

Spatial association in high-dimensional settings, with application to local cell aggregation in spatial transcriptomics

Koen Van den Berge¹

in collaboration with Sofía Prieto Leon², Ewoud De Troyer¹, Helena
Geys^{1,2}, Olivier Thas²

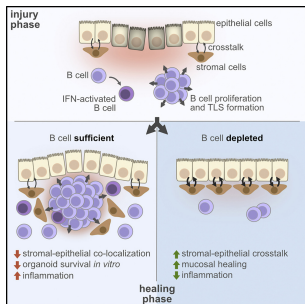
¹ Johnson & Johnson ²Hasselt University

September 2026

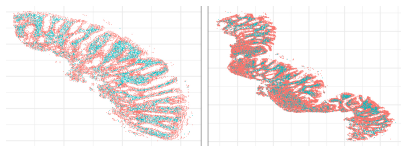
Good intentions gone wrong: The B cell block to epithelial repair

Emilie K. Grasset^{1,2,4} and Theresa Alenghat^{3,5,*}

Cellular dynamics that influence mucosal healing are not well understood. In this issue of *Immunity*, Frede, Czarzewski, Monasterio et al. find that B cells accumulate in the colon following intestinal injury. These B cells impair epithelial repair by hindering local stromal-epithelial interactions.



We collected 145 colon biopsies from IBD patients, for spatial omics:



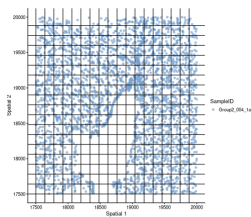
Goal is to identify B-cell aggregates and link them to clinical response.

Hot- and cold spots are detected using methods from the nineties

- ▶ Consider data in 2D (x, y) -space, such as
 - ▶ Bicycle accidents in a city.
 - ▶ Gene expression in biological tissue.
- ▶ Task is to locate clusters of unusually high or low values: hot spots and cold spots.
 - ▶ Dangerous crossroads.
 - ▶ Immune cell aggregates.

Hot- and cold spots are detected using methods from the nineties

- ▶ Consider data in 2D (x, y) -space, such as
 - ▶ Bicycle accidents in a city.
 - ▶ Gene expression in biological tissue.
- ▶ Task is to locate clusters of unusually high or low values: hot spots and cold spots.
 - ▶ Dangerous crossroads.
 - ▶ Immune cell aggregates.
- ▶ Getis & Ord (1992); Ord & Getis (1995) introduce the local Getis-Ord (GO) statistic, later formalized by Anselin (1995).
- ▶ Such methods “... give an indication of the extent of significant spatial clustering of similar values around that observation.” (Anselin, 1995).
- ▶ All methods work by imposing a grid on the 2D space.



The Getis-Ord (GO) test statistic and inference

The Getis-Ord (GO) statistic, with w_{ij} a binary neighbor indicator, is defined as

$$G_i^* = \frac{\sum_j w_{ij} x_j}{\sum_j x_j}.$$

Common approach resorts to standardizing the GO statistic, assuming asymptotic normality via the combinatorial CLT, i.e.,

$$z(G_i^*) = \frac{G_i^* - E_0[G_i^*]}{\sqrt{\text{Var}_0[G_i^*]}} \sim N(0, 1)$$

Three approaches exist for statistical inference.

The Getis-Ord (GO) test statistic and inference

The Getis-Ord (GO) statistic, with w_{ij} a binary neighbor indicator, is defined as

$$G_i^* = \frac{\sum_j w_{ij} x_j}{\sum_j x_j}.$$

Common approach resorts to standardizing the GO statistic, assuming asymptotic normality via the combinatorial CLT, i.e.,

$$z(G_i^*) = \frac{G_i^* - E_0[G_i^*]}{\sqrt{\text{Var}_0[G_i^*]}} \sim N(0, 1)$$

Three approaches exist for statistical inference.

1. **Analytical permutation:** Closed form, exact moments by reasoning about all $n!$ rearrangements (Getis & Ord, 1992).

The Getis-Ord (GO) test statistic and inference

The Getis-Ord (GO) statistic, with w_{ij} a binary neighbor indicator, is defined as

$$G_i^* = \frac{\sum_j w_{ij} x_j}{\sum_j x_j}.$$

Common approach resorts to standardizing the GO statistic, assuming asymptotic normality via the combinatorial CLT, i.e.,

$$z(G_i^*) = \frac{G_i^* - E_0[G_i^*]}{\sqrt{\text{Var}_0[G_i^*]}} \sim N(0, 1)$$

Three approaches exist for statistical inference.

1. **Analytical permutation:** Closed form, exact moments by reasoning about all $n!$ rearrangements (Getis & Ord, 1992).
2. **Monte Carlo permutation:** Actually permute the data, to estimate null mean/variance (Anselin, 1995).

The Getis-Ord (GO) test statistic and inference

The Getis-Ord (GO) statistic, with w_{ij} a binary neighbor indicator, is defined as

$$G_i^* = \frac{\sum_j w_{ij} x_j}{\sum_j x_j}.$$

Common approach resorts to standardizing the GO statistic, assuming asymptotic normality via the combinatorial CLT, i.e.,

$$z(G_i^*) = \frac{G_i^* - E_0[G_i^*]}{\sqrt{\text{Var}_0[G_i^*]}} \sim N(0, 1)$$

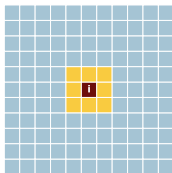
Three approaches exist for statistical inference.

1. **Analytical permutation:** Closed form, exact moments by reasoning about all $n!$ rearrangements (Getis & Ord, 1992).
2. **Monte Carlo permutation:** Actually permute the data, to estimate null mean/variance (Anselin, 1995).
3. **Monte Carlo permutation null:** Same permutation, but used to build an empirical null distribution without standardizing G_i^* (Ord & Getis, 1995).

Problem 1: Permutation schema assumptions do not hold in spatial omics

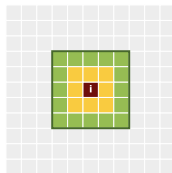
- ▶ The permutation argument requires global exchangeability.
- ▶ Data in spatial omics are typically autocorrelated → not exchangeable.
- ▶ This leads to invalid inference: null distribution too narrow.
- ▶ **Solution:** We suggest a new permutation schema that only assumes exchangeability within a neighborhood.

Standard: draw from the **WHOLE** grid



Reshuffles values across the entire domain → assumes complete spatial randomness. Destroys the ambient autocorrelation, so the null is too narrow → hotspots flagged from global structure alone.

Proposed: draw from a local window ($m > k$)



Resamples only within the m -order window → keeps the local level, trend and correlation, Tests 'is this k -neighbourhood exceptional vs. its own surroundings?' (window large enough for stable moments).



focal cell i



k -order neighbourhood



global pool



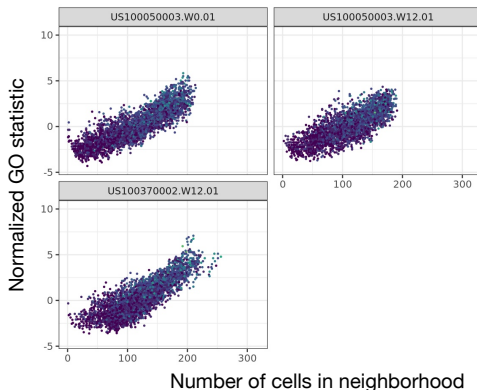
local window pool



excluded

Problem 2: GO statistic is affected by technical variation (cell density)

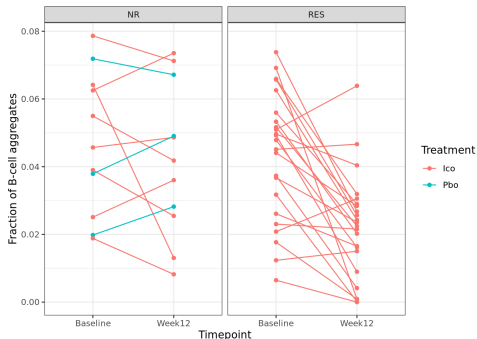
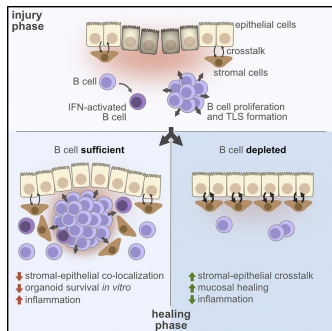
- ▶ The GO statistic increases as cell density of its neighborhood $\sum_j w_{ij}$ increases.
- ▶ This is also expected under H_0 , but not effectively normalized due to underestimation of $E_0(G_i)$.
- ▶ Plot below shows association under H_0 .
- ▶ **Solution:** Calculate GO statistic using relative abundances.



Problem 3: Interpretation of significant Getis-Ord tests in the wild

- ▶ Literature usually interprets via autocorrelation: a cell's value influences the value of other cells.
- ▶ But, significant G_i^* can be due to (a) a mean trend; (b) autocorrelation.
 - ▶ In reality, it only means "*This neighborhood departs from spatial randomness*".
- ▶ The mechanism (trend vs autocorrelation) is not separately identifiable from a single realization — you must formalize by assuming which one you're testing!
- ▶ **Solution:** We devise two combined sets of null hypothesis and required assumptions to aid interpretation.

B-cell aggregates decrease in responders to Icotrokinra treatment in inflammatory bowel disease



Thank you.

Local statistic (focal included)

w_{ij} : neighbour indicator; G_i excludes focal cell ($j \neq i$).

$$G_i^* = \frac{\sum_j w_{ij} x_j}{\sum_j x_j}$$

Null mean (exchangeability)

expected neighbour share under spatial randomness

$$E_0[G_i^*] = \frac{W_i}{n}, \quad W_i = \sum_j w_{ij}$$

Null variance (analytical permutation)

$$\sigma_0^2 = \frac{1}{n} \left(\sum_r x_r^2 - \bar{X}^2 \right)$$

$$\text{Var}_0[G_i^*] = \frac{\sigma_0^2 W_i (n - W_i)}{n^2 (n - 1) \bar{X}^2}$$

Standardised score

asymptotic normality via the combinatorial CLT

$$z(G_i^*) = \frac{G_i^* - E_0[G_i^*]}{\sqrt{\text{Var}_0[G_i^*]}} \sim N(0, 1)$$

Monte-Carlo permutation p-value

rank-based; l permutations, l_{ext} as-or-more extreme

$$p_i^{\text{perm}} = \frac{l_{\text{ext}} + 1}{l + 1}$$

Johnson & Johnson
Innovative Medicine

Eqs (3)–(4), (7), (18), (19), (22). Null moments hold only under exchangeability.