Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A method-independent recoverability ceiling for transcript-to-cell assignment in imaging-based spatial transcriptomics

Rui Yamasaki

rui.yamasaki0724@gmail.com

Independent Researcher https://orcid.org/0009-0000-1277-6197

Research Article

Keywords: spatial transcriptomics, imaging-based spatial transcriptomics, transcript-to-cell assignment, cell segmentation, Xenium, MERFISH, Bayes-optimal assignment, benchmarking, recoverability ceiling, spurious co-expression

Posted Date: June 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10093252/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

In imaging-based spatial transcriptomics, assigning transcripts to their cell of origin underlies most downstream analysis, yet there is no method-independent measure of how accurately it can be recovered. Using synthetic tissue with known ground truth, we compute the best assignment accuracy achievable against the generating process, a ceiling no method can exceed. Calibrated to real data, with displacement pinned by a non-circular validity test, the ceiling falls below 0.9 in dense tissue, to between 0.66 and 0.81, replicated on Xenium breast and lung. The best fairly configured published methods leave a further 0.066 to 0.110 unrecovered beneath it. We package the ceiling as an open diagnostic that flags spatial conclusions the data cannot support, and apply it to a published tumor co-expression claim: it identifies the claim as unrecoverable, independent data contradict it, and a denser corroborated claim shows the diagnostic does not condemn findings indiscriminately.

Introduction

Imaging-based spatial transcriptomics measures the location of individual RNA transcripts within intact tissue, then assigns each transcript to the cell that produced it [1, 2]. This assignment step underlies nearly every downstream analysis, from cell-type calling to inferences about which cells interact [3, 4], yet it is also the least certain. Transcripts are detected as points, cell boundaries are rarely imaged directly, and in dense tissue the molecules of neighboring cells lie within micrometres of one another, so a transcript's measured position need not indicate its cell of origin.

The field has responded with a steady stream of segmentation and reassignment methods [3, 5, 6, 7, 8], each aiming to recover the true assignment more faithfully than the last, and with quality metrics that compare methods to one another. What has been missing is a reference point that does not depend on any method: a statement of how accurately assignment can be recovered at all, given the physical conditions of the measurement. Without such a reference it is impossible to know whether a method is near the limit of what the data allow or whether substantial accuracy remains to be recovered, and impossible to know which biological conclusions rest on assignments the data cannot actually support.

We provide that reference point. By generating synthetic tissue in which the true cell of origin of every transcript is known, then computing the best assignment any method could achieve against the generating process, we obtain a method-independent ceiling on assignment accuracy as a function of transcript density, cell packing, and displacement. Calibrated to real data, the ceiling shows where assignment is recoverable and where it is not, places existing methods relative to that limit, and yields a diagnostic that flags which spatial conclusions in dense tissue require independent confirmation.

Results

A by-construction ceiling on transcript-to-cell assignment

The accuracy of transcript-to-cell assignment is limited not only by the segmentation method applied but by the information present in the measurement itself. That limit can be computed exactly when the true cell of origin of every transcript is known. To obtain such ground truth, we built a generative model of an imaging-based

spatial transcriptomics field in which cells are placed at a controllable packing density, assigned cell types, and emit transcripts that are displaced from their point of origin by a controllable amount before detection. Because the field is generated rather than measured, the cell that emitted each transcript is recorded exactly, which no real experiment can provide (Fig. 1a).

Given this known generative process, the assignment that maximizes expected accuracy assigns each detected transcript to the cell with the highest posterior probability of having emitted it. This Bayes-optimal assignment is the best achievable against the model, so its accuracy is a ceiling: no segmentation or reassignment procedure, however sophisticated, can exceed it on data drawn from the same process. We call this quantity the oracle accuracy (Fig. 1b).

Sweeping the generative parameters over a grid of transcript density, cell packing, and displacement produced a smooth, monotone surface. Oracle accuracy fell as cells were packed more tightly and as displacement increased, and rose as transcripts became more abundant (Fig. 1c). Across the full grid it ranged from 0.234 to 0.957, so the identical task is almost perfectly recoverable in the easiest regimes and largely irrecoverable in the hardest. We refer to this surface as the answerability frontier, since it states, for each physical regime, the best assignment accuracy the data can support (Extended Data Fig. 1).

To confirm that the oracle is a genuine ceiling and not an artifact of its own construction, we compared it at every grid point to a naive baseline that assigns each transcript to the nearest cell center. The oracle equaled or exceeded this baseline at all 125 grid points, so the simplest reasonable method never beats the optimum (Fig. 1d).

For the frontier to bear on real measurements, the generator must reproduce real tissue. Calibrated to public Xenium and MERFISH fields, it matched their median transcripts per cell to within 1.52% and reproduced their cell packing exactly (Extended Data Fig. 2). The regime that real tissue occupies falls in the informative interior of the frontier, well below complete recovery, which is where a fundamental limit on assignment would constrain practice rather than being irrelevant. Locating that limit precisely on real data requires fixing the one generative parameter that cannot be read directly from a cell-by-gene matrix, the transcript displacement, which we determine next.

A data-pinned displacement places the bound on real tissue

The frontier locates a regime once three quantities are known. Transcript density and cell packing can be read directly from a processed dataset, but transcript displacement, the distance a transcript travels from its cell of origin before detection, cannot: a cell-by-gene matrix records where transcripts were assigned, not where they began. Estimating displacement from apparent co-expression of cell-type markers is circular, because genuine biological co-expression and displacement-driven contamination produce the same signal.

To break the circularity we measured displacement only from markers that demonstrably carry a displacement signal, selected without reference to the displacement they imply. For each candidate marker we computed the excess of its expression in cells immediately adjacent to a marker-positive cell over its expression in distant cells, and admitted the marker only if this adjacent-minus-distant signal exceeded a position-permutation null at $P < 0.05$ over 1,000 permutations. Admissibility therefore depends on whether a marker responds to displacement at all, never on the value it yields, so the estimate is non-circular by construction. We then pinned

displacement to the aggregate leakage of the admissible markers, with a confidence interval from a bootstrap of 2,000 resamples over markers and cells.

On a public Xenium breast cancer dataset (167,780 cells) [2, 9], 28 of 38 candidate markers were admissible and the pinned displacement was 1.99 micrometres (95% CI 1.43 to 2.63). At this displacement, in dense tissue at about 13,600 cells/mm2, the oracle assignment accuracy was 0.663 to 0.805 across the displacement CI, entirely below 0.9 (Fig. 2a-c). The same pipeline applied to an independent Xenium lung cancer dataset (150,365 cells) [10] admitted 31 of 32 markers, pinned displacement at 1.83 micrometres (CI 1.62 to 2.08), and gave a dense oracle accuracy of 0.746 to 0.796. The best-possible accuracy of transcript-to-cell assignment in dense tissue is therefore bounded below 0.9 on two independent tissues.

The contrast between the two datasets is itself evidence that the estimate is not an artifact of marker selection. In breast, 10 of 38 markers were rejected, so selection did substantial work; in lung, only 1 of 32 was, so selection barely operated (Extended Data Fig. 5). If the displacement in breast were inflated by retaining only strong-signal markers, lung should land elsewhere. Instead the two agree, at 1.99 and 1.83 micrometres.

The bound survived the assumptions a skeptical reader challenges first (Fig. 2d). Replacing the idealized marker model with realistic overlapping expression left the assignment frontier essentially unchanged, with a mean absolute difference of 0.006 and the oracle-minus-naive gap preserved at 0.071 against 0.066 (Extended Data Fig. 3). Adding realistic within-type expression variation, modeled as negative-binomial with dispersion near 4.19, shifted the dense-tissue accuracy by only − 0.016. The bound also held under non-Voronoi cell geometry and non-Gaussian displacement (Extended Data Fig. 4), and was stable across cell-typing choices, with the pinned displacement ranging over 1.99 to 2.65 micrometres and dense accuracy never exceeding 0.803 (Extended Data Fig. 6).

Current methods leave a measurable fraction of the ceiling unrecovered

A ceiling is informative only if methods in use fall short of it. We therefore ran three widely used segmentation and reassignment methods, Baysor, Proseg and ComSeg [3, 5, 6], on synthetic dense-tissue fields where the true cell of origin was known, each tuned to its best documented configuration. To compare them against the oracle on equal terms, we supplied every method with the true nuclei as a prior, so that each solved the same assignment-given-locations problem the oracle bounds and was not charged for segmentation errors the oracle never incurs.

Across all three over-segmentation-robust metrics, the best fairly configured method remained 0.066 to 0.110 below the oracle in dense tissue (Fig. 3a,b; full metric tables in Extended Data Fig. 8). A measurable fraction of the achievable accuracy is thus left unrecovered by current practice, and the oracle marks the target that fraction is measured against.

Whether these methods improve on the simplest possible baseline, nearest-nucleus assignment, depended on the metric and stayed within the run-to-run noise of the methods themselves, about 0.012. Under a one-to-one matched metric the best method sat 0.004 to 0.009 below nearest-nucleus; under a many-to-one homogeneity metric it sat up to 0.009 above; under a co-assignment metric, up to 0.019 above (Fig. 3c). Reassignment given true nuclei is therefore close to a wash against nearest-nucleus, and the meaningful comparison is not to that baseline but to the ceiling above all of them. When the nuclei prior was withdrawn, only one of the three

methods (Baysor) ran de novo, and its accuracy fell below even nearest-nucleus, quantifying the additional cost of inferring cell boundaries rather than assignment alone.

A diagnostic applied to a published claim

The bound implies a practical test: a spatial conclusion that depends on transcript assignment, drawn from a regime above the ceiling, cannot be trusted from the spatial data alone and requires orthogonal confirmation. We applied this test to a published claim from a widely cited Xenium breast cancer study [2], which reported a small ductal carcinoma in situ region, roughly 0.1 mm2, in which cells co-expressed the receptors ERBB2, ESR1 and PGR (HER2, ER and PR), even though the tissue block was clinically annotated HER2-positive, ER-positive and PR-negative.

First, is the claim in an unrecoverable regime? The co-expressing cells sat in tissue denser than the section median, at a local packing of 7,985 against 6,534 cells/mm2. At the data-pinned displacement, the oracle assignment accuracy there was 0.799, with an optimistic confidence-interval end of 0.86; at the densest tumor-nest packing of 9,329 cells/mm2 it was 0.783. Under every definition of the region the accuracy stayed below 0.90, and below 0.95, across the displacement confidence interval (Fig. 4a). The triple-positive signal therefore sits where assignment is not recoverable. PGR was the limiting marker, detected in only 2.9% of cells, consistent with the negative clinical PR status; apparent triple-positive cells numbered 165, or 0.098% of the section.

Second, can transcript displacement account for the signal? We built a synthetic region matching the measured packing and composition but containing, by construction, no truly triple-positive cell, then asked how often pure displacement manufactured an apparent one. At zero displacement the rate was 0%; at the data-pinned displacement, among cells adjacent to a progesterone-receptor source, it was 1.26%, rising to 1.96% at the upper end of the displacement interval. In the same adjacency condition the real data show 24.2% of such cells as apparent triple-positives, so pure displacement at the measured value under-produces the observed local rate roughly nineteenfold (Fig. 4b; full mechanism table and region localization in Extended Data Fig. 10). Displacement is therefore a contributing mechanism rather than a sufficient one: the apparent triple-positives are concentrated exactly where a proximity-driven misassignment artifact would place them, adjacent to progesterone-receptor sources, but their magnitude far exceeds what the best-possible assignment would leak, so most of the spurious signal arises from misassignment beyond the irreducible displacement floor rather than from displacement itself.

Third, what does independent evidence say? The original study's own matched dissociated single-cell data [2] found roughly 30 progesterone-receptor-positive cells, and these did not co-express ERBB2 or ESR1; the tissue is clinically PR-negative. Both point in the direction misassignment predicts, that the spatial measurement over-reports co-expression. Reduced sensitivity in the dissociated data does not explain this: ERBB2 was detected in 87% of cells, so its absence in those cells is not plausibly a dropout, and the clinical PR-negative status is immune to sequencing dropout entirely. The co-expression is therefore unrecoverable from the spatial data, consistent with a misassignment artifact, and independently contradicted.

Being above the ceiling is not by itself a verdict, which a contrasting case makes plain (Fig. 4c). Colorectal tumor tissue sits even deeper above the ceiling: on a public colorectal Xenium dataset, the densest we examined (median packing 17,424 cells/mm2) [11], oracle accuracy was 0.77 at the section median and 0.68 in

the densest tissue. A separately published colorectal claim of macrophage-tumor co-localization, in comparably dense tissue, is nonetheless corroborated by its own orthogonal data, matched single-cell and high-resolution spatial sequencing [12], so it is supported rather than flagged. A MERFISH hypothalamus claim of hybrid neurons [13], in a sparser regime where the same data give an oracle accuracy of 0.836 [14], is likewise corroborated by matched single-cell data. The ceiling identifies where a conclusion is unrecoverable; the orthogonal data, not the ceiling, decides whether it is spurious.

Discussion

Work on transcript assignment in imaging-based spatial transcriptomics has concentrated on building better segmentation and reassignment methods [3, 4, 5, 6, 15]. Our results add a prior question: in a given physical regime, how accurately can assignment be recovered at all? Because the oracle is optimal against a generative model calibrated to real data, its accuracy is a ceiling that bounds every method at once, and in dense tissue that ceiling lies well below complete recovery, between roughly 0.66 and 0.81. The best fairly configured published methods leave a further measurable fraction, 0.066 to 0.110, unrecovered beneath it.

The practical consequence is a diagnostic rather than another method. Given a dataset and a claim that depends on assignment, one can ask whether the claim sits above the ceiling, and if it does, treat the spatial measurement as insufficient on its own. The diagnostic does not declare such claims false, and the worked example shows why the distinction matters. A claim above the ceiling can still be true, as in the corroborated colorectal and hypothalamus cases, or artifactual, as the orthogonal evidence indicates for the breast triple-positive. The ceiling says where to be cautious; only orthogonal data settles the question. For a field that draws many co-expression and cell-cell interaction conclusions from dense tissue [4], a principled way to flag which of those conclusions require independent confirmation is the contribution we intend.

Several limitations bound the claim. The oracle is optimal relative to the generative model, so the ceiling transfers to real data only insofar as the model captures the features that govern assignment; we calibrated density, packing, displacement and expression overlap to real measurements and confirmed the bound survives realistic departures from the model's idealizations, but it remains a model-relative statement. The bound applies to the dense-tissue tail, where cells are packed most tightly, not to imaging-based spatial transcriptomics in general; in sparser tissue the ceiling rises. The displacement estimate, drawn only from markers with significant displacement signal, could in principle be biased upward by that selection, which would make the limit appear more severe; the bias is not decisive, since the displacement would have to fall roughly fivefold to lift dense accuracy to where the limit ceases to bite (Extended Data Fig. 7), and the lung dataset, where selection barely operated, gives the same displacement. Finally, the mechanism analysis shows that transcript displacement contributes to the spurious co-expression but does not by itself account for its magnitude, which exceeds what the best-possible assignment would leak by more than an order of magnitude; the conclusion that the co-expression is an artifact rests on its unrecoverability and the orthogonal contradiction, not on the displacement mechanism.

Methods

Synthetic generator

We generate a two-dimensional tissue field by sampling cell centers at a chosen packing density, partitioning space into cells, assigning each cell a type from a reference composition, and emitting transcripts per cell from per-type expression profiles. Transcripts are displaced from their cell of origin by a chosen displacement model before detection, and the true cell of origin of every transcript is retained. The three exposed parameters are mean transcripts per cell (density), cell packing (cells per unit area), and displacement. Expression profiles are realistic and overlapping, derived from real references or from clustering the real fields; emission is Poisson by default and negative-binomial when within-type variance is modeled. Geometry and displacement-shape variants (non-Voronoi cells, non-Gaussian displacement) are used for robustness. Exact parameter values, reference compositions, and grid definitions are in the committed code.

Oracle and the recovery ceiling

Given the known generative model, the oracle assigns each transcript to the cell of maximum posterior probability, combining the spatial likelihood under the displacement model with the gene likelihood under the per-type profiles. Oracle accuracy is the fraction of transcripts assigned to their true cell, evaluated on a converged spatial grid. At every grid point the oracle was confirmed to equal or exceed the naive nearest-center baseline.

Displacement pinning and the marker-validity test

Displacement is pinned from a spatial marker-leakage statistic: for each marker, the excess of its signal in cells adjacent to a marker-positive cell over distant cells. A marker is admissible only if this signal exceeds a position-permutation null (1,000 permutations, $P < 0.05$), a criterion defined without reference to the resulting displacement. Displacement is fit to the aggregate leakage of admissible markers, with a 2,000-resample bootstrap over markers and cells for the confidence interval. The test was applied identically across datasets and across alternative cell-typing choices.

Datasets and access

All datasets are public. Xenium human breast cancer (GEO GSE243280) [9]; Xenium human lung cancer (10x Genomics sample Xenium_V1_hLung_cancer_section, Human Lung Cancer Preview Data, Xenium Human Multi-Tissue and Cancer Panel, release 1.5.0) [10]; a 10x Genomics public human colorectal cancer Xenium release as the colorectal proxy (sample Xenium_V1_Human_Colorectal_Cancer_Addon_FFPE, Xenium V1 Human Colorectal Cancer Add-on FFPE, Xenium Onboard Analysis 2.0.0) [11]; and MERFISH mouse hypothalamus (Dryad) [14]. Cell types were obtained by clustering each dataset; markers were defined per dataset. Per-dataset cell counts, packing, and nearest-neighbor distances are summarized in Extended Data Fig. 9.

Segmenter-headroom protocol

Baysor (0.7.1) [3], Proseg (3.1.1) [5] and ComSeg (1.8.5) [6] were run on synthetic dense-tissue fields with known ground truth, each in its best documented configuration. In the primary comparison every method received the true nuclei as a prior, isolating assignment from segmentation error; a secondary free-segmentation comparison was run where supported. Methods were scored against the known truth under one-to-one matched accuracy, many-to-one homogeneity, and a co-assignment metric.

Worked lever analysis

The region of interest was located from the published spatial description and the released matrix, never by re-assigning transcripts. The oracle ceiling was evaluated at the region's own local packing. The mechanism demonstration used a synthetic region matched to the region's packing and composition with no true triple-positive cell by construction, a zero-displacement control, and an adjacency-conditioned spurious rate compared locally to the observed region. Orthogonal evidence was taken from the original study's matched dissociated data and the clinical annotation. No re-assignment of real transcripts was used to adjudicate the real claim.

Reproducibility, seeds, and determinism

All randomness uses fixed, recorded seeds. Results are append-only and regenerate byte-identically; a single command re-derives every committed result and figure. Cross-gate regeneration was verified to produce zero differences.

Use of AI assistance

An AI assistant (Claude, Anthropic) was used to support drafting and language editing of this manuscript. All study design, analyses, interpretation, and conclusions are the author’s own; the author reviewed and approved the final text and is solely responsible for its content.

Declarations

Code availability

The analysis code is available at https://github.com/RuiYamasaki/st-recoverability under the MIT License, with an installation guide and a demo on example data. A versioned archive of the release is deposited at Zenodo under the concept DOI https://doi.org/10.5281/zenodo.20760704, which resolves to the latest version.

Data availability

All datasets analyzed are publicly available, with accessions listed above. Only small derived summary files are distributed with the code; raw data are streamed from their original repositories.

References

1. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X (2015) Spatially resolved, highly multiplexed RNA profiling in single cells. Science 348:aaa6090. 10.1126/science.aaa6090
2. Janesick A et al (2023) High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. Nat Commun 14:8353. 10.1038/s41467-023-43458-x
3. Petukhov V, Xu RJ, Soldatov RA et al (2022) Cell segmentation in imaging-based spatial transcriptomics. Nat Biotechnol 40:345–354. 10.1038/s41587-021-01044-w

4. Mitchel J, Gao T, Petukhov V, Cole E, Kharchenko PV (2026) Impact and correction of segmentation errors in spatial transcriptomics. Nat Genet 58:434–444. 10.1038/s41588-025-02497-4
5. Jones DC, Elz AE, Hadadianpour A, Ryu H, Glass DR, Newell EW (2025) Cell simulation as cell segmentation. Nat Methods 22:1331–1342. 10.1038/s41592-025-02697-0
6. Defard T, Laporte H, Ayan M et al (2024) A point cloud segmentation framework for image-based spatial transcriptomics. Commun Biol 7:823. 10.1038/s42003-024-06480-3
7. Park J, Choi W, Tiesmeyer S et al (2021) Cell segmentation-free inference of cell types from in situ transcriptomics data. Nat Commun 12:3545. 10.1038/s41467-021-23807-4
8. Fu X et al (2024) BIDCell: biologically-informed self-supervised learning for segmentation of subcellular spatial transcriptomics data. Nat Commun 15:509. 10.1038/s41467-023-44560-w
9. Janesick A et al (2023) Integrated single-cell, spatial and in situ analysis of human breast cancer (FFPE): Xenium, Visium and Chromium scFFPE-seq [dataset]. Gene Expression Omnibus GSE243280
10. 10x Genomics (2026) Human Lung Cancer Preview Data, Xenium Human Multi-Tissue and Cancer Panel, release 1.5.0 (sample Xenium_V1_hLung_cancer_section) [dataset]. https://cf.10xgenomics.com/samples/xenium/1.5.0/Xenium_V1_hLung_cancer_section/
11. 10x Genomics (2026) Xenium V1 Human Colorectal Cancer Add-on FFPE, Xenium Onboard Analysis 2.0.0 (sample Xenium_V1_Human_Colorectal_Cancer_Addon_FFPE) [dataset]. https://cf.10xgenomics.com/samples/xenium/2.0.0/Xenium_V1_Human_Colorectal_Cancer_Addon_FFPE/
12. Oliveira MF, Romero JP, Chung M et al (2025) High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer. Nat Genet 57:1512–1523. 10.1038/s41588-025-02193-3
13. Moffitt JR, Bambah-Mukku D, Eichhorn SW et al (2018) Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science 362:eaau5324. 10.1126/science.aau5324
14. Moffitt JR, Bambah-Mukku D, Eichhorn SW et al (2018) Data from: Molecular, spatial and functional single-cell profiling of the hypothalamic preoptic region [dataset]. 10.5061/dryad.8t8s248. Dryad
15. Bilous M, Buszta D, Bac J et al (2026) Resolving sensitivity, specificity and signal contamination in Xenium spatial transcriptomics. Nat Methods. 10.1038/s41592-026-03089-8

Figures

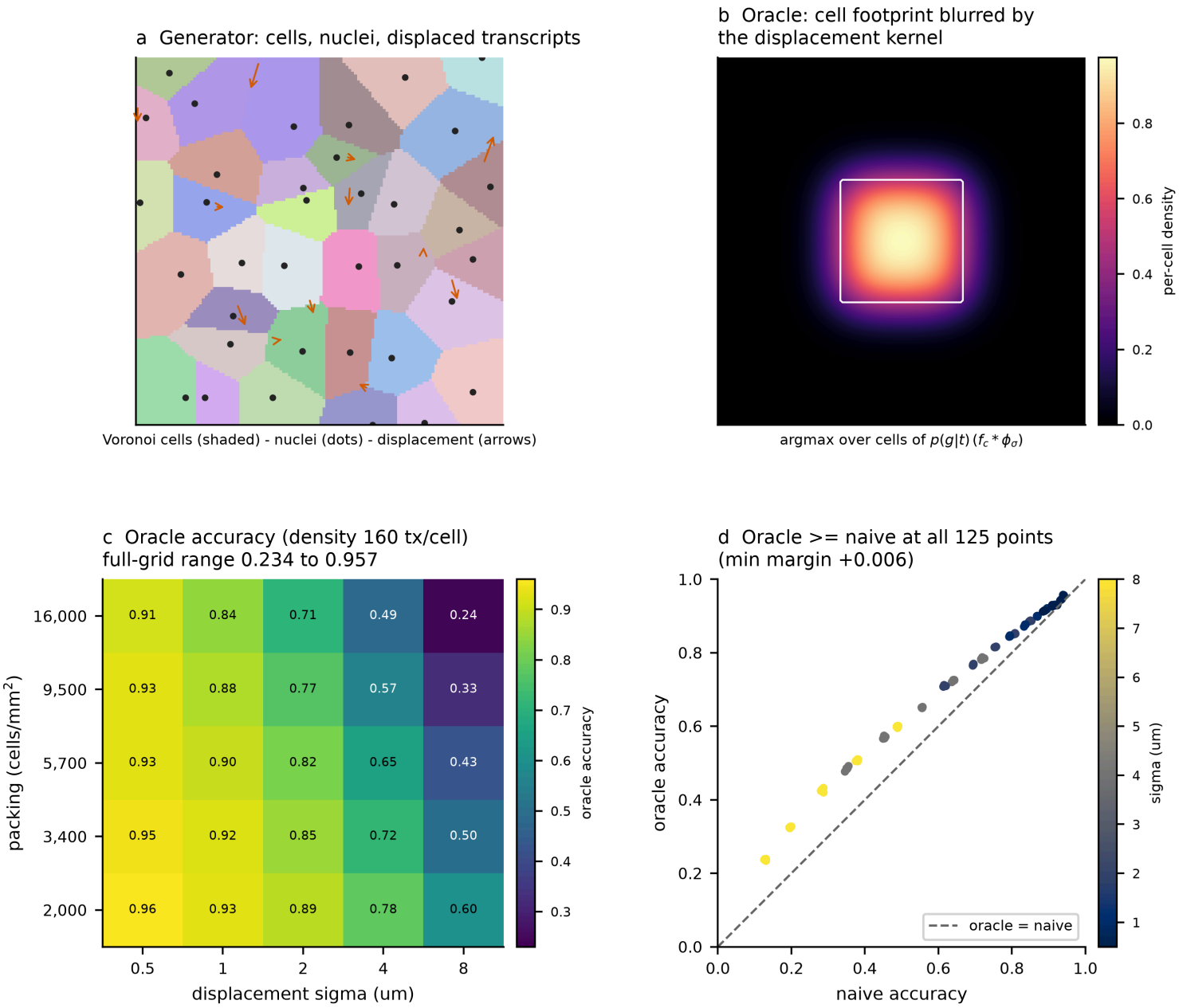


Figure 1

A by-construction ceiling on transcript-to-cell assignment. (a) The generative model: cells are placed at a set packing density, assigned types, and emit transcripts that are displaced from their cell of origin before detection; the true cell of origin of every transcript is recorded. (b) The oracle assigns each transcript to the cell of maximum posterior probability under the known model, the best accuracy achievable against that model. (c) The answerability frontier: oracle accuracy across a grid of transcript density, cell packing, and displacement, ranging from 0.234 to 0.957. (d) The oracle equals or exceeds a naive nearest-cell-center baseline at all 125 grid points.

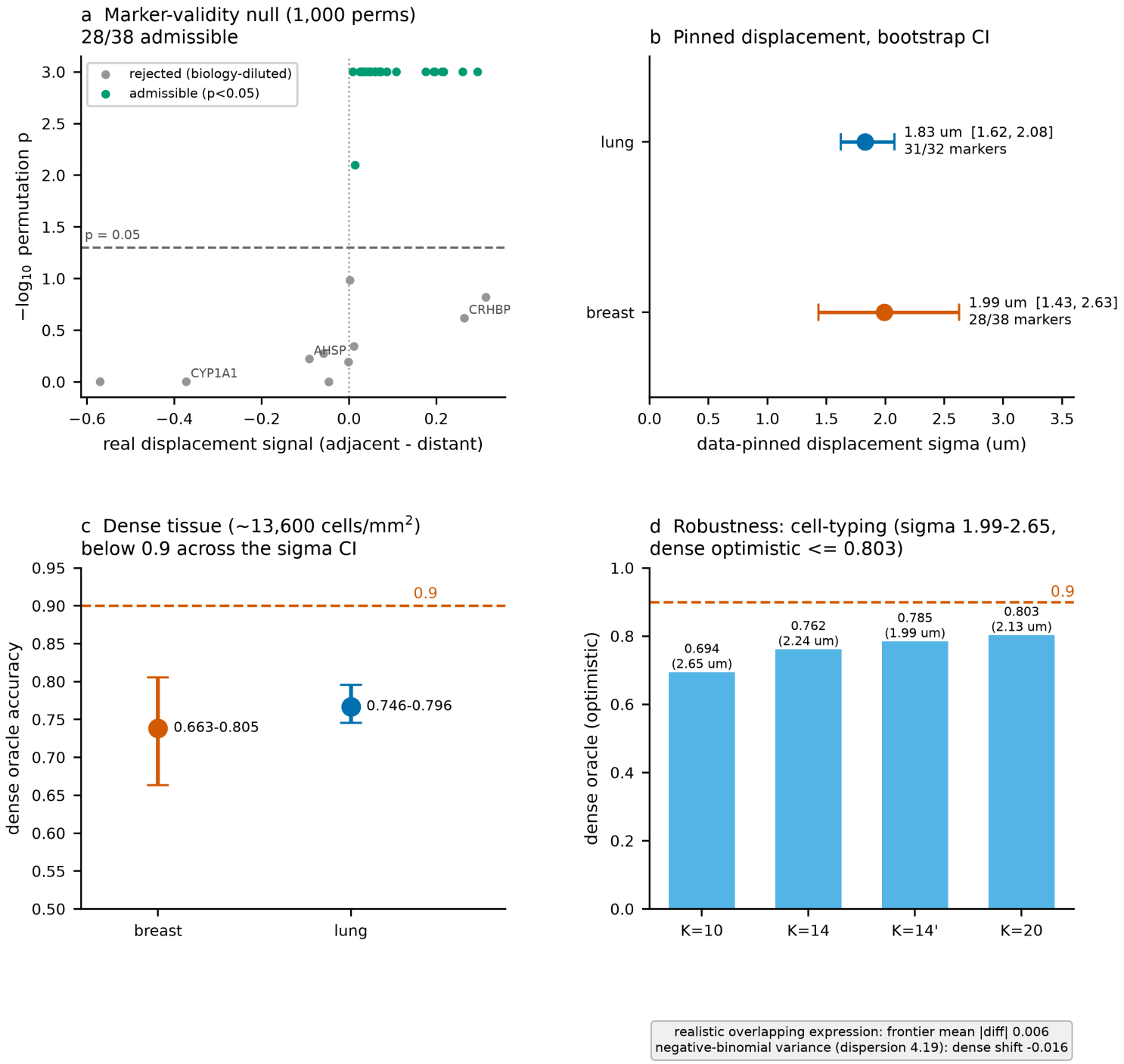


Figure 2

A data-pinned displacement places the bound on real tissue. (a) The marker-validity test: a marker is admissible only if its adjacent-minus-distant expression signal exceeds a position-permutation null ($P < 0.05$, 1,000 permutations), independent of the displacement it implies. (b) Displacement pinned from admissible markers with a 2,000-resample bootstrap confidence interval: 1.99 micrometres (95% CI 1.43 to 2.63) on Xenium breast (167,780 cells; 28 of 38 markers admissible) [9] and 1.83 micrometres (CI 1.62 to 2.08) on Xenium lung (150,365 cells; 31 of 32 admissible) [10]. (c) Dense-tissue oracle accuracy at about 13,600 cells/mm2 across the displacement CI: 0.663 to 0.805 (breast) and 0.746 to 0.796 (lung), all below 0.9. (d) The bound under realistic overlapping expression (assignment-frontier mean absolute difference 0.006; oracle-minus-naive gap 0.071 versus 0.066), negative-binomial within-type variance (dispersion near 4.19; dense shift

-0.016), and across cell-typing choices (displacement 1.99 to 2.65 micrometres; dense accuracy at most 0.803).

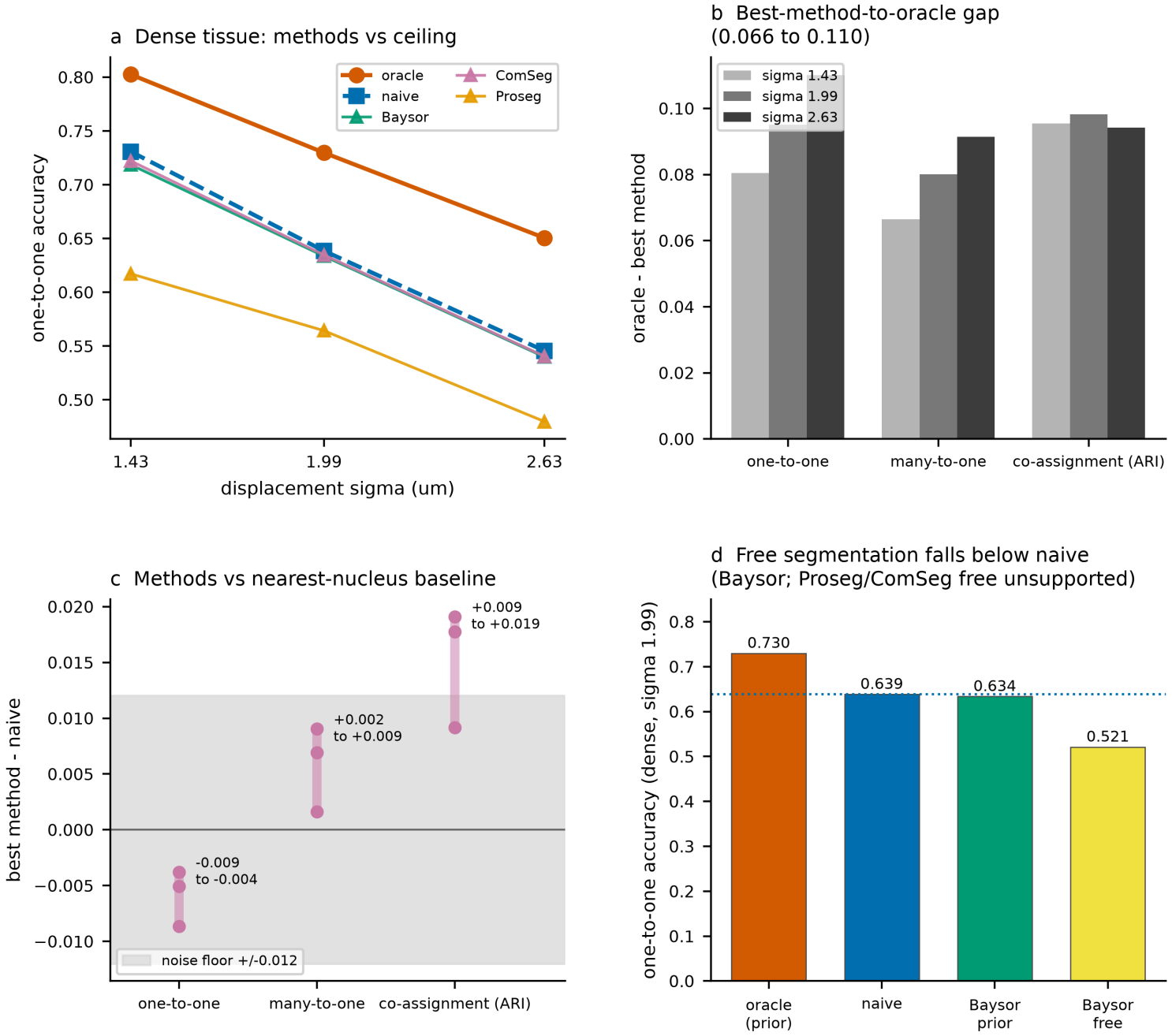
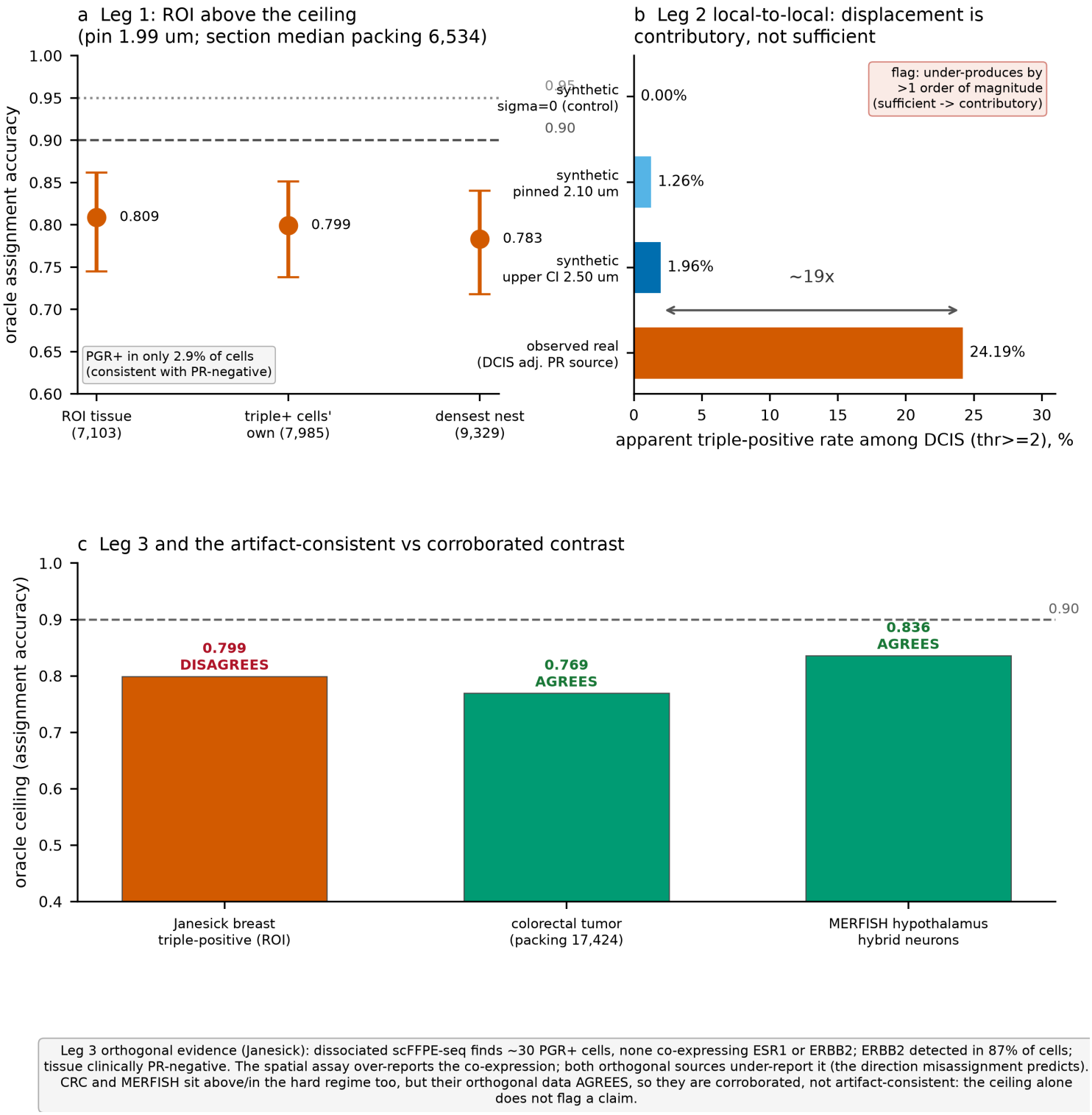


Figure 3

Current methods leave a measurable fraction of the ceiling unrecovered. (a,b) Baysor, Proseg and ComSeg [3,5,6], each in its best documented configuration and supplied with true nuclei, against the oracle in dense tissue; the best method remains 0.066 to 0.110 below the oracle across three over-segmentation-robust metrics. (c) Comparison with nearest-nucleus assignment is metric-dependent and within the methods' run-to-run noise (about 0.012): the best method is 0.004 to 0.009 below under a one-to-one metric, up to 0.009 above under many-to-one homogeneity, and up to 0.019 above under co-assignment. Without a nuclei prior, only one method runs de novo and falls below nearest-nucleus.



Leg 3 orthogonal evidence (Janesick): dissociated scFFPE-seq finds ~30 PGR+ cells, none co-expressing ESR1 or ERBB2; ERBB2 detected in 87% of cells; tissue clinically PR-negative. The spatial assay over-reports the co-expression; both orthogonal sources under-report it (the direction misassignment predicts). CRC and MERFISH sit above/in the hard regime too, but their orthogonal data AGREES, so they are corroborated, not artifact-consistent: the ceiling alone does not flag a claim.

Figure 4

A diagnostic applied to a published claim. (a) Leg 1: the triple-positive region sits in dense tissue (local packing 7,985 versus a section median of 6,534 cells/mm2), where oracle accuracy is 0.799 (optimistic CI end 0.86) and 0.783 at the densest nest (9,329 cells/mm2), below 0.90 and 0.95 across the displacement CI; PGR is detected in 2.9% of cells. (b) Leg 2: a synthetic region with no true triple-positive cell by construction; pure displacement produces 0% spurious apparent triple-positives at zero displacement and 1.26% to 1.96% among cells adjacent to a progesterone-receptor source at the pinned displacement, against 24.2% observed in the

same adjacency condition in the real data, a roughly nineteenfold under-production that marks displacement as a contributing but not sufficient mechanism. (c) Leg 3 and contrast: the original matched dissociated data [2] show roughly 30 PGR-positive cells not co-expressing ERBB2 or ESR1 (ERBB2 detected in 87% of cells; tissue clinically PR-negative); a public colorectal dataset is deeper above the ceiling (oracle 0.77 median, 0.68 dense; median packing 17,424 cells/mm2) [11], where a separately published colorectal co-localization claim is orthogonally corroborated [12], and a MERFISH hypothalamus claim [13] is corroborated where the same data give an oracle accuracy of 0.836 [14].

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- extendeddata.docx
- supplementaryinformation.docx