A056 · Consistent and scalable detection and comparison of spatial patterns

Presenter: Jiayu Su — Broad Institute

Authors: Jiayu Su, Jun Hou Fung, Haoyu Wang, Dian Yang, Xiao Wang, David A. Knowles, Raul Rabadan

Session: Day 1 · Thu Oct 1, 2026 · 9:45–10:45 AM

Since Robert Hooke sketched cells in cork in 1665, biological discovery has depended on recognizing spatial organization. Spatial omics brings this quest to molecular scale, but noisy, sparse data pose two computational challenges: distinguishing patterns from noise within one sample, and determining how they change across samples. Existing spatial-variability tests address detection but often disagree without explanation, while comparison methods require coordinate registration, restricting analysis to matched sections.

We introduce SONIC (Spatial Organization through Nonrandom-pattern Inference and Comparison), a harmonic framework that addresses both. For detection, we show that virtually all approaches—including Moran's I, Gaussian processes, regression models, and dependency tests—share the quadratic form Q = z^T Kz. Characterizing their effective kernels in the frequency domain explains why tests disagree: no weighting is optimal for all patterns, while mixed signs in Moran's I permit spectral cancellation, causing inconsistency and power loss. Guided by this theory, we redesign spatial-variability tests with positive-definite kernels, permutation-free calibration, and accelerated algorithms that scale to millions of locations.

This spectral lens extends to multi-sample comparison. Each pattern's power spectral density provides a shared physical-frequency coordinate system. Our differential-frequency test compares spectra across complex designs without tissue registration, while DF-norm isolates frequency composition from total spectral power. Across applications, our approach identifies two plaque-induced response programs in Alzheimer's disease, IDH-associated extracellular-matrix and coagulation programs relevant to venous thromboembolism in glioma, and conserved spatiotemporal gene dynamics during mammalian embryogenesis.

Together, SONIC unites single-sample detection and registration-free comparison within one harmonic framework for large-scale spatial discovery.

Consistent and scalable detection and comparison of spatial patterns

Jiayu Su[1,2,6,7,*], Jun Hou Fung[1,2], Haoyu Wang[2,3], Dian Yang[2,3],
Xiao Wang[7], David A. Knowles[2,4,6], and Raul Rabadan[1,2,5]

[1]Program for Mathematical Genomics; [2]Department of Systems Biology; [3]Department of Molecular Pharmacology and Therapeutics;
[4]Department of Computer Science; [5]Department of Biomedical Informatics, Columbia University, New York, NY, USA
[6]New York Genome Center, New York, NY, USA; [7]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[*]Presenting author; sujiayu@broadinstitute.org

Motivation. Since Robert Hooke sketched cells in cork in 1665, biological discovery has depended on recognizing spatial organization. Spatial omics brings this quest to the molecular scale, but noisy, sparse measurements pose two computational challenges: distinguishing patterns from noise within one sample (*detection*) and determining how they change across samples (*comparison*). Existing spatial-variability (SV) tests address detection but often disagree, with no unified theory explaining why; cross-sample methods generally require registration to common coordinates, restricting comparison to consecutive or closely matched sections. We introduce SONIC (Spatial Organization through Nonrandom-pattern Inference and Comparison), a unified harmonic framework for both challenges.

Single-sample detection: a unified theory of SV tests. For expression $\mathbf{x}$ observed at coordinates $\mathbf{s}_1, \ldots, \mathbf{s}_n$, SV testing asks whether $\mathbf{x}$ varies systematically with location rather than being randomly distributed. We show that major approaches—including Moran's I, Gaussian processes, regression models, and nonparametric dependency tests—share the quadratic form $Q = \mathbf{z}^\top K \mathbf{z}$ for centered, unit-variance expression $\mathbf{z}$ and can detect only changes in mean expression across locations (Figure 1A). In the frequency domain, the effective kernel K is simply a weighting of spatial frequencies, explaining why tests disagree: each favors different patterns, and none is uniformly best. Zero eigenvalues create blind spots, while mixed-sign weights permit spectral cancellation, making methods such as Moran's I susceptible to inconsistency and power loss. In single-cell lineage-tracing data, where the goal is to identify heritable expression along a phylogenetic tree, Moran's I missed rare-subclone markers because high-frequency variation within clones obscured sparse signals on the lineage graph.

Scalable implementation. Guided by this theory, we redesign SV tests with consistent, positive-definite kernels with scalability in mind: Sparse precision operators reduce quadratic memory to $O(n \log n)$ on single-cell graphs, while FFTs and NUFFTs reduce quadratic computation to $O(n \log n)$ for spatial data, enabling million-spot screening under one second.

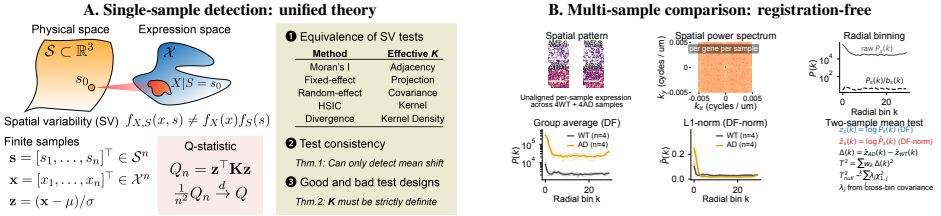

Figure 1. A, Major SV tests share a quadratic form whose frequency weighting determines their statistical behavior. **B,** Spatial power spectra place patterns from spatially unregistered samples in a shared physical-frequency coordinate system for multi-sample comparison and rigorous statistical testing.

Multi-sample comparison without registration. This spectral representation extends naturally to multi-sample comparison (Figure 1B). For each feature and sample, we aggregate squared Fourier magnitudes into radial bins by frequency magnitude, producing a one-dimensional power spectrum $P_s(k)$ indexed in physical units. These spectra provide shared coordinates for patterns from heterogeneous sections without registering their original spatial coordinates. Our *differential-frequency (DF)* test compares log spectra across samples with a covariance-aware quadratic statistic; its L1-normalized counterpart, *DF-norm*, tests frequency composition independently of total expression change. Across applications, the approach identifies two plaque-induced response programs in Alzheimer's disease, IDH-associated extracellular-matrix and coagulation programs relevant to venous thromboembolism in glioma, and conserved spatiotemporal gene dynamics during mammalian embryogenesis.

Conclusion. Together, SONIC unites single-sample detection and registration-free comparison within one harmonic framework for large-scale spatial discovery. The detection theory and benchmarks are described in arXiv:2602.02825.