Two largely independent axes of drug transcriptional response in Tahoe-100M

Huazhang Shen

Independent Researcher · hzshen@gmail.com · ORCID 0009-0006-8208-3111

Preprint, not peer reviewed. Version date: 2026-08-22.

Abstract

Some drugs produce nearly the same transcriptional fingerprint in every cell line; others act only where biological context permits, so their response shifts from one background to the next. Large-scale drug-perturbation atlases are used mainly to train and benchmark predictive models, not to audit response structure itself; we take the descriptive stance instead.

Using Tahoe-100M — the largest single-cell drug-perturbation atlas, spanning 50 cancer cell lines, 379 small molecules and 3 concentrations — we show a drug's pseudobulk response resolves onto two nearly independent axes (Spearman = -0.03, n = 268). **Context-robustness** is the fraction of a drug's response that is a background-independent conserved "core" rather than a cell-line-specific "periphery"; its ranking matches pharmacological expectation and is well-calibrated under a full-library permutation check (91% of drugs with zero empirical false-core genes; = 0.99 against the reported ranking). **Dose-emergence** orders each drug's genes by the concentration at which they turn on, recovers known mechanism at the gene level, and resolves "stress" into two programs with opposite dose behaviour. **Target transcriptional depth** — how close a drug's target sits to the transcription machinery — has no significant direct effect on context-robustness (an initial reading via core pathway coherence did not survive correction), but a real, moderate effect on dose-emergence: proximal-target drugs show fewer low-dose-emergent genes than distal-target drugs ($p = 0.0013$, $\Delta R^2 = 6.3\%$, $n = 157$). A Type III variance partition shows the stress response is dominated by drug identity and cell-line background over nominal dose (partial $^2 \approx 0.53$–0.66 vs 0.02), with a substantial drug×dose interaction. Response strength also tracks a known clinical resistance mechanism: KRAS-mutant lines respond more weakly to the EGFR inhibitor Simotinib (FDR = 0.001). The contribution is an interpretable coordinate system for response structure, with code and derived tables openly available (CC0 source data).

---

1. Introduction

The public release of Tahoe-100M — a single-cell drug-perturbation atlas of ~100 million cells spanning 50 cancer cell lines, 379 small molecules and three concentrations [1] — has been followed by a rapid wave of computational work. That work is overwhelmingly *predictive*: foundation models, diffusion and flow-matching architectures, "virtual-cell" simulators, and the benchmarks built to score them. The organizing question of the field is, given a drug and a cellular context, to predict what the transcriptional response will look like.

A more basic, descriptive question has received far less attention. For any given drug, how much of its measured transcriptional response is **conserved across cellular backgrounds** — a background-independent mechanistic fingerprint — and how much is **context-specific**, varying from one cell line to the next? And what governs that ratio? This is a conservation–divergence audit rather than a prediction task: the object of study is the response that has already been measured, not a model of it. A phrase-restricted survey of the indexed literature (arXiv and PubMed, July 2026) is consistent with the predictive/benchmarking work being crowded and this descriptive angle under-explored. An arXiv search for "virtual cell" + "perturbation" returns 36 papers, overwhelmingly proposing or benchmarking predictive models (diffusion, flow-matching, world-model and foundation-model architectures); a direct "Tahoe-100M" search returns 6; searches combining "conserved" or "context-dependent" with "drug perturbation" or "single-cell" return 0–2 hits, none of which decompose measured response into conserved and context-specific structure. We are careful throughout not to overstate this into a claim that conservation–divergence decomposition has never been attempted

— the absence of matching search results is not proof of absence, and data-provider technical notes have gestured at related intuitions (non-peer-reviewed) — but rather to position the specific contribution as making context-robustness a *quantifiable drug property across the full library, with a test of its mechanistic cause.*

We report two empirical response axes and test one candidate mechanistic correlate — target transcriptional depth — against each:

- **Context-robustness** (empirical axis 1). For each drug we separate its cross-cell-line conserved core from its context-specific periphery using a consistency statistic tailored to within-drug, across-cell-line agreement, and rank drugs by the fraction of their response that is core. This ranking is a new, interpretable drug property.
- **Dose-emergence** (empirical axis 2). Along the three concentrations we order each drug's genes by the dose at which they turn on, separating candidate on-mechanism (early) from candidate generic-stress (late) genes, and characterise what tunes the shared stress response. We show directly that these two axes are near-independent, so they carry complementary information rather than restating one another.
- **Target transcriptional depth** (candidate correlate, tested against both axes). We ask whether a drug's position on each empirical axis tracks how deep its target sits in the signalling-to-transcription hierarchy. Once tested with a directly controlled regression and honest multiple-testing correction, depth has no significant direct effect on context-robustness, but it does have a real, moderate direct effect on dose-emergence (proximal-target drugs have a lower fraction of low-dose-emergent genes, indicating broader or later recruitment across the tested dose range). Depth is therefore an axis-specific correlate, not a shared mechanistic origin for both axes.

The result is an interpretable, two-axis coordinate system for drug-response structure on public-domain data, together with a scoped account of what depth does and does not explain about it — a resource for, and a proposed ground truth against, the predictive-modelling community.

---

2. Methods

2.1 Data and preprocessing

We analysed the Tahoe-100M atlas — Hugging Face dataset `tahoebio/Tahoe-100M`, configuration `pseudobulk_differential_expression`, comprising 1,026 shards (88.9 GB). Each shard provides, for one combination of cell line, plate and concentration, the DESeq2 [2] log2 fold-change of drug-treated versus vehicle control per gene. The full release covers **50 cancer cell lines** (13 organs) × **379 drugs** × **3 concentrations** (0.05, 0.5, 5.0 µM), each (drug, concentration) supported by multiple plate replicates. We retained the **12,310 genes** with mean baseMean > 10 as expressed. Unless noted, the primary concentration is 5.0 µM (fullest response, broadest positive-control coverage), with 0.5 and 0.05 µM used for concentration-robustness cross-checks.

Analysis scope (declared). All three analyses were computed on the **full 50-cell-line release**. Context-robustness (§2.2–2.4) and target depth (§2.7) were computed on the 50-line release throughout. Dose-emergence (§2.5) and the stress-axis decomposition (§2.6) were first developed on a 22-cell-line pilot subset and then **re-run on the full 50 lines**; unless stated otherwise, every dose-emergence and stress number in this paper is the 50-line value. The two scales agree on the classifier (per-drug low-dose-emergent fraction, 50-vs-22 Spearman = 0.64, n = 379; median 0.43 (50-line) vs 0.44 (22-line pilot)) but differ on the stress-axis variance partition, where the larger cell-line panel changes the ranking of factors (§3.4); we report the 50-line partition as primary and note the pilot value where it differs. The 50-line dose-emergence and stress-axis tables were produced by re-aggregating the per-(drug, gene, cell-line, concentration) means underlying the context-robustness analysis, so all three analyses rest on the same 50-line pseudobulk collapse.

2.2 A cross-cell-line consistency statistic

The conservation–divergence question here is not "how do classes differ" (the setting of permutation-based class-divergence tests) but "for a single drug, how consistent is the response across cell lines versus how

dispersed." This is closer in spirit to a variance-component / intraclass-correlation decomposition, so we defined a dedicated statistic (validated to perfectly separate core / periphery / none on synthetic data). For each drug and gene, given per-cell-line mean log2FC μ_c (after collapsing plate replicates) and within-replicate variance:

- **effect** (response magnitude) = mean_c $|\mu\_c|$ — using per-cell-line absolute values so that opposite-direction responses are not miscounted as "no response";
- **direction-consistency** = |mean_c μ_c| / effect (all same-direction ≈ 1, cancelling ≈ 0);
- **magnitude-consistency** = effect^2 / (effect^2 + excess cross-line variance), where excess variance = cross-line variance - plate-replicate noise floor;
- **consistency** = the product of the two;
- **z** from a permutation null that shuffles cell-line assignment (preserving each line's marginal, breaking gene identity).

A drug × gene is **core** if effect $\geq$ threshold and z $\geq$ threshold; **periphery** if responsive but not core; **none** if unresponsive.

2.3 Multiple stringency levels

Because core calls depend on threshold, we report three levels in parallel: loose (effect ≥ 0.5, z ≥ 2), **medium (effect $\geq$ 1.0, z $\geq$ 3; primary)**, and strict (effect ≥ 1.5, z ≥ 3). All main-text figures use the medium level at 5.0 µM.

2.4 Universal axis and controlled context-robustness ranking

We define the **ubiquitous conserved set** as the genes that are core in $\geq 50\%$ of drugs — small nuclear RNAs, a mitochondrial pseudogene, a pseudoautosomal lncRNA and unannotated loci with no established growth-arrest identity, so we do not describe this set as a functional "growth-arrest axis"; at the medium level (5.0 µM) this is a compact set of **6 genes** (7SK, ASMTL-AS1, MTND2P28, RNU4-2, and two unannotated loci ENSG00000278996, ENSG00000289901). A drug's **drug-specific core** is its core minus this ubiquitous conserved set. The **context-robustness score** is the drug-specific core fraction, residualised by ordinary least squares against log(response magnitude) and cell-line count n; a floor of ≥ 100 responsive genes was applied (277/379 drugs pass). Two nearly identical versions of this residual are used, differing only in whether the OLS is fit before or after the floor is applied (Spearman = 0.99 between them): the ranking display (Fig 2) uses the on-floored-set fit, and the cross-axis analyses (Figs 5–6) use the full-set fit; the choice does not affect any conclusion.

> **Reconciliation of the two "universal axis" definitions.** This 6-gene universal axis (used here for the drug-specific-core subtraction) is not in conflict with the 103-gene "stress axis" used in §2.6. They are two nested thresholds of the same phenomenon: the 6 genes are the $\geq 50\%$-of-drugs core at medium stringency, and are a strict subset of a broader 103-gene conservation set (adding ribosomal, mitochondrial and integrated-stress genes such as *PPP1R15A*/GADD34 and *HSPA8*) obtained at a looser conservation threshold. The 6-gene set serves as a conservative subtraction so that "drug-specific core" is not dominated by a handful of pan-cytotoxic loci; the 103-gene set serves as a dose-independent probe of stress-response *magnitude*. We use each for its stated role and do not mix them.

2.5 Dose-emergence classifier

For each drug × responsive gene we computed a continuous **emergence score** = the dose at half-maximal |log2FC|, interpolated on log-dose across the three concentrations (lower = activates at lower dose = more potent), with within-drug normalisation, a monotonicity flag, and a baseMean detectability control. Discrete low/mid/high bins are a secondary view. Absolute potency differs across drugs, so emergence is normalised within each drug; drugs that saturate at the lowest dose (no dynamic range) self-flag and are excluded rather than forced.

Pathway enrichment along the dose axis used continuous directional GSEA: responsive genes ranked by emergence, tested against 1,516 size-filtered Reactome pathways (per-drug Mann–Whitney, BH-FDR [3] within drug) against the Tahoe-tested-gene background (7,134 genes), not the genome default.

2.6 Stress-axis magnitude decomposition

Using the dose-independent 103-gene stress axis (inherited from the conservation analysis, which is why the dose-link test below is non-circular), we decomposed the variance of per-sample stress-axis magnitude across drugs by a drug × cell-line × dose ANOVA-style sum-of-squares partition, and ranked mechanism-of-action classes by their stress-axis baseline.

2.7 Functional enrichment and core coherence (target depth)

For each drug we took its drug-specific core (§2.4) and tested it for pathway enrichment. Mixed identifiers in the core tables were unified by mapping Ensembl IDs to HGNC symbols with mygene.info (133 of 511 Ensembl IDs resolved; the remainder are non-coding/pseudogene loci absent from the pathway libraries). Enrichment used the hypergeometric test against a **Tahoe-expressed-gene background** (N = 11,151 symbols — the subset of the 12,310 baseMean-filtered expressed genes (§2.1) that appears in the core/periphery label table across the drug panel; the two counts are different, both-correct denominators for different steps of the pipeline, not an inconsistency), with GO Biological Process (2023) [4,5], Reactome (2022) [6] and KEGG (2021) [7] libraries, size-filtered to 5–500 genes (6,026 sets); enrichment was run with GSEApy [8]. Per drug, p-values were BH-corrected; a pathway was called enriched at FDR < 0.1 with an observed overlap ≥ 2 genes. Enrichment ran for all 374 drugs with a core-specific set; 348 (93%) had a core large enough (≥ 5 mapped genes) to be enrichable.

Per-drug pathway p-values are BH-corrected across the full size-filtered 6,026-set library, not merely the subset of pathways that happen to pass an overlap ≥ 2 filter, so that the multiple-testing universe matches the tested library; a pathway is called enriched at FDR < 0.1. Under this procedure, 132 of 374 drugs (35%) have at least one BH-significant pathway, and 77 of these retain a depth label and a large enough enrichable set to support the coherence–depth test of §3.6.

We initially planned a target-recall statistic (does a drug's core recover its own annotated target pathway) but abandoned it: drugs with distal targets have cores too small for any pathway to survive multiple-testing correction, so recall degenerates into a restatement of core size (and hence of context-robustness). We instead defined a size-robust, orthogonal readout — **core coherence** — as the fraction of a drug's FDR-significant pathways supported by more than three core genes: whether a core parses into connected pathway programs (high) or dissolves into fragile two-to-three-gene hits (low), independent of whether those pathways match the annotated target.

Each drug was assigned a **target transcriptional depth** from its fine-grained mechanism-of-action annotation (moa-fine; 26 classes, 377/379 drugs) by a fixed mapping, prospectively specified before the library-scale analysis. *Proximal* targets act at or near the transcription/chromatin/translation machinery or on immediate growth-signalling hubs (HDAC, DNMT, MEK, RAF, RAS, other MAPK, PI3K/AKT, mTOR, GSK3, JAK/STAT, CDK, protein-synthesis and proteasome inhibitors). *Distal* targets act far from transcription (DNA synthesis/repair, EGFR/ERBB and other tyrosine kinases, microtubules, Hedgehog/SMO, adrenoceptors, cyclooxygenase, transporters). Ligand-activated nuclear receptors (glucocorticoid, androgen, retinoic-acid receptors) are mechanistically proximal transcription factors but their effect is gated by receptor expression, so they form a separate *nuclear-receptor* tier. Drugs with unclear MOA (198) were resolved where possible by their annotated target gene; 191 drugs received a depth label (72 proximal, 101 distal, 18 nuclear-receptor).

2.8 Confound control and bias testing

Because coherence rises with core size, all depth comparisons control for it: we fit coherence ~ log10(core size) by OLS and compared residuals across depth tiers (Mann–Whitney), and separately fit coherence ~ log10(core size) + depth for the partial effect of depth. Because 183 drugs lack a depth label, we tested for

selection bias by comparing labelled vs unlabelled drugs on core size, coherence and context-robustness, raw and size-corrected.

Direct regressions bypassing the coherence intermediate. Because coherence is itself a size-dominated, enrichment-dependent statistic (§2.7), we additionally tested target depth directly against each empirical axis without routing through coherence: `core_frac_resid` (context-robustness) and `frac_low` (dose-emergence) were each regressed on depth (proximal indicator) + log(core size) + n_cell_lines by OLS, with the proximal-tier coefficient tested by both the regression coefficient p-value and a Mann–Whitney test on the covariate-adjusted residuals (n = 157 depth-labelled drugs with both axis values available).

2.9 External molecular validation (DepMap)

To test whether the response-strength readout encodes known molecular determinants, we mapped the 50 Tahoe cell lines to DepMap 24Q4 [9] model identifiers (49/50 mapped; hTERT-HPNE is a normal control absent from DepMap). We screened nine targeted drugs (Simotinib/EGFR, Sonidegib/SMO, olaparib/PARP1-2, Pimitespib/HSP90, Volasertib/PLK1-3, Silodosin/ADRA1, Hydroxyfasudil/ROCK, plus Panobinostat and Trametinib as broad-mechanism controls for contrast); of these, three (Simotinib, Trametinib, Volasertib) had a DepMap hotspot-mutation gene with $\geq$3 mutant lines available for testing, yielding five drug-mutation pairs. We associated per-cell-line response strength with DepMap hotspot-mutation status (542-gene matrix) for these five pairs by Mann–Whitney with BH-FDR, checking the significant hit for organ confounding by re-examining direction within organ.

---

3. Results

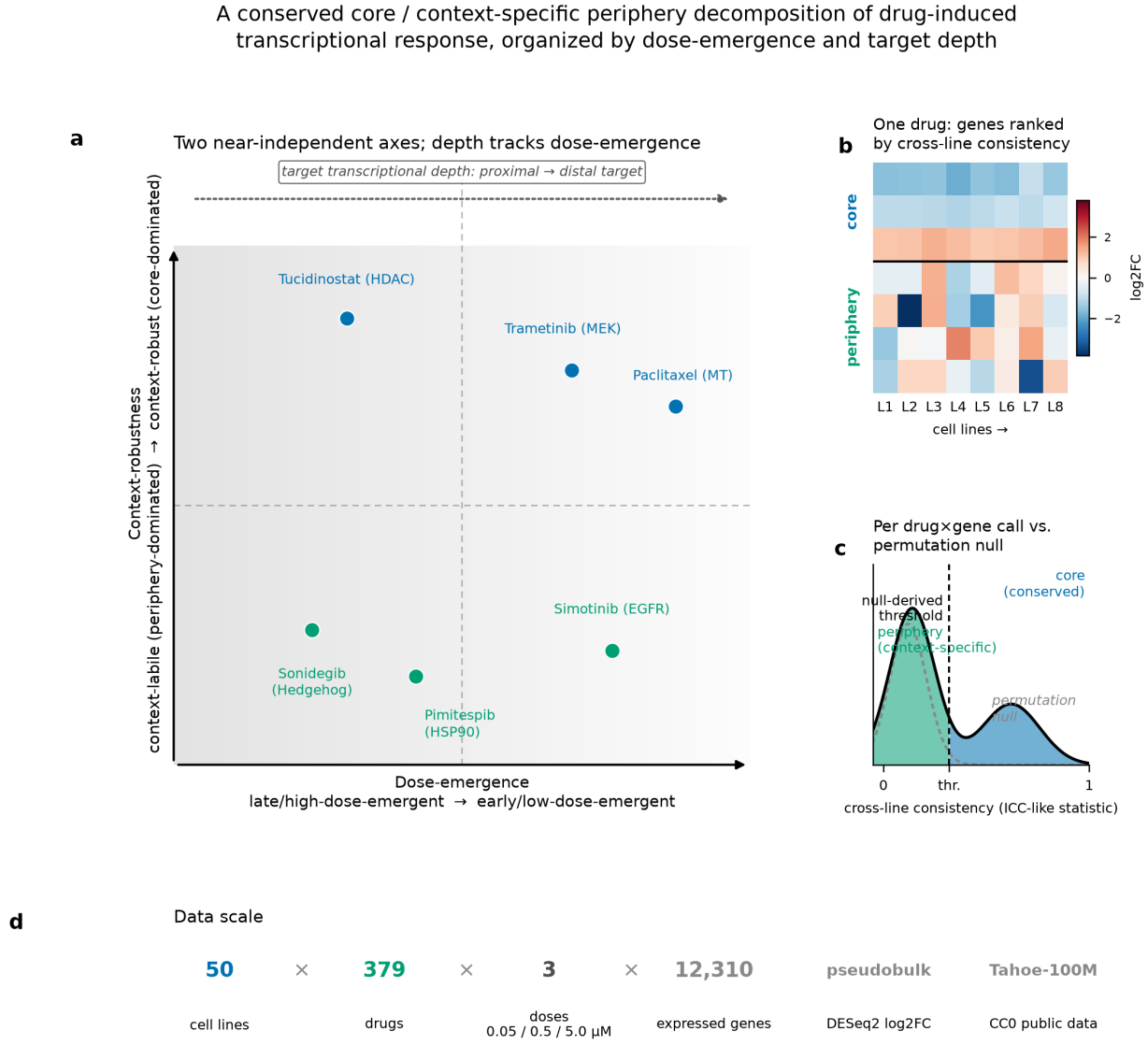


Figure 1. A conserved-core / context-specific-periphery decomposition of drug-induced transcriptional response, organized by dose-emergence and target depth. (a) The two empirical response axes — context-robustness (vertical) and dose-emergence (horizontal) — are near-independent (Spearman = -0.03; Fig 5), so illustrative drugs are spread across the dose-emergence range within each robustness band rather than along a diagonal. Target transcriptional depth (horizontal bracket) tracks the dose-emergence axis, not context-robustness — see §3.6 for the corrected analysis and the retraction of an earlier coherence-mediated context-robustness link. (b) Schematic per-drug log2FC across cell lines: a conserved core block versus a context-specific periphery block. (c) The per-drug×gene core/periphery call from a cross-cell-line consistency statistic against a permutation null. (d) Data scale: 50 cell lines × 379 drugs × 3 doses × 12,310 expressed genes, pseudobulk DESeq2 log2FC, CC0 Tahoe-100M.

3.1 The conserved core is overwhelmingly drug-specific

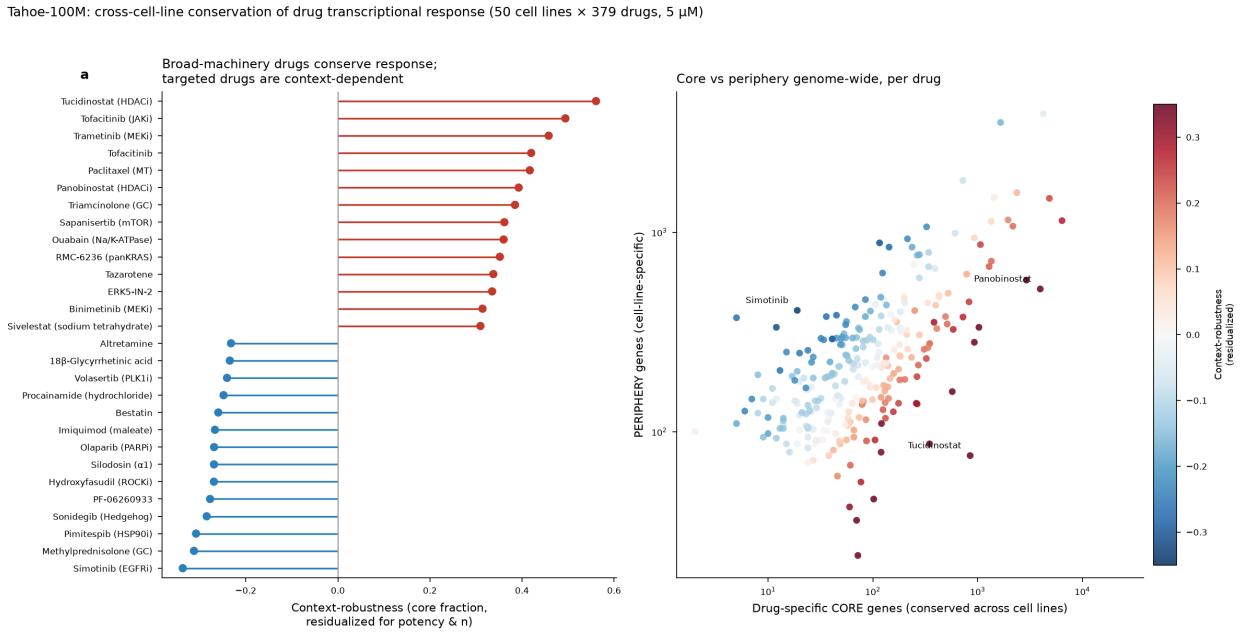


Figure 2. Context-robustness ranking and the core/periphery decomposition across 50 cell lines. Left: the most context-robust (core-dominated, top) and most context-labile (periphery-dominated, bottom) drugs by controlled residual; only the extremes are shown for readability. Right: the genome-wide core/periphery landscape.

At the medium level (5.0 µM) the atlas contains **72,529** drug-specific core, **1,644** universal core, and **89,649** periphery drug × gene calls. The ubiquitous conserved set — genes that are core in $\geq 50\%$ of drugs — comprises only **6 genes** (a mitochondrial pseudogene, two small nuclear RNAs, a pseudoautosomal lncRNA, and two unannotated loci), stable across concentration (all six appear in the 0.05 µM set). None of the six has an established growth-arrest identity, so we do not describe this set as a functional "growth-arrest axis." The decisive observation is that the overwhelming majority of core calls are *drug-specific*: the drug-specific-core-fraction distribution shows a cliff at 0.5, so the conserved core of a drug's response carries drug-specific mechanistic information rather than a uniform, functionally interpretable signature. A conservation–divergence decomposition is therefore meaningful and information-rich on drug-perturbation data.

3.2 Context-robustness is a continuous, interpretable drug property

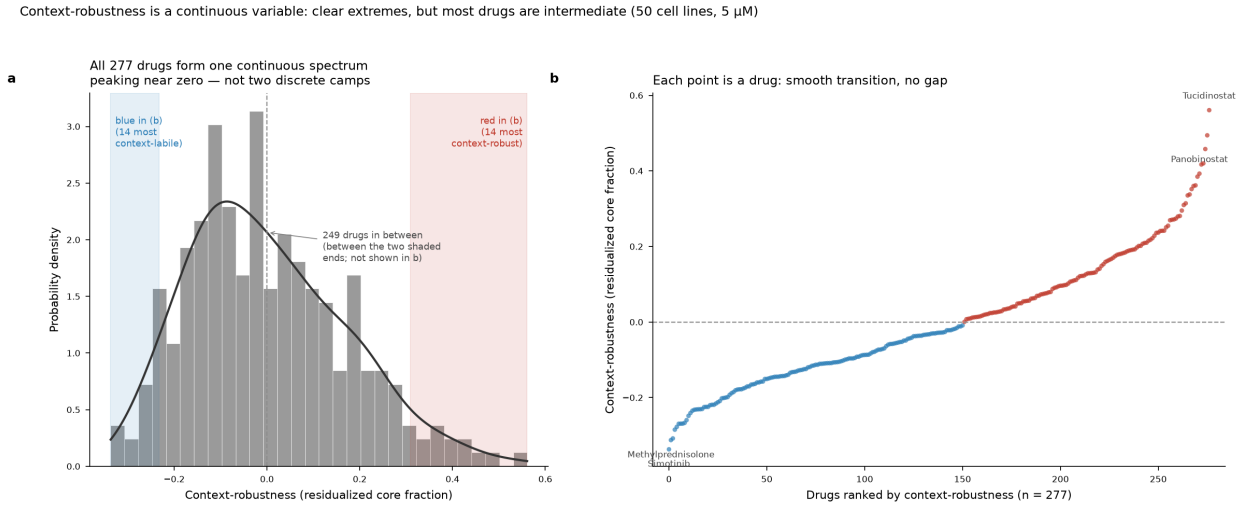


Figure 3. Context-robustness is a continuous unimodal spectrum, not two discrete classes. The full distribution across 277 drugs peaks near zero; the extreme-only ranking in Fig 2 should not be read as a bimodal split.

The most context-robust drugs (highest controlled residual) are inhibitors of broad signalling hubs and cell machinery — HDAC (Tucidinostat, residual +0.561, core 849; Panobinostat +0.393, core 3,963), MEK (Trametinib +0.458, core 347), mTOR/PI3K (Sapanisertib +0.362; RMC-6236 +0.352), JAK (Tofacitinib +0.420), Na/K-ATPase (Ouabain +0.360) and microtubules (Paclitaxel +0.417). The most context-labile are agents whose effect depends on a specific target or lineage — EGFR (Simotinib, residual -0.337, core 19), HSP90 (Pimitespib -0.308), PARP (olaparib -0.269), Hedgehog (Sonidegib -0.285), ROCK (Hydroxyfasudil -0.269) and glucocorticoids (Methylprednisolone succinate -0.313, polarised by receptor expression). This ordering matches pharmacological priors and is strong evidence of method validity.

Two honest boundaries accompany this result. First, context-robustness is a **continuous, unimodal spectrum peaking near zero** (Fig 3), not two discrete robust/labile camps; the extreme-only ranking of Fig 2 omits 249 intermediate drugs and must not be read as bimodal. Context-robustness should be used as a continuous variable. Second, the ranking carries a real concentration dependence (on the 277-drug floored set, 5 vs 0.5 µM Spearman $= 0.52$; 5 vs 0.05 µM $= 0.35$), because a single nominal concentration sits at different points on each drug's dose–response curve; the 5 µM result is primary and this dependence is a stated caveat, not a masked flaw.

3.3 Dose-emergence recovers mechanism, and "stress" is two opposite-dose programs

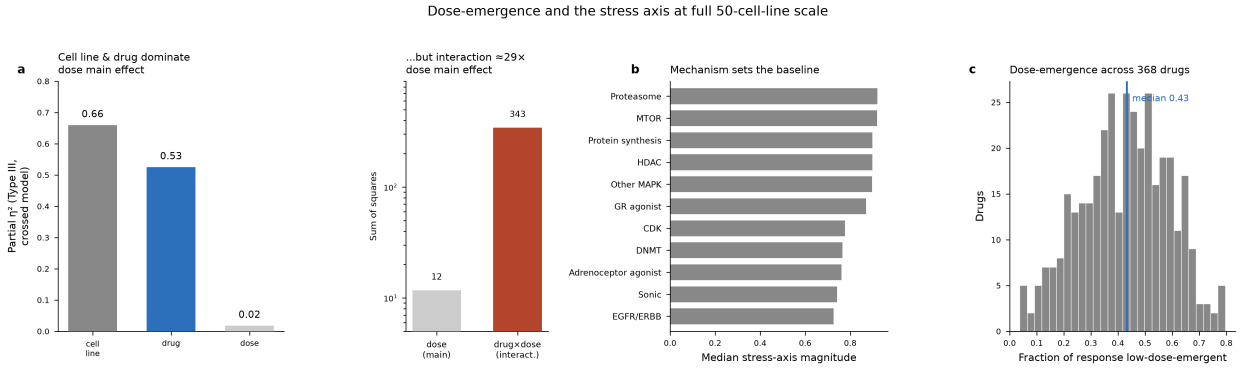


Figure 4. Dose-emergence and the stress axis at full 50-cell-line scale. (a) Under a crossed (Type III) variance partition, stress-axis magnitude varies far more across cell lines and drugs than across nominal dose (partial $^2 \approx$ 0.53–0.66 vs 0.02), but the drug×dose interaction term is itself ~29-fold the dose main effect — dose's effect is highly drug-dependent, not a fixed background rate. (b) Median stress-axis magnitude by mechanism class: highest in proteasome, mTOR, protein-synthesis and HDAC inhibitors, lowest in EGFR/ERBB, Hedgehog and DNMT agents. (c) Distribution of the per-drug low-dose-emergent fraction across 368 classifiable drugs, median 0.43.

At full 50-line scale, **368 of 379 drugs are classifiable** by dose-emergence (11 saturate at the lowest dose and self-flag); the median fraction of a drug's responsive genes that are low-dose-emergent is 0.43 (Fig 4c). The classifier is stable between scales — the per-drug low-dose fraction agrees between the 22-line pilot and the full 50 lines (Spearman = 0.64, n = 379). Positive controls pass at the gene level: Trametinib's MAPK-output genes (*FOS, EGR1, DUSP4/6, SPRY2/4*) and Bortezomib's proteasome subunits (*PSMB1/2/5/6, PSMD1*) are all correctly low-dose-emergent, and the emergence score's partial correlation with baseline expression is small and sign-inconsistent ($|\text{partial r}| \leq 0.15$), so early emergence is not a detectability artifact.

At the pathway level, **18 of 258 classifiable drugs (7%)** recover their own annotated target pathway on the early/low-dose side at FDR < 0.1. This is a narrow-versus-broad-target split, not a failure: single-gene/small-family targets recover cleanly (Auranofin → TXNRD1/2, FDR 1×10^{-10}; Idarubicin → TOP2A, 4×10^{-7}; Bortezomib → proteasome, 9×10^{-5}; Tucidinostat → HDACs, 6×10^{-4}), whereas family/complex targets (Panobinostat's seven HDACs, Trametinib's MAP2K1/2, EGFR and rapalogs) are diluted across many pathways even though their marker genes are correctly early. The gene-level emergence score is the robust primary output; pathway-FDR recovery is a conservative secondary layer that by construction rewards narrow targets.

We read "stress" itself as **two programs with opposite dose behaviour** — a reading that stitches two analyses together rather than one single readout, which we state plainly. An early, on-mechanism **translation-attenuation/growth-arrest** program (the conserved-axis genes: ribosomal, mitochondrial, *PPP1R15A*/GADD34, *HSPA8*) is enriched *low-dose* across classifiable drugs, while a late, high-dose **cytotoxic-death cascade** dominates the stress *pathways* (DNA-damage 85% late, p53 88% late, apoptosis 76% late, heat-shock 62% late; pathway-direction analysis). Combining the two, the pattern is sequential recruitment, not a switch: the growth-arrest program is already on at low dose (an early consequence of potently hitting the target), and raising the dose progressively adds the death cascade on top. The intuition "high dose recruits generic stress" is therefore right for the death cascade and wrong for the growth-arrest program. Throughout, "late-emerging" denotes a within-drug relative-potency ranking at a single steady-state timepoint and is consistent with generic high-dose stress *or* slow on-target downstream kinetics; late ≠ proven-nonspecific.

3.4 Stress-axis magnitude is set by drug identity and cell-line background, but dose's effect is itself drug-dependent

We decomposed the variance of per-sample stress-axis magnitude across the full 50 lines with a fully crossed (Type III) drug × cell-line × dose model, including drug×dose and cell-line×dose interaction terms (sum-to-zero contrasts, 1,283 parameters); a scalar "drug/dose" ratio requires this crossed specification, since an order-dependent sequential partition is not a valid basis for one (see correction-notes.md for the superseded sequential-partition figures). Cell line and drug are the dominant factors (partial $^2 = 0.660$ and 0.525 respectively, versus 0.018 for dose) — cell-line background and drug identity dominate over a fixed dose effect. Note the model tests individual *drug identity* (379 levels), not annotated mechanism-of-action class; Fig 4b's mechanism-class medians are a separate, descriptive summary and should not be read as apportioning this drug-identity variance component to "mechanism" as a formal factor. The qualitative reading is in line with the atlas paper's own characterisation that cell-line identity and cell-cycle phase are the major factors while drug treatment and dose have a more modest effect [1]. The drug×dose interaction sum of squares (≈343) is, however, ~**29-fold the dose main-effect sum of squares (≈12)**, and the cell-line×dose interaction (≈19) is also non-negligible relative to the dose main effect. Dose's effect on the stress axis is therefore highly drug-specific — some drugs show strong dose-dependence, others almost none — so no single fixed "drug/dose" ratio is statistically defensible once this interaction is accounted for; we report the partial-^{2} comparison above in its place. Mechanism still sets the per-drug baseline: stress-axis magnitude is highest in proteasome, mTOR, protein-synthesis and HDAC inhibitors, lowest in EGFR/ERBB and Hedgehog inhibitors and DNMT and adrenoceptor agents (Fig 4b); the per-gene dose-emergence axis remains usable because it is a within-drug, within-gene contrast rather than a global embedding factor.

3.5 The two axes are near-independent

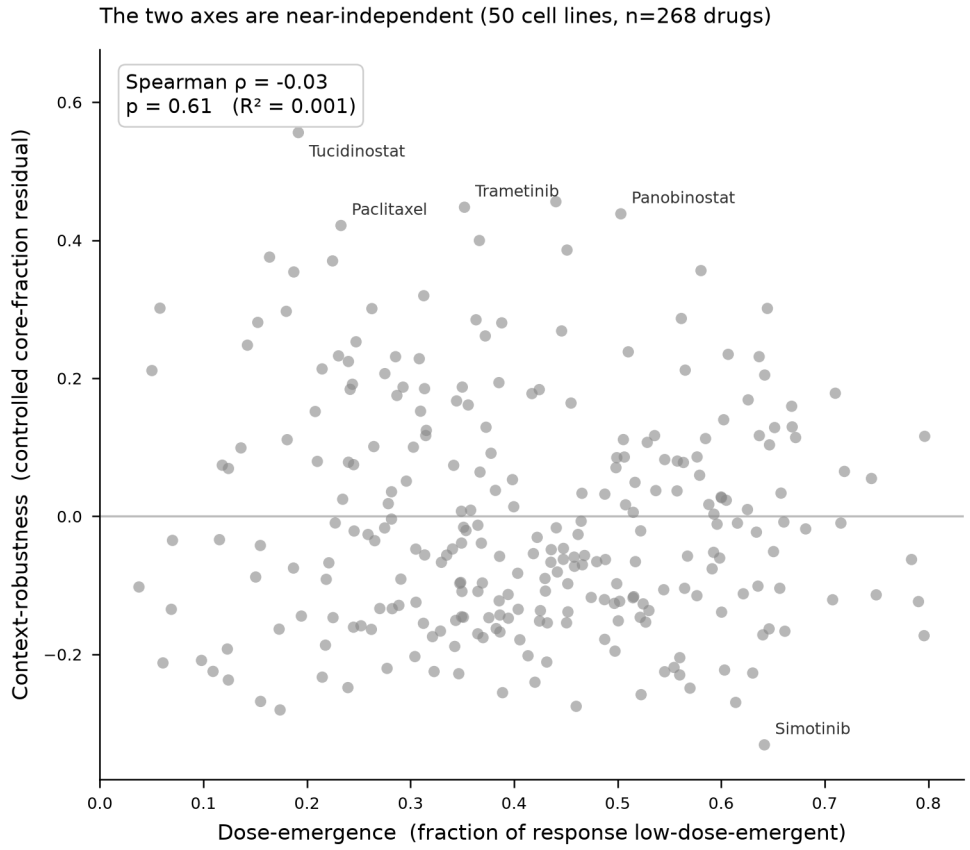


Figure 5. The two response axes are near-independent. On the 268 drugs classifiable on both axes and drawn from the 50-cell-line release (per-drug coverage 17–50 lines, median 49; 114/268 drugs have full 50-line coverage), context-robustness (controlled core-fraction residual) shows essentially no monotonic relationship with dose-emergence (fraction of response that is low-dose-emergent): Spearman = -0.03 (p = 0.61, n = 268); restricted to the 114 full-coverage drugs, = -0.15 (p = 0.10). Illustrative drugs are labelled.

A two-axis framing is only useful if the axes are actually distinct. We tested this directly on the 268 drugs classifiable on both axes and drawn from the 50-cell-line release; per-drug cell-line coverage in this set ranges 17–50 (median 49), and 114/268 drugs have full 50-line coverage. Context-robustness and dose-emergence are near-independent: the low-dose-emergent fraction has essentially no monotonic association with the context-robustness residual across the full set (Spearman = -0.03, p = 0.61, R^2 = 0.001, n = 268; Fig 5), and the continuous median-emergence-dose carries only a weak one (= 0.22, p = 3×10^{-4}, R^2 = 0.05, n = 268). Coverage-matched subsets give the same qualitative reading: restricted to the 197 drugs with $\geq$ 48 cell lines, = -0.03 (p = 0.68) and = 0.23 (p = 0.001) respectively; restricted to the 114 drugs with full 50-line coverage, = -0.15 (p = 0.10) and = 0.27 (p = 0.004). The frac_low-vs-robustness association stays near zero and non-significant at every coverage threshold; the median-emergence-dose association is weak but strengthens slightly, not away, as coverage tightens. Sparse-coverage drugs are not manufacturing the apparent independence. A drug can be robust and early-emergent (Tucidinostat), robust and later (Trametinib), labile and early, or labile and late — all four quadrants are populated. Whether a drug's response is background-independent, and the dose at which that response turns on, are two separable properties. This is what licenses treating them as two axes rather than one.

3.6 Target transcriptional depth and dose-emergence

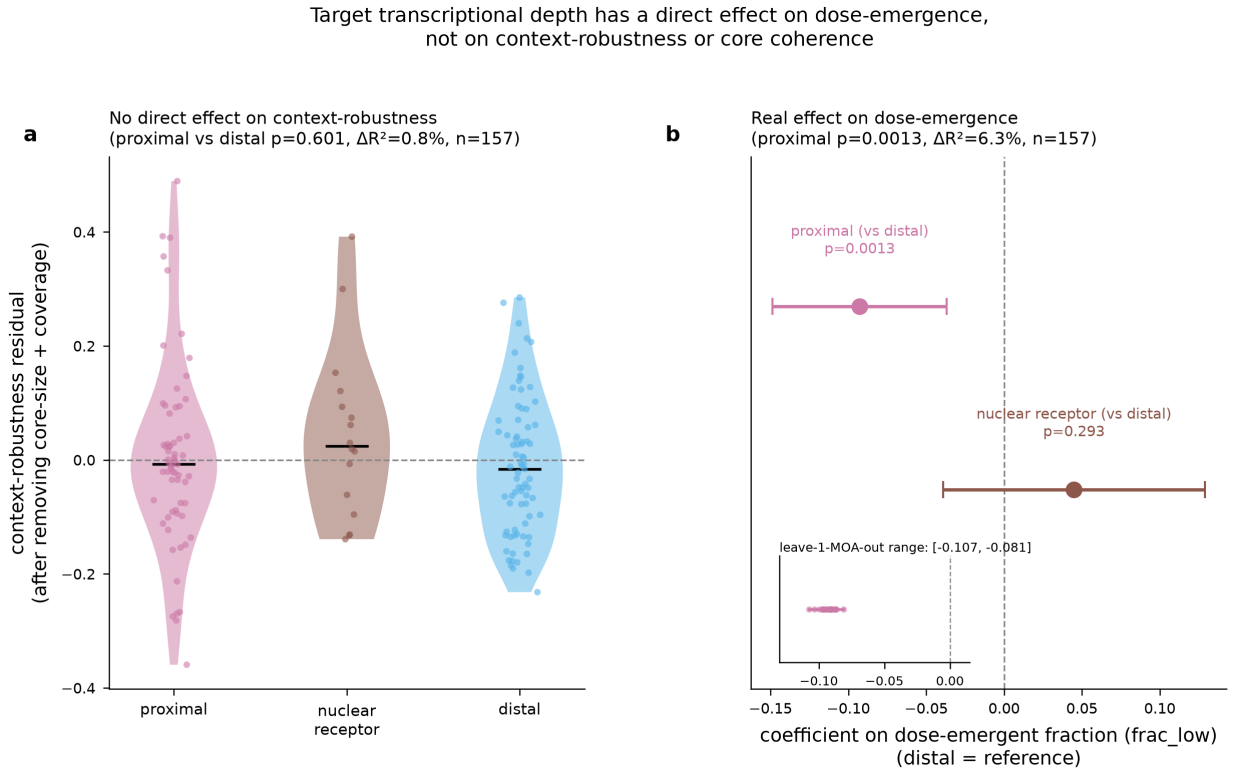


Figure 6. Target transcriptional depth has a direct effect on dose-emergence, not on context-robustness or core coherence. (a) Target-depth tier has no significant direct effect on context-robustness (residual core_frac_resid after removing core-size and cell-line-coverage effects; proximal vs distal p = 0.601, ΔR^2 = 0.8%, n = 157) — depth does not organize context-robustness, either through coherence or directly. (b) Target-depth tier has a real, moderate direct effect on dose-emergence (frac_low ~ depth + log(core size) + n_cell_lines; proximal coefficient = -0.093 vs. distal, p = 0.0013, ΔR^2 = 6.3%, n = 157), robust to cluster-robust standard errors, a magnitude-controlled alternative specification, and leave-one-MOA-class-out refitting (proximal coefficient range [-0.107, -0.081] across 26 refits, always p < 0.01; inset). The originally reported coherence–depth relationship (proximal > distal core coherence at matched size) did not** survive an honest Benjamini–Hochberg correction across the full pathway library and is retracted (see text below and Supplementary Figure S2 for the original-vs-corrected comparison).**

Across the corrected analysis, core coherence remains strongly determined by core size (Spearman = 0.70, p = 2×10^{-20}, n = 132 drugs with at least one BH-significant pathway; $R^2 \approx 0.30$–0.47 depending on subset), confirming that any statistic built on absolute pathway recovery largely re-encodes core size and hence context-robustness; a naïve target-recall analysis (does a drug's core recover its own annotated target pathway) adds nothing beyond the context-robustness axis for the same reason.

An initial reading of depth via core pathway coherence did not survive correction. At equal core size, proximal-target drugs showed apparently more coherent conserved cores than distal-target drugs under an initial enrichment procedure; re-deriving the pathway enrichment with honest Benjamini–Hochberg correction across the full size-filtered 6,026-pathway library (§2.7) leaves this effect **not significant** (R^2 rises from 0.457, size only, to 0.468 with depth added; ΔR^2 = 1.1%, coefficient-test p = 0.234, Mann–Whitney-on-residuals p = 0.165, n = 77). We therefore do not report "target transcriptional depth organizes core coherence" as a finding; the full original-versus-corrected comparison is archived in correction-notes.md and Supplementary Figure S2.

Target depth has a direct effect on dose-emergence, not on context-robustness. Bypassing the coherence intermediate entirely, we tested target depth directly against each empirical axis by OLS, controlling for log(core size) and cell-line coverage (§2.8; n = 157 depth-labelled drugs with both axis values). Depth has **no significant direct effect on context-robustness** (`core_frac_resid` ~ depth + log(core size) + n_cell_lines: proximal coefficient p = 0.601, Mann–Whitney-on-residuals p = 0.692, $\Delta R^2 = 0.8\%$) — target depth does not organize context-robustness, either through coherence or directly. Depth **does** have a real, moderate direct effect on **dose-emergence**: `frac_low` (fraction of responsive genes that are low-dose-emergent) ~ depth + log(core size) + n_cell_lines gives proximal coefficient = -0.093, p = 0.0013, Mann–Whitney-on-residuals p = 0.00065, $\Delta R^2 = 6.3\%$ — proximal-target drugs have a lower low-dose-emergent fraction (fewer genes respond at low dose; equivalently, they show a broader dose range of activation) than distal-target drugs at matched core size and coverage. This is the paper's mechanistic-depth result: **target transcriptional depth is linked to dose-emergence, not to context-robustness or to core coherence.**

Core coherence and context-robustness remain positively associated (Spearman $= 0.32$, $p = 6\times10^{-4}$, n = 112), consistent with both being driven substantially by core size; we make no depth-mediated mechanistic claim about that association, and read context-robustness and core coherence as reflecting shared size-dependence rather than a common causal driver.

3.7 What did not replicate

The distal-cores-are-death-driven pattern from an early pilot failed at library scale and is reported as a negative: across all enriched drugs, distal cores were no more death-enriched than proximal cores (Mann–Whitney p = 0.72), and the pilot signal traces to small-sample variance in two drugs. We therefore removed the "distal cores are death-driven" claim entirely.

We also found no selection bias from unlabelled drugs. The 183 depth-unlabelled drugs had marginally smaller cores and lower raw coherence (p = 0.004, 0.008 under the original uncorrected enrichment), but this gap disappeared after size correction (residual coherence p = 0.44), and the groups did not differ in context-robustness (p = 0.53) or enrichable rate. Excluding unlabelled drugs does not bias the depth-axis relationships reported above in either direction.

3.8 Methodological caveats: the periphery layer, and permutation-based core-gene calibration

The periphery layer (89,649 calls) is not, on inspection, a set of cross-drug "context-dependent response hubs" — its skeleton is a **cell-line-identity/secretome layer**, and we surface this as a caveat for anyone reusing these tables rather than as a biological result. Three baseMean-free diagnostics establish it (Supplementary Fig S1): the top-100 periphery genes are almost never core in any background (median core-fraction 0.042); periphery is ~2.2× more concentrated than core; and the hubs are overwhelmingly lineage/secretome markers (*ALB/AFP/AHSG*, *SOX10/DCT*, *S100A9*, mucins, *TFF3/PSG9*). Highly expressed lineage/secretory genes carry systematic technical fluctuation in pseudobulk DESeq2 log2FC that the consistency statistic misreads as "inconsistent response." The practical lesson: pseudobulk log2FC cross-cell-line consistency analyses must baseMean-decontaminate the low-consistency layer before treating it as biology, and GO-term filtering is not enough (it misses classic loci like *ALB/AFP*).

Permutation-null calibration of core-gene calls. The $z \geq 2$ (loose) and $z \geq 3$ (medium/strict, §2.3) thresholds used to call a drug×gene "core" were checked against a null-shuffle calibration: an initial scoped check on 6 drugs, which silently used the function's loose default (effect ≥ 0.5, $z \geq 2.0$), and a full-library check across all 277 context-robustness drugs (5 independent null replicates per drug) at the manuscript's actual reporting stringency (medium: effect ≥ 1.0, $z \geq 3.0$, used for `core_frac_medium`/`core_frac_resid` and Figs 2, 3, 5, 6).

At the loose threshold, the null is measurably non-Gaussian and right-skewed (skew ≈ 0.9–1.5; Kolmogorov–Smirnov/Shapiro–Wilk reject normality at $p < 0.001$), with $z \geq 2$ type-I error running ~1.9× nominal and implied maximum FDR reaching ~14–45% for robust drugs and >100% of the reported core-gene count for labile drugs (Simotinib, Hydroxyfasudil). This is a real property of the loose threshold and is retained here as a caveat on any analysis using it.

At the manuscript's actual medium stringency, the full 277-drug recalibration is materially better-calibrated: **253 of 277 drugs (91%) have zero empirical false-core genes** across 5 null replicates, the worst case (Belzutifan) has an implied maximum FDR of **2.5%**, and the null-adjusted, residualized ranking correlates with the reported `core_frac_resid` ranking at **Spearman = 0.99** (n = 277). Bootstrap resampling (200 cell-line resamples) confirms the top- and bottom-ranked illustrative drugs used in Figs 1–3 (Tucidinostat, Trametinib, Paclitaxel, Panobinostat vs. Simotinib, Pimitespib, Sonidegib) keep their extreme positions and non-overlapping confidence intervals under this correction. **We therefore treat the medium-stringency context-robustness ranking (Figs 2, 3, 5, 6) as well-calibrated** — the loose-stringency calibration concern applies specifically to analyses run at that looser, non-default setting, not to the manuscript's headline reporting stringency. Individual core-gene *lists* at the loose threshold should still be read as descriptive rather than FDR-controlled.

3.9 Response strength encodes a known molecular determinant: KRAS and anti-EGFR resistance (external validation)

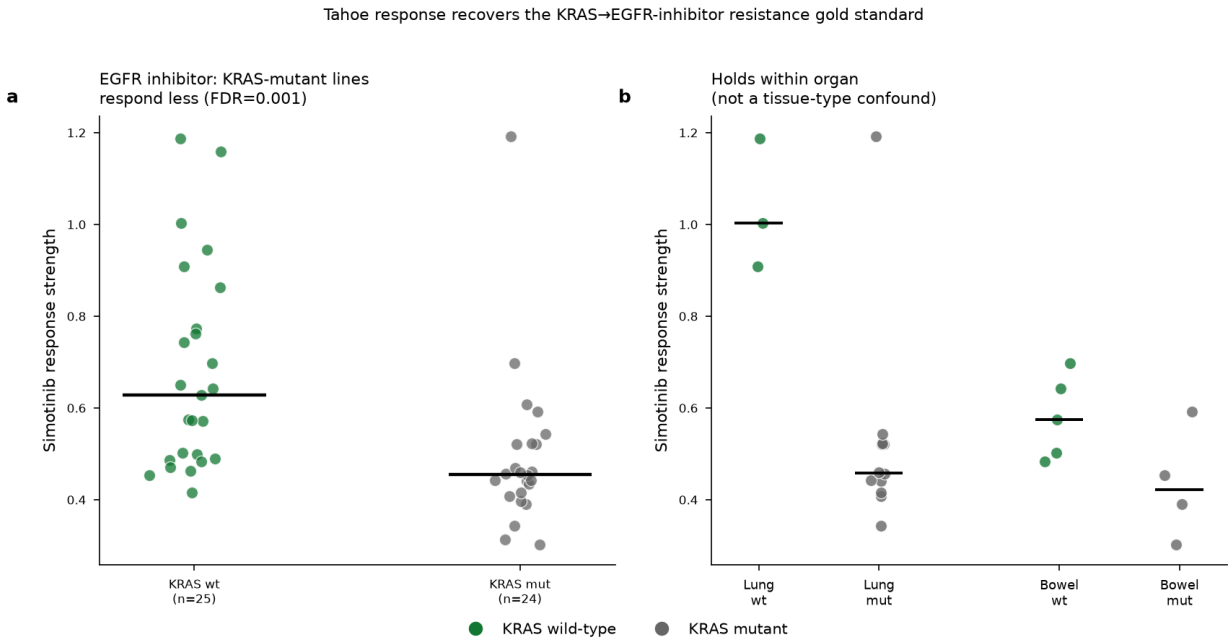


Figure 7. External molecular validation. KRAS-mutant cell lines respond more weakly to the EGFR inhibitor Simotinib (median 0.456 vs 0.628 wild-type; p = 0.0002, FDR = 0.001), consistent with the known KRAS-associated resistance pattern to anti-EGFR therapy (established for cetuximab/panitumumab in colorectal cancer; here observed with a small-molecule EGFR inhibitor across multiple organs); the same direction is present within both lung and colorectal subgroups (informal check, not a joint model).

An earlier attempt to find molecular determinants of targeted-drug response from Tahoe's own coarse features (organ + single driver mutation) was a nominal-only negative. Replacing those with DepMap 24Q4 molecular annotations turns it positive: **KRAS-mutant lines respond significantly more weakly to the EGFR inhibitor Simotinib** (median response 0.456 vs 0.628 in wild-type; p = 0.0002, **FDR = 0.001**), consistent with the known KRAS-associated resistance pattern to anti-EGFR therapy — established for the monoclonal antibodies cetuximab and panitumumab in colorectal cancer, where KRAS mutation is a clinical exclusion criterion [10,11] — here observed with a small-molecule EGFR inhibitor and across multiple organs, not colorectal cancer specifically. The same direction is present within both the lung subgroup (mutant 0.459 vs wild-type 1.003) and the colorectal subgroup (0.422 vs 0.574) considered separately; this is an informal within-subgroup check, not a joint statistical model with organ as a covariate, and the per-organ sample sizes are too small to support a formal interaction test. We state the boundary honestly: of five mutation tests only

this one is significant (the MEK inhibitor Trametinib's BRAF/KRAS association is not, plausibly because transcriptional response magnitude ≠ proliferation inhibition, or because n = 7 BRAF lines is too small). Screening covered nine targeted drugs in total (seven target-specific plus two broad-mechanism controls for contrast); of these, three (Simotinib, Trametinib, Volasertib) had a DepMap hotspot-mutation gene with ≥ 3 mutant lines available, yielding the five drug-mutation pairs tested. This upgrades the biomarker analysis from an honest negative to an external demonstration, anchored to a known clinical association, that Tahoe response strength is molecularly interpretable, and provides evidence for using response strength as a predictive-model ground truth — without claiming to have reproduced the clinical rule itself or to have formally excluded an organ effect.

As a further descriptive check, context-robustness is unrelated to clinical approval status (approved vs experimental, p = 0.83, FDR = 0.87) — a drug's conservation–divergence mode does not determine whether it reaches market — while experimental drugs show marginally stronger transcriptional responses than approved drugs (median |log2FC| 0.45 vs 0.42, p = 0.007, FDR = 0.033), consistent with approved libraries containing many mild, target-specific legacy agents.

---

4. Discussion

We have shown that drug transcriptional responses in the largest available perturbation atlas admit an interpretable structural decomposition along two empirical axes that are near-independent (Spearman = -0.03 for the frac_low operationalization of dose-emergence; a weak but real = 0.22–0.27 for the continuous median-emergence-dose operationalization, so "near-independent" is best read as applying most cleanly to the frac_low reading) — context-robustness and dose-emergence. The core is overwhelmingly drug-specific; context-robustness is a continuous, pharmacologically sensible drug property; and the shared stress response reads as two opposite-dose programs whose magnitude is set by mechanism and cell-line background rather than by nominal dose, though dose's effect is itself substantially drug-dependent (§3.4). We tested whether a single mechanistic correlate — target transcriptional depth — organizes a drug's position on *both* axes, and found that it does not: depth has no significant direct effect on context-robustness (either through the coherence intermediate, which does not survive honest multiple-testing correction, or directly), but it does have a real, moderate direct effect on dose-emergence (proximal-target drugs have a lower fraction of low-dose-emergent genes than distal-target drugs; $\Delta R^2 = 6.3\%$, p = 0.0013, n = 157). Depth is therefore an axis-specific correlate of dose-emergence, not a shared mechanistic origin for both empirical axes.

Honest positioning. The contribution is deliberately not a predictor. In a field whose published Tahoe-100M work is dominated by predictive foundation models, "virtual-cell" simulators and their benchmarks, the systematic quantification of conservation–divergence structure as a drug property is an under-developed angle rather than virgin territory — data-provider notes have gestured at related decompositions, and we do not claim priority over the general idea. What we add is a two-axis coordinate system with analysis code and derived tables openly available, and a scoped, honestly-corrected account of what target depth does and does not explain about it. The window for descriptive/mechanistic analyses of this specific atlas is real but time-limited: the data provider, with the Arc Institute and the Biohub, has announced (January 2026 press release) a forthcoming dataset the provider describes as substantially more perturbation-rich than Tahoe-100M — a company projection of a not-yet-released resource rather than an audited comparison, but a clear indication that the descriptive ground will keep shifting.

A proposed ground truth for predictive models. The core/periphery split offers a directly useful evaluation dimension: when a model predicts a drug's response, is it getting the robust, conserved core right, or the noisy, context-specific periphery? Scoring predictors separately on core versus periphery — and, given §3.9, on whether they reproduce response-strength associations with known molecular determinants — would sharpen a benchmark landscape that currently mixes the two. Any such use of the core/periphery gene lists should carry the calibration caveat of §3.8: the aggregate context-robustness *ranking* is comparatively robust, but individual core-gene lists have not been corrected for multiple testing across the full gene panel and should be treated as descriptive rather than FDR-controlled.

Limitations. (i) The depth mapping is hand-built. Target transcriptional depth is our own fixed

proximal/distal/nuclear-receptor assignment of the atlas's mechanism-of-action labels, prospectively specified before the library-scale analysis but not a measured quantity; we use causally neutral language ("is associated with") and note that a different binning of borderline classes could shift the effect, which — after correction — is null for context-robustness and modest for dose-emergence ($\Delta R^2 = 6.3\%$). (ii) Concentration: three nominal doses give response *shape*, not dose–response curves; context-robustness carries a real concentration dependence ($\approx$ 0.4–0.5 across doses) because a fixed concentration sits at different points on each drug's curve — so context-robustness measured at 5 µM is a property of the response *at that concentration*, not a concentration-free constant. (iii) Steady-state snapshot: "late-emerging" cannot be separated from slow on-target kinetics at a single timepoint. (iv) Periphery contamination: the periphery skeleton is dominated by lineage/secretome identity and requires baseMean decontamination before biomarker use. (v) The stress-axis "two programs" reading combines a gene-level low-dose enrichment with a pathway-level high-dose enrichment; it is a synthesis of two analyses, not a single measured switch, and the mechanism/dose variance partition itself required a crossed model to expose the drug×dose interaction (§3.4). (vi) The DepMap validation is a single significant mutation association among five tests screened from nine targeted drugs (§2.9, §3.9); it is consistent with, but does not itself reproduce, the clinical KRAS/anti-EGFR resistance rule established for monoclonal antibodies in colorectal cancer, and the organ check is informal rather than a joint statistical model. (vii) Permutation-based core-gene calling: the $z \geq 2/z \geq 3$ thresholds are not well-calibrated against a Gaussian null (§3.8); individual core-gene lists carry a material, quantified false-positive risk across the whole context-robustness spectrum, and should be used as descriptive rankings, not FDR-controlled discovery sets, until a formal multiple-testing correction across the full ~12,310-gene × 379-drug panel is performed.

Future work. Completing the DepMap expression-feature validation across the nine targeted drugs; a sensitivity analysis of the depth mapping under alternative class bins; a formal FDR-controlled recalibration of the core/periphery permutation statistic at full library scale; and building the explicit core/periphery-stratified benchmark interface for the predictive-modelling community.

Conclusion. On public-domain data with strict confound discipline — including negative and retracted results reported in full — a drug's transcriptional response resolves into two largely independent, empirically grounded axes. Target transcriptional depth is not the shared mechanistic origin of both axes we originally proposed; it is a real but axis-specific correlate of dose-emergence alone. We offer the two-axis coordinate system, together with this corrected and more limited account of what target depth explains, as a complement to the predictive models that currently dominate the use of drug-perturbation atlases.

Data and code availability

Source data. Tahoe-100M is available at the Hugging Face Hub, `tahoebio/Tahoe-100M`, configuration `pseudobulk_differential_expression` (CC0 1.0 public-domain dedication). DepMap 24Q4 Public is available from the Broad Institute Cancer Dependency Map portal (`depmap.org`) and Figshare+ (doi:10.25452/figshare.plus.27993248).

Derived tables (this study). The per-drug results tables are released as supplementary files. Because these tables are the reusable product of the paper, we document their columns:

- `context_robustness.csv` — one row per drug. `core_frac_medium`: fraction of the drug's responsive gene calls that are conserved-core at medium stringency (5 µM). `core_frac_resid`: context-robustness score = `core_frac_medium` residualised (OLS) against log response-magnitude and cell-line count; this is the ranking variable. Higher = more context-robust.
- `full50_dose_emergence_summary.csv` (50-line) — one row per drug. `frac_low`: fraction of responsive genes that are low-dose-emergent; `median_emergence_dose` (µM): median interpolated dose at half-maximal |log2FC|; `saturation_verdict`: CLASSIFIABLE / UNCLASSIFIABLE_SATURATED_LOW / LOW_SIGNAL.
- `full50_stress_axis_drivers.csv` (50-line) — one row per drug. `stress_baseline`: mean stress-axis magnitude (mean |log2FC| over the 103-gene conservation set) across samples; `dose_slope`: within-drug slope of stress magnitude on log-dose.
- `full50_axis_projection.csv` (50-line) — one row per (drug, cell line, concentration): `axis_mean_abs`, `axis_mean_signed`, `nonaxis_mean_abs` — the per-sample stress-axis projections underlying the vari-

ance decomposition.

- `axis_orthogonality.csv` — the merged per-drug table (context-robustness residual × dose-emergence) plotted in Figure 5.
- `orthogonality_coverage_sensitivity.csv` — the axis-independence correlations (frac_low and median_emergence_dose versus context-robustness residual) recomputed on the all-drugs, ≥48-line, and full-50-line coverage subsets underlying §3.5 and Figure 5's coverage caveat.
- `enrichment_metrics_corrected.csv` — one row per drug: core size, BH-corrected core coherence (`coherence_corrected`, re-derived across the full 6,026-set pathway library per drug, superseding the originally released `enrichment_metrics.csv`), target-depth tier, and the covariates (`core_frac_resid`, `n_cell_lines`, `frac_low`) used in the direct depth-vs-axis regressions of §3.6.
- `depmap_mutation_response_assoc.csv` — one row per (drug, gene) mutation test: mutant/wild-type median response, Mann–Whitney p, BH-FDR.
- `tahoe_stress_axis_variance_partition_corrected.csv` — the Type I/II/III stress-axis variance-partition table (§3.4).
- `calibration_type1_error_check.csv`, `calibration_qq_zscore_normality_check.csv`, `calibration_bootstrap_core_stability.csv`, and `calibration_check_summary.csv` (a per-drug merge of the preceding three) — the four permutation-calibration diagnostic tables (type-I error, QQ/normality, bootstrap stability; §3.8).

Code. The analysis scripts (streaming pseudobulk aggregation, cross-cell-line consistency statistic, dose-emergence classifier, stress-axis decomposition, enrichment/coherence, DepMap association, and figure generation), together with the post-review remediation scripts that produced the corrected tables above, are archived on Zenodo: https://doi.org/10.5281/zenodo.21449974 (concept DOI, always resolves to the latest version; the current version is v2, DOI https://doi.org/10.5281/zenodo.21531168). Tahoe pseudobulk aggregation is a streaming pass over the 1,026 shards; the 50-line dose-emergence and stress-axis tables were produced by re-aggregating the per-(drug, gene, cell-line, concentration) means built for the context-robustness analysis. Derived outputs and analysis code are available as described above; one gap is documented rather than hidden: the full-scale 50-line re-aggregation script was run on a remote host whose working directory was later cleared and could not be recovered verbatim, and some figure-generation scripts required import fixes to run standalone outside their original environment — the known aggregation parameters are recorded in the archive's README in its place, and this is noted here rather than described as fully reproducible end-to-end.

Competing interests

The author declares no competing interests.

Funding

This work received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions

H.S. conceived the study, performed all analyses, and wrote the manuscript.

Computational tools and AI-assisted analysis

Data assembly, statistical analysis, and figure generation were implemented in Python with substantial code-development and computational assistance from Claude Science (Anthropic), an AI-based scientific computing agent, used interactively throughout data assembly, analysis, and figure preparation. All analysis code, computed values, and statistical results were reviewed, checked for internal consistency, and independently spot-verified by the author against source data; the author takes full responsibility for the accuracy of all reported results and conclusions. Consistent with ICMJE and COPE guidance, this AI tool is not credited as an author, as it cannot take accountability for the work.

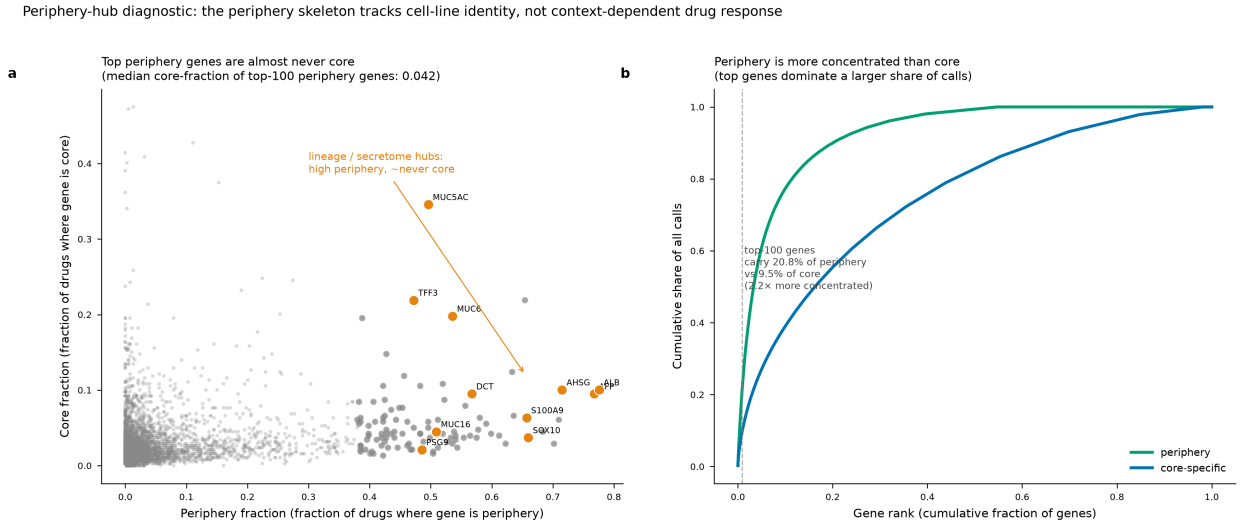
Figures

- **Figure 1** — Conceptual overview: the two empirical axes, data scale, and method.
- **Figure 2** — Context-robustness ranking and the core/periphery landscape (50 cell lines).
- **Figure 3** — Context-robustness is a continuous, unimodal spectrum.
- **Figure 4** — Dose-emergence and the stress axis at full 50-cell-line scale: crossed variance partition (drug identity and cell-line background dominate over nominal dose, but dose's effect is itself drug-dependent), mechanism-class stress ranking (descriptive), and the dose-emergence distribution.
- **Figure 5** — The two response axes are near-independent (= -0.03, n = 268).
- **Figure 6** — Target transcriptional depth has a direct effect on dose-emergence (proximal coefficient = -0.093, p = 0.0013, n = 157) and no direct effect on context-robustness (p = 0.601); the originally reported coherence–depth link does not survive honest multiple-testing correction and is retracted (Supplementary Figure S2).
- **Figure 7** — DepMap external validation: KRAS mutation and weak Simotinib response.
- **Supplementary Figure S1** — Periphery-hub diagnostic (methodological caveat): the periphery layer tracks cell-line identity.
- **Supplementary Figure S2** — Coherence–depth relationship, original (uncorrected) versus BH-corrected across the full pathway library; shows the basis for the §3.6 retraction of the coherence–depth claim.

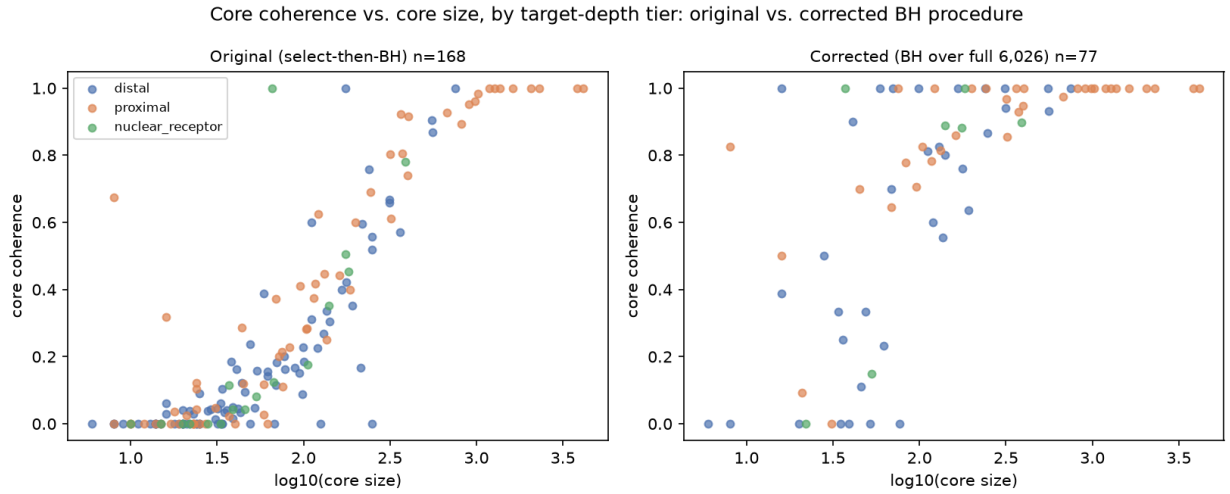
References

1. Zhang J, Ubas AA, de Borja R, Svensson V, Thomas N, Thakar N, et al. Tahoe-100M: A Giga-Scale Single-Cell Perturbation Atlas for Context-Dependent Gene Function and Cellular Modeling. *bioRxiv* 2025. doi:10.1101/2025.02.20.639398
2. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8. PMID 25516281
3. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x
4. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nature Genetics*. 2000;25(1):25-29. doi:10.1038/75556. PMID 10802651
5. The Gene Ontology Consortium. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224(1):iyad031. doi:10.1093/genetics/iyad031. PMID 36866529
6. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Research*. 2024;52(D1):D672-D678. doi:10.1093/nar/gkad1025. PMID 37941124
7. Kanehisa M, Furumichi M, Sato Y, Matsuura Y, Ishiguro-Watanabe M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Research*. 2025;53(D1):D672-D677. doi:10.1093/nar/gkae909. PMID 39417505
8. Fang Z, Liu X, Peltz G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics*. 2023;39(1):btac757. doi:10.1093/bioinformatics/btac757. PMID 36426870
9. Broad DepMap. DepMap 24Q4 Public [data set]. *Figshare+*. 2024. doi:10.25452/figshare.plus.27993248
10. Karapetis CS, Khambata-Ford S, Jonker DJ, O'Callaghan CJ, Tu D, Tebbutt NC, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. *New England Journal of Medicine*. 2008;359(17):1757-1765. doi:10.1056/NEJMoa0804385. PMID 18946061
11. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *Journal of Clinical Oncology*. 2008;26(10):1626-1634. doi:10.1200/JCO.2007.14.7116. PMID 18316791

Supplementary



Supplementary Figure S1. Periphery-hub diagnostic. Top periphery genes are almost never core in any background (median core-fraction 0.042) and are ~2.2× more concentrated than core, and are dominated by lineage/secretome markers — the periphery skeleton tracks cell-line identity, not context-dependent drug response. This is a methodological caveat for reuse of the periphery tables, not an independent biological result.



Supplementary Figure S2. Coherence–depth relationship, original versus BH-corrected. Coherence versus core size by target-depth tier under the originally reported (uncorrected, overlap-then-BH) enrichment procedure (left) and under the corrected procedure that BH-corrects across the full 6,026-pathway library per drug (right). The proximal/distal separation visible on the left collapses once the correction is applied honestly (right; Mann–Whitney on size-adjusted residuals $p = 0.165$, $n = 77$) — this is the basis for the §3.6 retraction.