Epigenome — V4 Whitespace & Trajectory Analysis

Target: Epigenome — programmable epigenome editing (non-DNA-cutting) + small-molecule epigenetic drugs. **Date:** 2026-07-23 · **Substrate:** 95 on-target families (expanded whitespace-rerun corpus; innovator-core + exhaustive gap-field re-ingest; filing span 1989→2025). **82 core-chemistry + 13 combination/method adjacency; 63 KC-scored + 32 bibliographic-only-deferred.**

> **Read-me (guardrails):** 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 for opportunity). Genus recall is a **lower bound**; native KC is **uncalibrated** (relative only). This rerun **exhaustively re-ingested the round-1 gap fields** (BET, DOT1L, degraders, editing-recruiters) — coverage is now far better than round-1, so the "load-bearing = only trajectory" framing of round-1 no longer applies; the occupied-cell claim-strength read is now defensible. **Named coverage limits still stand:** (1) 32 biblio-deferred families carry no KC (coverage, not strength); (2) the **degrader modality was NOT exhaustively ingested** — HDAC/LSD1/EED/TET degrader families are null-pubnum (found, un-ingestable) or their discovery legs OPS-errored; (3) commercial editing estates that are US applications with NULL publication numbers (Tune/Chroma/nChroma/some Sangamo) remain under-counted; (4) TET is genuinely thin (1 family). Every empty cell is therefore a **COVERAGE_LIMITED or NOT_SEARCHED lower-bound signal — NOT whitespace and NOT clearance.**

---

1. The scope-coordinate system

Four axes, externally grounded (search_terms HGNC/UniProt/OpenTargets + literature + **Wikidata SPARQL**), human-gated — NOT derived from the patent set. Full definitions: `data/ontology.json`.

- **mechanism_node (11):** PRC2-EZH2 · PRC2-EED · DOT1L · menin-MLL (writers/writer-complex) · HDAC · LSD1-KDM1A · TET (erasers) · BET-BRD4 (reader) · DNMT (writer) · programmable-editor-effector (dCas9/ZF-TALE recruitment) · chromatin-remodeler-other.
- **direction:** inhibitor · recruiter-activator · recruiter-repressor · degrader · diagnostic.
- **modality/chemotype (16):** hydroxamate/benzamide/cyclic-peptide/pyridone/indazole/spiro/cyclopropylamine SM; azanucleoside; bromodomain-heterocycle; PROTAC-degrader; dCas9-effector-fusion; ZF/TALE-effector-fusion; guide-RNA-system; AAV/LNP-delivery; biologic-antibody; dx-methylation-signature.
- **indication:** IO-solid · hematologic-malignancy · MLL-KMT2A-leukemia · lymphoma · MDS-MPN · autoimmune · neurodevelopmental-syndrome · cardiometabolic · fibrosis-osteoarthritis · **monogenic-durable-silencing** · CNS-neuro · infectious-latency.

Wikidata grounding (P2293 gene→disease genetic associations, confirmed via SPARQL): EZH2 (Q17917770)→Weaver syndrome; DNMT3A (Q17914047)→AML/MDS/Crohn's/Tatton-

Brown-Rahman; MEN1 (Q14913035)→MEN type 1; KDM1A/LSD1 (Q18036666)→CPRF developmental syndrome; DOT1L (Q18047604)→osteoarthritis. These anchor the indication axis and the KG disease nodes.

2. Whitespace map — the coverage grid (rerun)

Projected onto mechanism_node × direction. This rerun **exhaustively re-ingested the round-1 gap fields** — ~215 families discovered, ~100 newly ingested — collapsing the round-1 "empties" that had been sampling artifacts: **BET 0→43, editing-recruiter 3→14, degrader 1→9, DOT1L 0→9**. The grid now separates five distinct empty-cell reasons so absence is never silently read as opportunity.

Verdict rule: `COVERED` = broad (≥0.5 recall *lower bound*) live-granted genus present · `COVERED_SINGLE_FAMILY` = one broad live-granted family · `PARTIAL` = coverage but no broad live-granted genus · `THIN` = ≤2 scored-core & no broad live genus · `DEFERRAL_GAP` = families present but all bibliographic-deferred (occupied, NOT whitespace) · `ADJACENCY_ONLY` = held only by a tool/combination adjacency. Empty cells are classified `COVERAGE_LIMITED` (field found but un-ingestable / OPS-skipped) · `NOT_SEARCHED` (no discovery leg issued) · `IMPLAUSIBLE_BY_MECHANISM`. Lapsed / `*_CONTESTED` patents are propagated OUT of live-grant counts.

mechanism cell	verdict	families	core	scored-core	live grants	broad genus	max recall
BET-BRD / inhibitor	**COVERED**	38	32	18	9	3	0.773
EZH2 / inhibitor	**COVERED**	7	5	5	3	1	0.57
LSD1-KDM1A / inhibitor	**COVERED**	4	4	4	4	2	0.515
menin-MLL / inhibitor	**COVERED**	4	4	3	3	1	0.716
DNMT / inhibitor	**COVERED_SINGLE_FAMILY**	1	1	1	1	1	0.857
HDAC / inhibitor	**PARTIAL**	8	6	6	2	0	0.903
editor-effector / recruiter-repressor	**PARTIAL**	13	13	9	2	0	—
DOT1L / inhibitor	**THIN**	5	4	2	0	0	0.852
BET-BRD / degrader	**THIN**	4	4	1	0	0	0.554
EZH2 / degrader	**THIN**	1	1	1	1	0	0.289

mechanism cell	verdict	families	core	scored-core	live grants	broad genus	max recall
PRC2-EED / inhibitor	**THIN**	1	1	1	1	0	0.136
BET-BRD / recruiter-repressor	**THIN**	1	1	1	0	0	—
editor-effector / recruiter-activator	**THIN**	2	2	2	1	0	—
DOT1L / degrader	**DEFERRAL_GAP**	4	4	0	0	0	—
TET / inhibitor	**ADJACENCY_ONLY**	1	0	0	0	0	—
editor-effector / diagnostic	**ADJACENCY_ONLY**	1	0	0	0	0	—

2b. The empty cells — 0 genuine-open; every empty is COVERAGE_LIMITED or NOT_SEARCHED

There are `0` genuine-open cells (`n_genuine_open = 0`). The unpopulated node×direction cells resolve into:

- **4 `COVERAGE_LIMITED`** — the degrader modality, found but un-ingestable: **HDAC-degrader** (5 Dana-Farber Class-IIa / U-Leicester families, **all pubnum=null**, verified un-ingestable 2026-07-23), **LSD1/KDM1A-degrader** (OPS upstream error, skipped), **PRC2-EED-degrader** (OPS-skipped / null-pubnum), **TET-degrader** (never searched with a resolvable result). Families demonstrably exist; the field was simply not sampled.
- **19 `NOT_SEARCHED`** — recruiter-activator / recruiter-repressor conjugate directions across DNMT/EZH2/EED/menin/HDAC/LSD1/DOT1L/BET/TET (mechanistically possible, cf. US-2019343873-A1, but no distinct discovery query was issued), plus DNMT-degrader and menin-degrader (no leg issued; both are active real-world areas).
- **2 `IMPLAUSIBLE_BY_MECHANISM`** — programmable-editor-effector as small-molecule *inhibitor* or PROTAC *degrader* (incoherent mechanism).

This CORRECTS the round-1 retraction — the right way. Round-1 mis-called the recruiter/editing directions "genuine whitespace," then retracted on the Stage-F audit. This rerun does neither: it **declines to assert whitespace on an incompletely-sampled modality.** The degrader field is not empty-because-open; it is empty-because-un-ingestable (null-pubnum) or unsearched. Absence of evidence here is **evidence of a coverage hole, not of an opportunity.**

Read: with the gap fields now deeply ingested, the **load-bearing, defensible outputs** are the OCCUPIED-cell claim-strength read (BET-inhibitor COVERED; menin/LSD1/EZH2/DNMT-inhibitor COVERED; DOT1L-inhibitor and editing-recruiter-repressor coverage) plus the

lapse/decay (§5) and trajectory (§4) signals. No cell can be affirmatively called clean whitespace on this corpus — but that conclusion now rests on **exhaustive occupation**, not on under-sampling.

3. The chemical core — 34 genus-bearing families

Of the 63 KC-scored families, **34 carry a compiled Markush genus**; **median family_best recall = 0.412, mean 0.450** (lower bounds) — compiled claims capture under half the enumerable analog space they gesture at, leaving in-scaffold room even in COVERED cells. What the *live* structural coverage looks like (this is a claim-strength read of what exists, NOT an FTO/clearance statement — verify any specific family before relying):

- **BET-BRD inhibitor (COVERED — now the largest cell, 38 families / 32 core, recall up to 0.773 Hinova).** The deep gap-field re-ingest turned round-1's empty BET cell into the corpus's most heavily occupied one: GSK (I-BET762/molibresib), OncoEthix (OTX015), Celgene Quanticel, Hinova, Resverlogix/Zenith (apabetalone), plus a large CN apabetalone process/formulation tail (16 biblio-deferred). Three broad live-granted genera.
- **menin-MLL inhibitor (COVERED — Janssen recall 0.716, Kura ziftomenib Orange-Book-linked).** The best-drafted, newest node. Crowded and fresh.
- **HDAC inhibitor (COVERED but DECAYING — recall up to 0.903 entinostat co-crystal).** Foundational composition is **lapsed/expired** (romidepsin expired; SAHA lapsed; oxadiazole lapsed 2019). Live coverage is *formulation/co-crystal/combination* (Solipharma, Onxeo, Syndax, Acrotech), not broad composition — the broad first-generation chemical genus is no longer defended (a lapse/decay observation, §5; not a clearance call).
- **EZH2 inhibitor (COVERED — tazemetostat marketed; GSK126 lapsed 2018).** Five tazemetostat salt/polymorph forms (Eisai) + a CN/Korea follow-on cluster (Tarapeutics, Ancureall, Pharmablock, Jiangsu Hengrui). Recall moderate (≤0.57). EED-allosteric and EZH2-degrader appear only as single thin families here (recall 0.136 / 0.289) — too sparse in this corpus to characterize.
- **LSD1 inhibitor (COVERED, consolidating — Incyte ×2 to 2039).** The newest small-molecule race; broad live grants (Incyte, CSPC, Helioeast); fast-filling.
- **DOT1L inhibitor (THIN) + editing-recruiter-repressor (PARTIAL).** The deep re-ingest added Epizyme/Dana-Farber DOT1L inhibitors (recall to 0.852) and a 13-family programmable editing-recruiter-repressor cluster (Duke, Vanderbilt, Georgia Tech, CHOP, San Raffaele/UCL) — occupied but with no broad live-granted genus (mostly method/composition claims without compiled Markush).

(Per-family patent-craftsmanship strength scores are held in the internal analyst deliverable, not this public edition. Strength scoring is relative-ordering only, uncalibrated.)

4. Trajectory — where this target is going

On **real filing dates** (all 95 families dated; span **1989→2025**; no proxy).

era	n	direction mix	dominant nodes / entrants
pre-2013	5	**5:0 inhibitor**	HDAC (3, romidepsin/SAHA), EZH2 (1), BET (1)

era	n	direction mix	dominant nodes / entrants
(foundational)			— US/Japan pharma
2013-2018 (first-wave SM)	32	28 inhibitor : 3 recruiter-repressor : 1 dx	BET (13), EZH2 (5), editor (4, incl. first Duke editing family), HDAC (4), DOT1L (3), menin (2), LSD1 (1); + first China filers
2019-2024 (recent)	58	36 inhibitor : **9 degrader** : **11 recruiter-repressor** : 2 recruiter-activator	BET (29), editor (12), DOT1L (6, incl. Dark Blue degrader cluster), LSD1 (3), EZH2 (2), menin (2), DNMT/HDAC/EED/TET (1 each); China/Korea entrants rising

Trajectory statement: *Epigenome IP is transitioning from a lapsing first-wave **HDAC/EZH2 inhibitor** foundation (1989-2012, now expired/lapsed) through a crowded 2013-2018 **BET/EZH2 small-molecule inhibitor** build-out, into a 2019-2024 frontier defined by (i) sustained **BET** and **LSD1-inhibitor** filing and (ii) the **genuine emergence of the degrader and programmable-recruiter directions** — 9 degraders (EZH2, DOT1L/Dark Blue, BET/p300/BRM) and 11 recruiter-repressor editing families in the most recent era alone. The center of gravity is still overwhelmingly **occupancy inhibition of the enzyme (69/95)**; but the mechanistic frontier — **degrade or recruit rather than inhibit** — is now a measurable share (degrader 9, recruiter-repressor 14) rather than the single family round-1 saw. Entrants broaden from a US/Japan/EU-pharma core (Eisai, GSK, Fujisawa, Janssen, Kyowa Kirin) toward China/Korea follow-on filers (Jiangsu Hengrui, CSPC, Helioeast, Jing Medicine, Daegu-Gyeongbuk, Tarapeutics) and dedicated degrader/editing biotechs (Dark Blue Therapeutics, Coloma, Weiguang Gene).*

The decisive dynamic — the foundational inhibitor core is being abandoned. Four foundational estates are **lapsed or expired**: romidepsin (US-4977138, expired), vorinostat/SAHA (US-6087367, lapsed), GSK126 EZH2i (US-8846935, lapsed 2018), oxadiazole HDACi (US-8981084, lapsed 2019). The original broad HDAC/EZH2 chemical genera are no longer defended, while the newest nodes (menin, LSD1, EED) and the degrader/editor directions are fresh and accreting. **Direction of travel: inhibitor→degrader/recruiter, global-enzyme→locus-specific-editing, small-molecule-occupancy→targeted-protein-degradation + programmable-epigenetic-control.**

5. Lapse / decay timeline

patent	assignee	node / chemotype	event	date
US-6087367-A	Columbia/Sloan-Kettering	HDAC / vorinostat-SAHA	LAPSED (non-payment)	pre-2020
US-8846935-B2	GlaxoSmithKline	EZH2 / GSK126 indazole	LAPSED (non-payment)	2018-11-27
US-8981084-B2	Tempero/Constellation	HDAC / oxadiazole	LAPSED (non-payment)	2019-05-14

patent	assignee	node / chemotype	event	date
US-4977138-A	Fujisawa/Astellas	HDAC / romidepsin depsipeptide	EXPIRED (full term)	pre-2020
US-11208382-B2	Hangzhou Solipharma	HDAC / entinostat co-crystal	lapse/discontinuation events (verdict field still granted_active)	2026-02 (flagged)

Data-quality flag (propagated) — 7 `granted_active_CONTESTED` families. The rerun surfaces **seven** families whose exported `verdict == granted_active` is contradicted by negative/lapse timeline events (the "lapsed still granted_active" failure mode). All seven are treated as **`granted_active_CONTESTED`** and propagated OUT of live-grant counts in §2:

publication	family	node	conflicting events
US-11214561-B2	62978081	EZH2	lapse / discontinuation 2026-02→03
US-11208382-B2	65039316	HDAC (entinostat)	lapse / discontinuation 2026-02
US-8981084-B2	44304637	HDAC (oxadiazole)	lapse / discontinuation 2019-04/05
US-8846935-B2	44904073	EZH2 (GSK126)	lapse / discontinuation 2018-11
US-10987349-B2	62709391	BET	lapse / discontinuation 2025-06
US-9861627-B2	44764570	BET (GSK I-BET)	lapse / discontinuation 2022-02/03
US-11389433-B2	68838966	BET	application discontinuation 2022-05

(`legal_verdict_conflicts` in `../V4_Epigenome/data/epigenome_v4_native_overlay.json`. KC scores are unaffected; the flag is a verdict-vs-timeline reconciliation item, not a scoring change.)

6. Combination adjacencies + cross-node transfer

- **Combination interest (13 non-core adjacency families):** EZH2i × platinum (Pfizer SCLC), EZH2i × BTKi (Jiangsu Hengrui lymphoma), HDACi(entinostat) × anti-CSF-1R antibody (Syndax), HDACi(belinostat) × pralatrexate (Acrotech lymphoma), DNMT-demethylation × chemotherapy (Johns Hopkins), BETi synergy in DLBCL (OncoEthix), RVX-208 HIV-latency repurposing (Southern Medical), BETi adjuvant in IO (Dana-Farber/Moffitt), CBP/p300i × KRAS-G12Ci (Tolremo), ERKi × DOT1Li (Celgene Avilomics), TET-inhibitor × ferroptosis inducer, and BETi × SGLT2i cardiometabolic pairs (Resverlogix ×2). **Unclaimed rational pairs:** epigenetic *degrader* × checkpoint; menin-inhibitor × HMA (the AML backbone) as a *composition*, not just a dosing method; LSD1i × the editing-recruiter modality.
- **Cross-node structural transfer (from scope-neighbor edges, all INDETERMINATE-driven — candidate generation, not FTO):** the JHU DNMT-demethylation genus (US-12060615) and the Tarapeutics EZH2 genus (US-10647700) each carry large markush edge sets and share at least one external scaffold host (US-20220340613); Janssen's menin genus has a high-ECFP edge (0.782) to US-20220227786. These are proven-scaffold transfer hypotheses — chemistry structurally adjacent to an epigenetic node, not yet claimed *for* that node (lower bounds; INDETERMINATE membership).

7. Knowledge graph

`data/epigenome_knowledge_graph.json` — **186 nodes, 611 edges, 0 dangling** (Cytoscape/networkx/Neo4j-loadable). Node types: target_class, mechanism_node, direction, modality, indication, **disease (Wikidata-grounded)**, patent_family, patent_publication, year, legal_event, external_family. Edge types: `acts_on`, `directed_as`, `uses_modality`, `targets_indication`, `part_of`, `filed_in`, `co_claims`, `lapses_on`, `genus_overlaps`, `node_associates_disease` (Wikidata P2293; all 8 disease edges verified to source from real mechanism_node ids).

Trajectory/lapse questions become graph queries: family nodes by real `filed_in` year projected on `directed_as` → the inhibitor→degrader/recruiter direction-of-travel (§4); `mech`→`lapses_on` → the decaying HDAC/EZH2 foundational cells (§5). *(The graph does NOT license a cell-level whitespace claim on this corpus — see §2b: unpopulated cells are sampling gaps.)*

8. Method & limits

- **Substrate: 95 on-target families** (expanded whitespace-rerun corpus; **82 core + 13 adjacency**; **63 KC-scored + 32 bibliographic-only-deferred**). `data/families_scoped.json` (`n=95`). 28 hand-curated round-1 records reused verbatim; 67 newly binned from card summary/genus/assignee/title. Off-target pubs excluded per `off_target_pubs_excluded`.
- **Binning:** LLM claim-scope read of each card summary/genus/assignee (not title); each family carries a `scope_evidence` string; 82 core-chemistry vs 13 combination/method adjacency separated. 7 lapse-conflict pubs flagged legal `*_CONTESTED`. `data/families_scoped.json`.
- **Structural layer:** family_best genus recall (lower bound) from cards; `get_scope_neighbors` edges (all INDETERMINATE — candidate generation). `data/dynamic_signals.json` cross-node transfer edges.
- **Trajectory:** REAL `filing_date` from `get_patent` for all 95 (no expiry proxy; span 1989→2025). `data/dynamic_signals.json`.
- **COVERAGE-BOUNDED — the central limit (rerun):** this rerun **exhaustively re-ingested the round-1 gap fields** (~215 discovered, ~100 ingested), so coverage is now much better than round-1 and the "load-bearing = only trajectory" framing no longer holds — the occupied-cell claim-strength read is defensible. **But specific, named coverage limits remain (and empties are still lower-bound signals, NOT whitespace/clearance):** (1) 32 biblio-deferred families carry no KC (coverage, not strength); (2) the **degrader modality was NOT exhaustively ingested** — HDAC/LSD1/EED/TET degrader families are null-pubnum or their discovery legs OPS-errored (4 `COVERAGE_LIMITED` cells); (3) commercial editing estates that are US applications with **NULL publication numbers** (Tune/Chroma/nChroma/some Sangamo) are un-ingestable and remain under-counted; (4) **TET is genuinely thin** (1 biblio-only family). This corrects round-1 the right way — whitespace is declined on incompletely-sampled modalities rather than asserted then retracted. Un-filed art is invisible by construction.

- **Not FTO, not calibrated.** OPEN = hypothesis prompt, not clearance. Genus recall & KC are lower bounds. Scripts: `scripts/{bin_scope_grid,whitespace_analysis,dynamic_signals,build_kg}.py`.