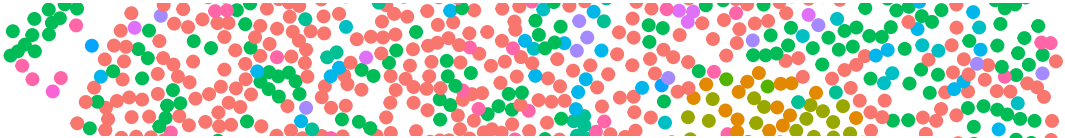
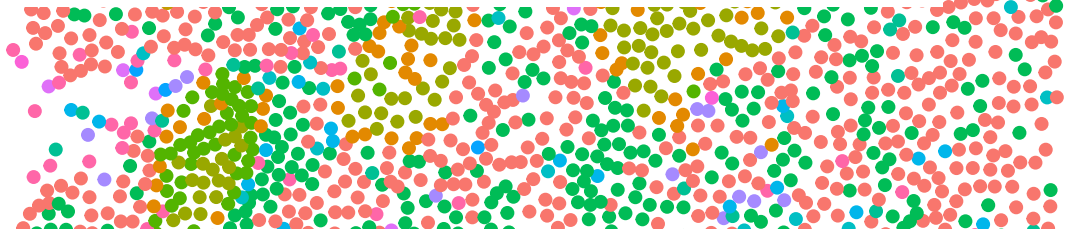


Sequencing-based Spatial Transcriptomics Data Analysis

Spatial Statistics for Differential Analysis

Martin Emons, Statistical Bioinformatics Group, UZH & SIB

December 10, 2025



Outline

- `pasta` - a resource for exploratory spatial analysis
- Lattice data analysis to compare spatial associatons on a grid
- Point pattern analysis for comparisons of stochastic processes
- Comparison of spatial statistics functions with `spatialFDA`

Cover picture data from [Damond et al., 2019]

Technological Overview

- High-throughput sequencing (HTS) of the transcriptome allows for high-throughput characterisation of cells
- Traditional HTS platforms work either on a homogenised sample (bulk) or on single cells (sc)
- These approaches come at the loss of spatial information
  - Tissue is either homogenised (bulk) or sorted by cells (sc)
- Recent advances allow for the characterisation of cells in their spatial environment

[Rao et al., 2021]

Imaging vs. HTS-based

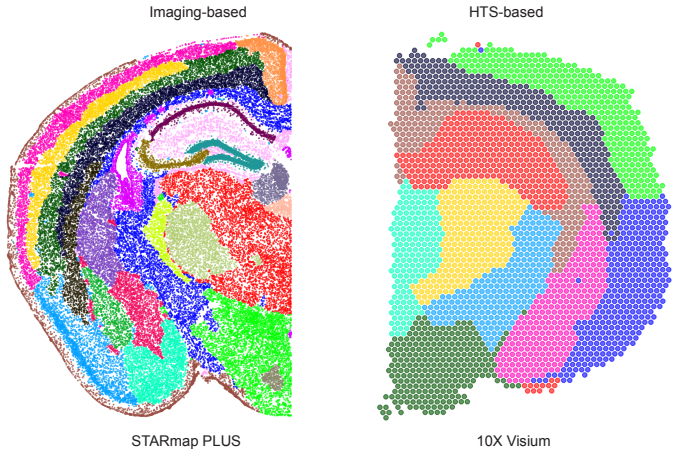


Figure: Data: [Shi et al., 2023, 10X, 2022], clustering: [Singhal et al., 2024]

Imaging vs. HTS-based

Imaging-based

Pros

- higher sensitivity
- better resolution
- adjustable area of detection

Cons

- lower number of features → targeted
- trade off area / acquisition time

→ hypothesis-testing

[Rao et al., 2021, Moffitt et al., 2022]

HTS-based

Pros

- higher number of features → untargeted

Cons

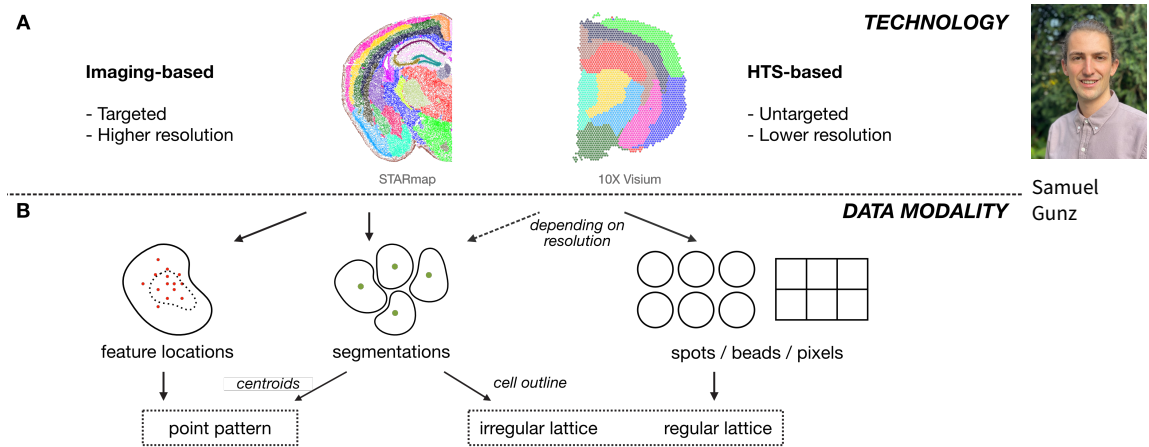
- lower sensitivity
- limited resolution (55 μm diameter Visium, 2 μm diameter Visium HD)
- standard area of detection (arrays)

→ hypothesis-generating

`pasta` a resource for exploratory spatial analysis

- `pasta` is a webpage with several vignettes in R and Python to introduce and organise exploratory spatial analysis
- shows code examples for different technologies
- Assumptions depending on the data type are discussed and put into context

Technologies differ in their data modalities



Samuel Gunz

Figure: [Emons et al., 2025, Rao et al., 2021], Data: [Shi et al., 2023, 10X, 2022]

Analysis options differ between technologies

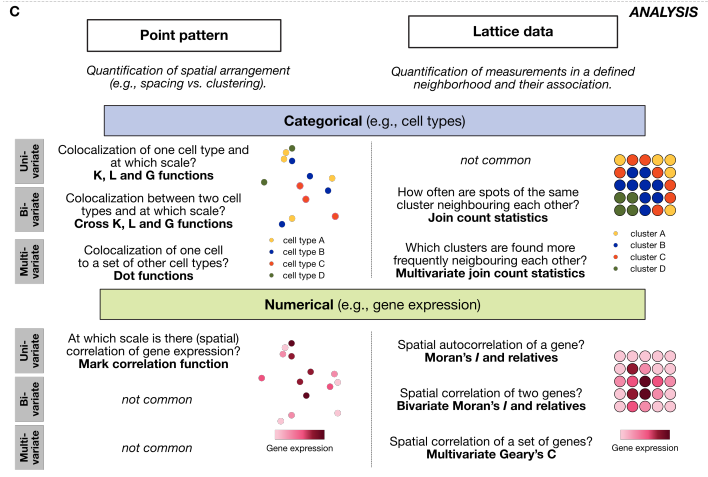
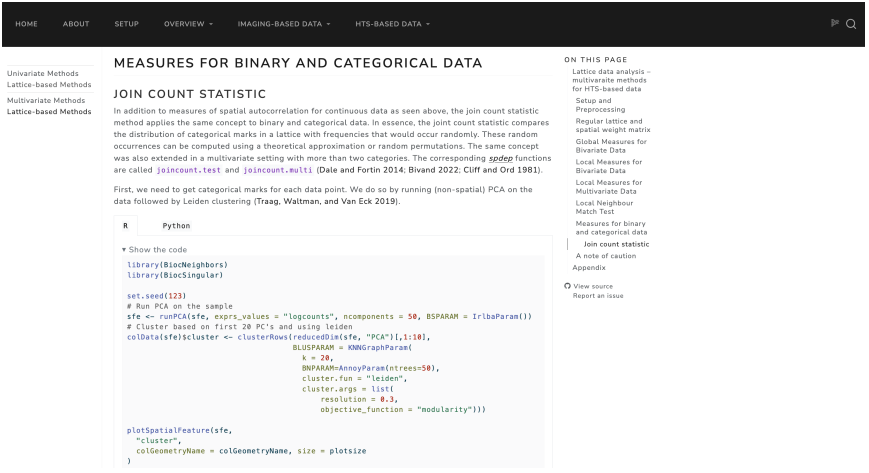


Figure: [Baddeley et al., 2015, Pebesma and Bivand, 2023, Emons et al., 2025]

Samuel Gunz

Vignettes in R and Python for analyses



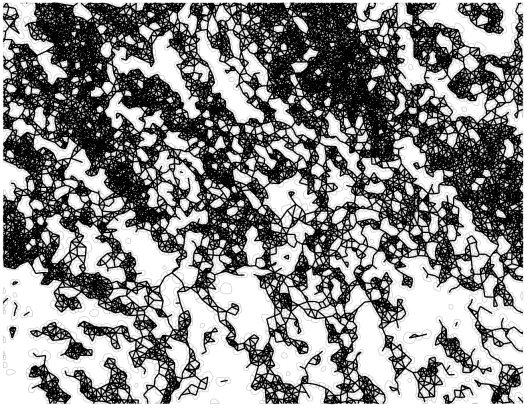
available under `robinsonlabuzh.github.io/pasta`

Spatial association measures per neighbourhood

$$\sum_i \sum_j f(x_i, x_j) w_{ij}$$

- Generally, lattice data analysis metrics are a double sum over all locations
- Neighbourhood is defined in a weight matrix w_{ij}.
- Typical choices are contiguity-, k-NN or distance based neighbourhoods.
- For a given neighbourhood, a measure of spatial association $f(x_i, x_j)$ is calculated
- More details in [Moses et al., 2023] and [Emons et al., 2025].

Weight matrix construction depends on the biological question



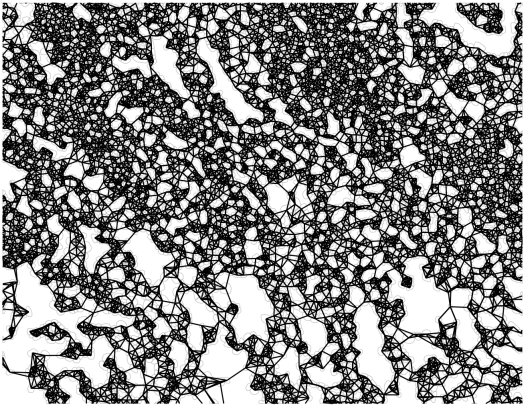


Figure: NSCLC CosMx Data from [He et al., 2022]

Weight matrix construction will influence results

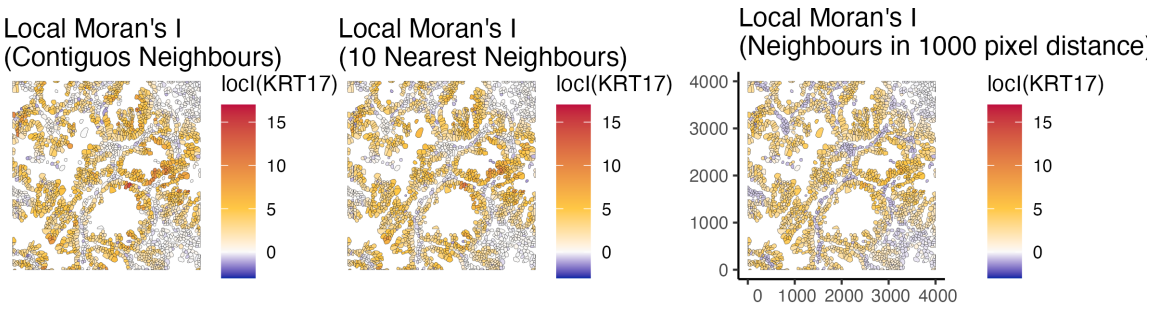


Figure: NSCLC CosMx Data from [He et al., 2022]

Spatial autocorrelation of gene expression

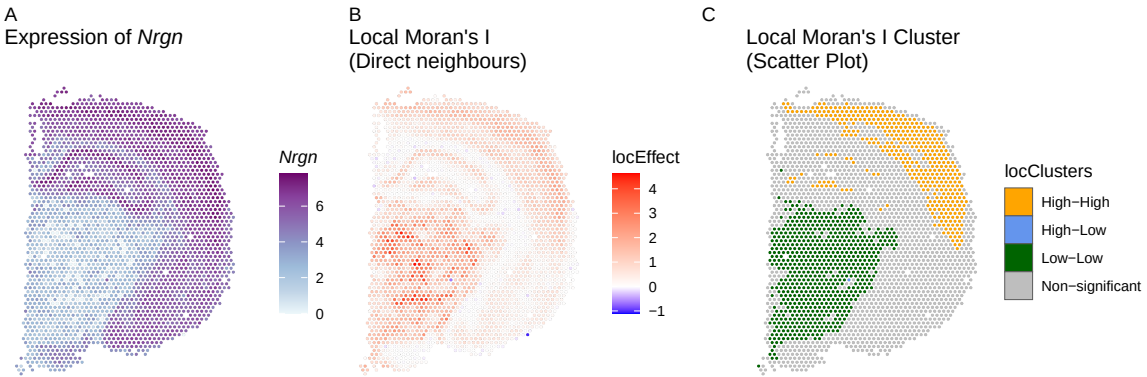


Figure: Mouse coronal brain 10x Visium data from [10X, 2022]

Lattice data analysis reveals triple positive receptor regions

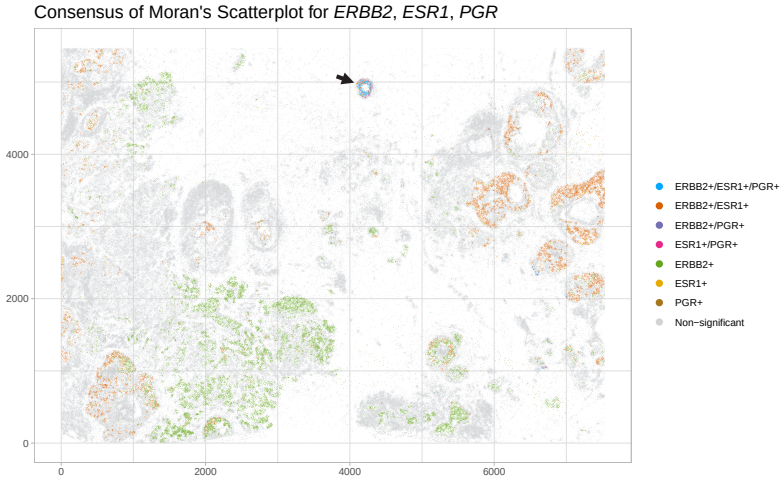
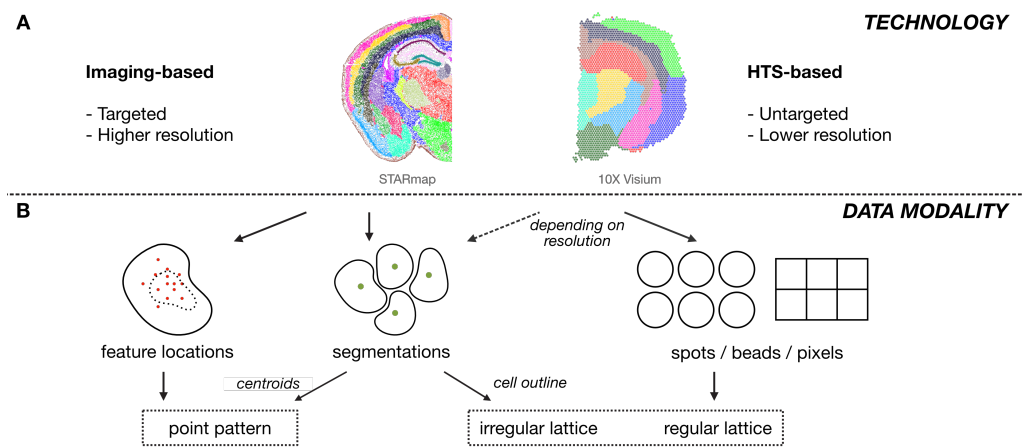


Figure: Breast cancer data from 10x Xenium [Janesick et al., 2023, Emons et al., 2025]

Technologies differ in their data modalities



Samuel Gunz

Figure: [Emons et al., 2025, Rao et al., 2021], Data: [Shi et al., 2023, 10X, 2022]

Point pattern analysis compute curves across radii r

$$\hat{K}(r) = \frac{|W|}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \{d_{ij} \leq r\} e_{ij}(r)$$

- Compute a measure of spatial assocation in a r-neighbourhood $\rightarrow$ distance-based neighbourhood
- Calculate the measure of spatial assocation across all radii r.
- Normalises for differences in field-of-view ($|W|$) and number of points (n)
- Corrects for edge-effects at the image boundary $e_{ij}(r)$

Edge effects are differences in association due to the FOV border

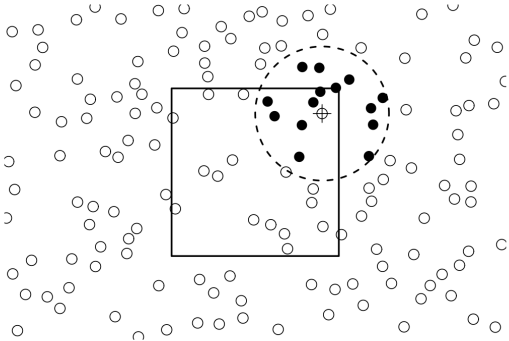


Figure 7.12. *Counting, without edge effects, the number of neighbours of each wildflower within a sampling frame (black rectangle) in a field of wildflowers.*

Figure: Taken from [Baddeley et al., 2015]

Point pattern analysis summarises cellular arrangements

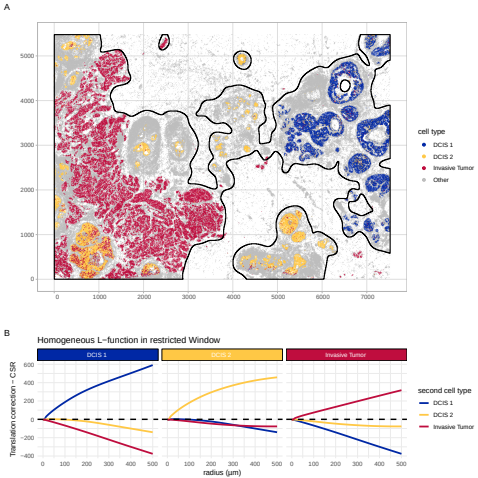


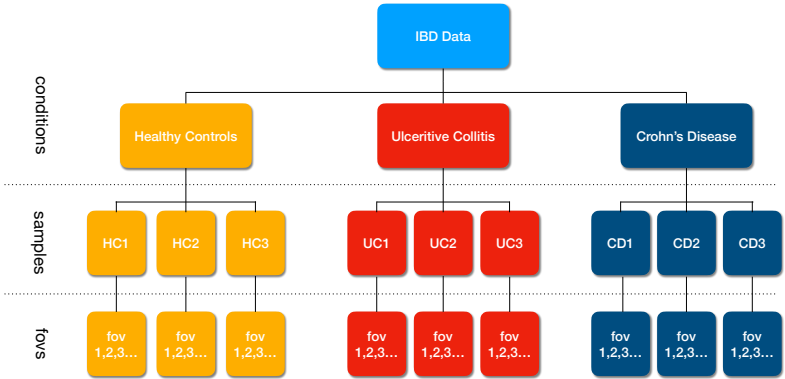
Figure: Breast cancer data from 10x Xenium [Janesick et al., 2023, Emons et al., 2025]

Conclusions `pasta`: Highlights the usability of spatial statistics in R and Python

- vignettes introduce and contextualise known concepts from spatial statistics and their usage for Omics data.
- point pattern analysis is shown in R with the package `spatialFDA`
- lattice datta analysis is shown in R with `Voyager` and in Python with `Pysal`
- paper available for further detail [Emons et al., 2025]

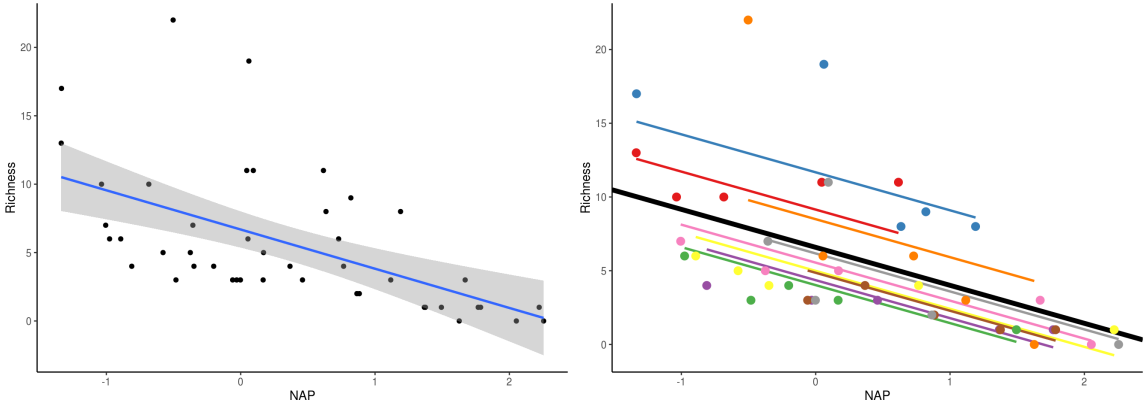
Modern datasets have complex nested variance structures

- imaging-based spatial transcriptomics (CosMX)
- compares three conditions - Healthy, Crohn's Disease, Ulceritive Collitis
- per condition three samples
- per sample $\sim$ 20 FOVs



[Garrido-Trigo et al., 2023]

Mixed-effects to account for nested covariance



Taken from [Santangelo, 2018], licensed as `CC BY 4.0`

`spatialFDA` set up

- Open question in the field is how to efficiently and sensitively compare different co-localisation of cells across samples and/or conditions
- There are only a few options $\rightarrow$ often limited to scalar comparisons [Canete et al., 2022, Hawinkel et al., 2025] or without random effect estimation [Seal et al., 2024]
- aim: build a framework to compare spatial statistics functions over their entire domain r.

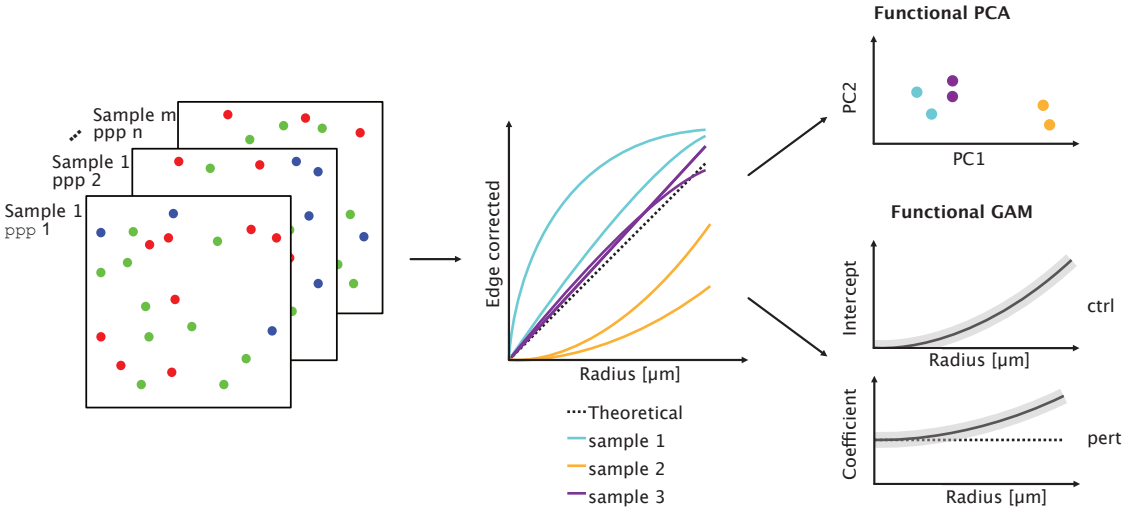
State of the art I

package	metric	full distribution	testing
spicyR	AUC of L function	no	linear (mixed effects) model
smoppix	AUC of G function	no	linear mixed effects model
SpaceANOVA	pair correlation function	yes	functional ANOVA

areas of improvement:

- improved functional testing
- systematic evaluation and validation of approaches
- larger flexibility
- software for simple installation and integration into existing analysis workflows → `Bioconductor` package

`spatialFDA`: Differential co-localisation across scales



Functional additive mixed model

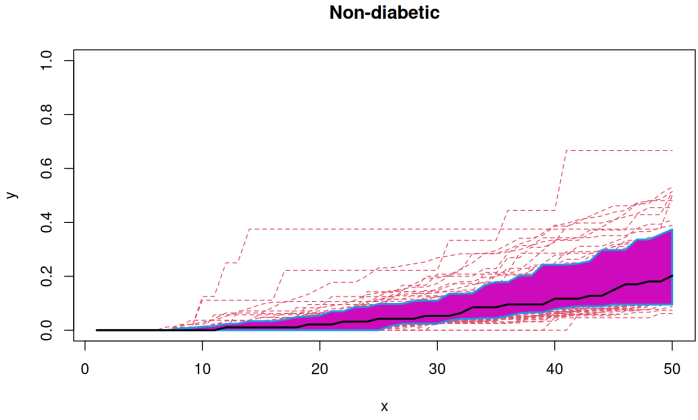
— Compare not a summary statistic between conditions but the function over the domain r
$\rightarrow$ Functional data analysis (FDA)

— Additive mixed model with a functional response

$$\underbrace{\mathbb{E}[y_i(r)]}_{\text{functions}} = \underbrace{\alpha(r)}_{\text{intercept}} + \underbrace{\beta_{0,g(i)}(r)}_{\text{random intercept}} + \underbrace{\beta_{0,i}}_{\text{offset}} + \underbrace{\sum_{j=1}^{J} f_j(X_{ji}, r)}_{\text{predictors}} + \underbrace{\epsilon_i(r)}_{\text{errors}}$$

[Scheipl et al., 2015, Scheipl et al., 2016]

Challenges in inference on functions

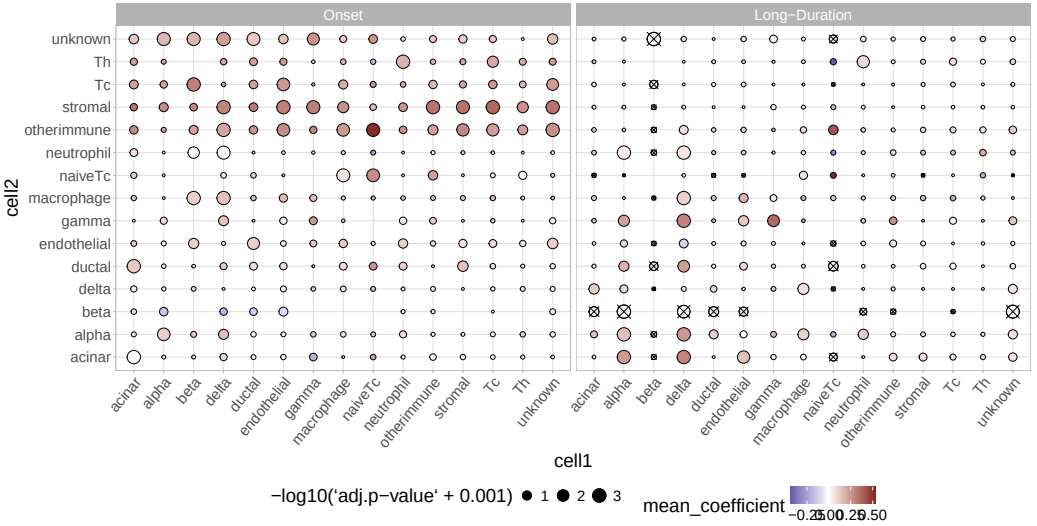


— Heteroscedasticity: Data-transformations and sandwich corrections

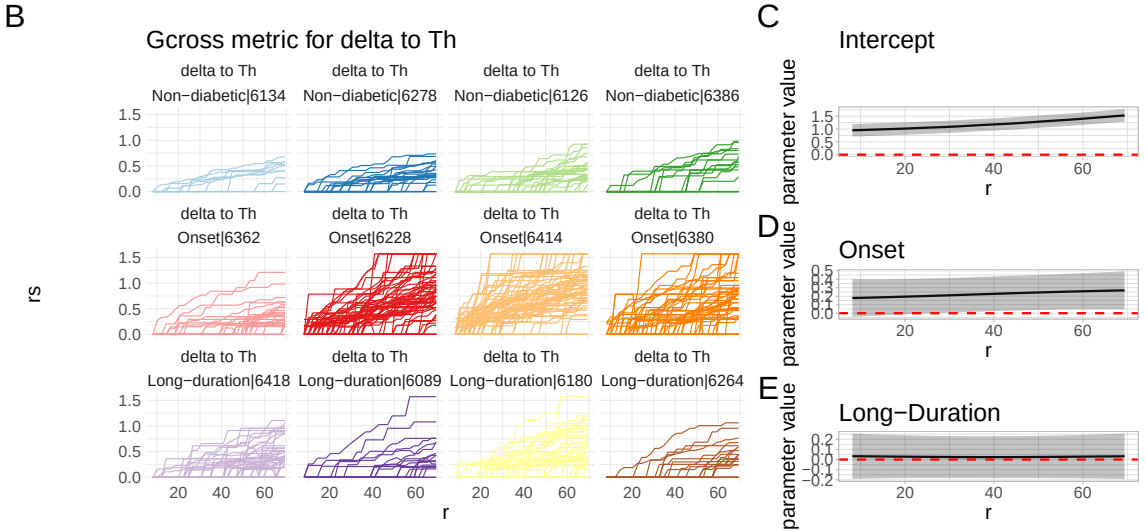
— Autocorrelation: AR(1) error model - residuals are correlated

Imaging mass cytometry dataset of type-I diabetes [Damond et al., 2019].

Pairwise differential co-localisation in type-I diabetes



Individual functional tests provide scale-effects



Conclusions

— Lattice data analysis compares relative locations of numeric marks on a grid → observation-based view

— Point pattern analysis is for inference of the stochastic process underlying the pattern → event-based view

— `spatialFDA` compares spatial statistics functions across spatial scales

References I

10X (2022).
Mouse brain section (coronal).
https://www.10xgenomics.com/datasets/mouse-brain-section-coronal-1-standard.
Accessed: 2024-08-15.

Baddeley, A., Rubak, E., and Turner, R. (2015).
Spatial point patterns: methodology and applications with R.
CRC press.

Canete, N. P., Iyengar, S. S., Ormerod, J. T., Baharlou, H., Harman, A. N., and Patrick, E. (2022).
spicyr: spatial analysis of in situ cytometry data in r.
Bioinformatics, 38(11):3099–3105.

Damond, N., Engler, S., Zanotelli, V. R. T., Schapiro, D., Wasserfall, C. H., Kusmartseva, I., Nick, H. S., Thorel, F., Herrera, P. L., Atkinson, M. A., and Bodenmiller, B. (2019).
A Map of Human Type 1 Diabetes Progression by Imaging Mass Cytometry.
Cell Metabolism, 29(3):755–768.e5.

References II

Emons, M., Gunz, S., Crowell, H. L., Mallona, I., Kuehl, M., Furrer, R., and Robinson, M. D. (2025).
Harnessing the potential of spatial statistics for spatial omics data with pasta.
Nucleic Acids Research, 53(17):gkaf870.

Garrido-Trigo, A., Corraliza, A. M., Veny, M., Dotti, I., Melón-Ardanaz, E., Rill, A., Crowell, H. L., Corbí, Á., Gudiño, V., Esteller, M., Álvarez-Teubel, I., Aguilar, D., Masamunt, M. C., Killingbeck, E., Kim, Y., Leon, M., Visvanathan, S., Marchese, D., Caratù, G., Martin-Cardona, A., Esteve, M., Ordás, I., Panés, J., Ricart, E., Mereu, E., Heyn, H., and Salas, A. (2023).
Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease.
Nature Communications, 14(1):4506.

Hawinkel, S., Yang, X., Poelmans, W., Motte, H., Beeckman, T., and Maere, S. (2025).
Unified nonparametric analysis of single-molecule spatial omics data using probabilistic indices.
bioRxiv.

He, S., Bhatt, R., Brown, C., Brown, E. A., Buhr, D. L., Chantranuvatana, K., Danaher, P., Dunaway, D., Garrison, R. G., Geiss, G., Gregory, M. T., Hoang, M. L., Khafizov, R., Killingbeck, E. E., Kim, D., Kim, T. K., Kim, Y., Klock, A., Korukonda, M., Kutchma, A., Lewis, Z. R., Liang, Y., Nelson, J. S., Ong, G. T., Perillo, E. P., Phan, J. C., Phan-Everson, T., Piazza, E., Rane, T., Reitz, Z., Rhodes, M., Rosenbloom, A., Ross, D., Sato, H., Wardhani, A. W., Williams-Wietzikoski, C. A., Wu, L., and Beechem, J. M. (2022).
High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging.
Nature Biotechnology, 40(12):1794–1806.

References III

Janesick, A., Shelansky, R., Gottscho, A. D., Wagner, F., Williams, S. R., Rouault, M., Beliakoff, G., Morrison, C. A., Oliveira, M. F., Sicherman, J. T., et al. (2023).
High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis.
Nature Communications, 14(1):8353.

Moffitt, J. R., Lundberg, E., and Heyn, H. (2022).
The emerging landscape of spatial profiling technologies.
Nature Reviews Genetics.

Moses, L., Einarsson, P. H., Jackson, K., Luebbert, L., Booeshaghi, A. S., Antonsson, S., Bray, N., Melsted, P., and Pachter, L. (2023).
Voyager: exploratory single-cell genomics data analysis with geospatial statistics.
bioRxiv.

Pebesma, E. and Bivand, R. (2023).
Spatial data science: With applications in R.
Chapman and Hall/CRC.

References IV


Rao, A., Barkley, D., França, G. S., and Yanai, I. (2021).
Exploring tissue architecture using spatial transcriptomics.
Nature, 596(7871):211–220.


Santangelo, J. S. (2018).
Linear mixed-effects models.
`https://uoftcoders.github.io/rcourse/lec08-linear-mixed-effects-models.html`.
Accessed: 2025-12-8.


Scheipl, F., Gertheiss, J., and Greven, S. (2016).
Generalized functional additive mixed models.
Electronic Journal of Statistics, 10(1):1455–1492.


Scheipl, F., Staicu, A.-M., and Greven, S. (2015).
Functional additive mixed models.
Journal of Computational and Graphical Statistics, 24(2):477–501.

References V

Seal, S., Neelon, B., Angel, P. M., O'Quinn, E. C., Hill, E., Vu, T., Ghosh, D., Mehta, A. S., Wallace, K., and Alekseyenko, A. V. (2024).
Spaceanova: Spatial co-occurrence analysis of cell types in multiplex imaging data using point process and functional anova.
Journal of Proteome Research, 23(4):1131–1143.

Shi, H., He, Y., Zhou, Y., Huang, J., Maher, K., Wang, B., Tang, Z., Luo, S., Tan, P., Wu, M., et al. (2023).
Spatial atlas of the mouse central nervous system at molecular resolution.
Nature, 622(7983):552–561.

Singhal, V., Chou, N., Lee, J., Yue, Y., Liu, J., Chock, W. K., Lin, L., Chang, Y.-C., Teo, E. M. L., Aow, J., et al. (2024).
Banksy unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis.
Nature Genetics, 56(3):431–441.