx Biosciences appears with **three separate families**, and Medshine/HelioEast, Scripps, Origenis and Kyowa Kirin with one each. That is a sponsor list a BD&L reader can act on; “unattributed” is not.
3. **The clean OUTs are all cross-sponsor** — Kyowa Kirin, Medshine and HelioEast, read anchor-as-query. Those are the three boundaries a reader can most defensibly read without an opinion, and they are exactly the ones the July tally buried among the Janssen self-comparisons.

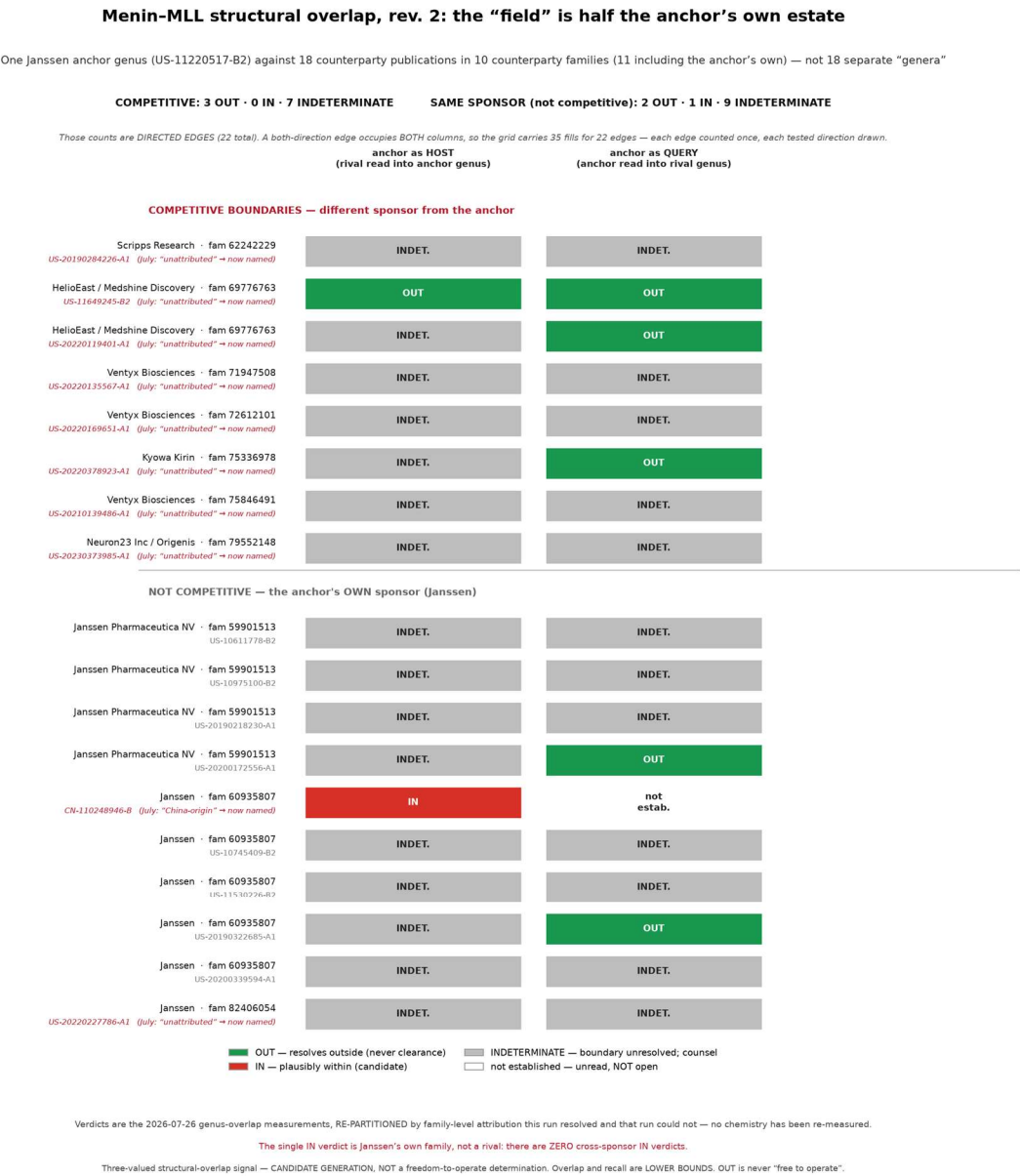
## 3. Direction still matters, and members disagree

The asymmetry finding stands and is strengthened. Family 60935807 alone carries **IN, OUT and INDETERMINATE across its five members** — which is why revision 2 keeps one row per publication inside family groups rather than collapsing each family to a single verdict.

Collapsing would have forced a pick, and the first draft of this figure did exactly that and silently overwrote the conflicts.

Read the direction that matches your question: anchor-as-host answers “does their chemistry fall in our genus?”; anchor-as-query answers “does our chemistry fall in theirs?”

4. The map



Menin-MLL structural-overlap exhibit, revision 2

Figure 1. The anchor genus against 11 distinct counterparty families, split into two blocks: competitive boundaries above, the anchor’s own sponsor below. Rows carry the resolved owner and family id, and rows the July edition drew as “unattributed” or “China-origin” are

*marked with the label they replace. Fills are categorical — IN, OUT, INDETERMINATE and an explicit “not established” — never a gradient, because a gradient would manufacture precision the three-valued reading does not have. The two blocks are tallied separately and the figure’s tallies are asserted equal to the underlying artifact at build time, so the figure cannot print a number the data does not carry. Per-edge structural-similarity values are internal and are suppressed in this public build by default. Candidate generation, not a freedom-to-operate determination.*

---

## PART II — METHOD & LIMITS

### The anchor and its quality — unchanged

US-11220517-B2 compiles cleanly: **precision 0.995** across **507 enumerated member structures**, **recall lower bound 0.706**, validation F1 0.826. Precision 0.995 is what lets an OUT be leaned on — when this genus says a structure is outside the claim, it is rarely wrong about that. Recall as a lower bound is the honest counterpart.

### How revision 2 was built

- **Attribution by direct lookup, at family level.** Each of the 19 July field publications was resolved to its assignees and INPADOC family. Ownership is taken at **family** level because a member can carry no assignee of its own: CN-110248946-B is the worked example — publication-level attribution leaves it “unattributed” and counts it as a rival, while its family 60935807 is Janssen.
- **Family deduplication.** 19 publications → 11 families. The July figure’s 36 directed cells therefore counted several families more than once. This is the same publication-versus-family conflation that the August corpus run found in its own ingest (451 card-present keys → 408 families; full corpus funnel: 586 discovery rows → 541 submitted → 453 collected → 451 card-present → 408 distinct families → 176 on-pathway → 165 gridded), from the same cause: an id that looks like a family id and is not.
- **Sponsor partition.** Same-sponsor and cross-sponsor edges are tallied separately. They answer different questions and must not share a denominator.
- **No re-measurement.** No genus was recompiled and no overlap re-run. Revision 2 is an attribution and partition correction over the July measurements.

### Limits — read before relying

- **Not FTO.** Three-valued verdicts are candidate generation. INDETERMINATE localises a boundary for counsel; OUT is never clearance.
- **Overlap and recall are lower bounds.** Recall 0.706 on the anchor is a floor.
- **“Not established” is uninformative, not open.** Families with no materialised comparison to the anchor are **unread** against it — neither outside nor whitespace.

- **Same-sponsor is not free ground.** A Janssen-vs-Janssen edge is not a competitive boundary, but Janssen still holds both genera; a third party clears neither.
- **Attribution is assignee-of-record, not beneficial ownership.** Licences, options and unrecorded assignments are invisible to it. Ventyx's three families are three *filings*, not a verified single programme.
- **One anchor, not a matrix.** This remains a star read from a single genus, not an all-pairs matrix. A full matrix is still not available: on the August corpus only **13 of 176 on-pathway families have a readable Markush fence** (66 degraded, 97 with no genus headline), so genus compile quality — not corpus size — remains the binding constraint.
- **Corpus-bounded.** The read is relative to the ingested V4 menin estate; genera without a materialised overlap edge are not represented.

### What a next revision should do

1. **Re-run the overlap with the anchor against the 7 cross-sponsor families only**, and drop the same-sponsor edges from the headline tally entirely.
2. **Attempt genus compiles on the Ventyx, Kyowa Kirin and Medshine/HelioEast families** — the competitive boundaries are where compile effort pays.
3. **Verify Medshine and HelioEast**, which share family 69776763 under two different assignees of record; that is either co-ownership or an assignment the register recorded twice, and it changes the sponsor count.

---

*Revision 2 restates the menin-MLL anchor-versus-field structural-overlap read with all 19 field publications attributed and family-deduplicated. Candidate generation, NOT freedom-to-operate.*