



# Finding structure in gene expression

As transcriptomics methodologies have matured, a major goal is to characterize tissues by identifying constituent cell types and analysing the spatial arrangements of cell types and gene expression patterns within a tissue. However, existing analytical tools for spatial gene expression data typically do not fully leverage the spatial information, instead performing spatially naive gene expression analyses on the single cells or tissue regions being sampled, before mapping the results back onto the spatial structure. Two new studies report bioinformatics analysis tools to directly identify genes with spatially structured expression patterns in tissues.

Both teams adapted statistical methodologies from physical science fields. Svensson, Teichmann and Stegle adapted Gaussian process regression from geostatistics to derive their SpatialDE method, in which gene expression variability is decomposed into spatial and non-spatial components. Gene expression variability that is sufficiently explained by the spatial component (the pairwise distance between cells or tissue regions) is used to call genes as spatially variable. In a related but distinct approach, Edsgård, Johnsson and Sandberg used marked point processes from geostatistics, astronomy and materials physics, in which points are used to represent the spatial locations of cells or tissue regions, and marks on these points represent expression levels. Their method, named trendsceek, tests points in a pairwise manner to identify when the expression pattern of a gene is dependent on the distance between the points being analysed. For both methods, data simulations were used to refine the algorithms

and to demonstrate that they could robustly identify different spatial patterns of gene expression.

The teams then applied their methods to available spatial gene expression data sets. Current data sets are largely generated by either RNA sequencing (RNA-seq) or single-molecule fluorescence in situ hybridization (smFISH), each with different complementary trade-offs.

Spatial RNA-seq-based methods provide expression data transcriptome-wide, but spatial resolution is typically limited to tissue regions encompassing ~10–100 cells for each sample. Both groups applied their analytical tools to the same mouse olfactory bulb RNA-seq data set, identifying 67 (for SpatialDE) and 35 (for trendsceek) genes with spatially structured expression patterns. The relevance of various identified genes was validated through agreement with their known spatial structure from standard single-gene histological staining, or because they are known marker genes for different components of tissue structure. A noteworthy extension of SpatialDE is ‘automatic expression histology’ (AEH), whereby following the identification of individual genes with spatially structured expression, genes are grouped into sets with similar spatial partners, thus yielding insights into tissue histology. As an example, in the olfactory bulb, there were five distinct types of spatial expression pattern, each shared by between 5 and 27 genes. For trendsceek, an automated procedure identifies the cells or tissue regions contributing to the upregulated expression patterns for each identified gene.

Both methods were also applied to RNA-seq data sets from breast cancer tissue, identifying 115 (for SpatialDE) and 14 (for trendsceek)

spatially structured genes. Again, the identities of various genes were consistent with relevant tissue substructure: both methods identified extracellular matrix components as likely architectural substructure, and SpatialDE identified cytokines and interleukin receptors potentially representing structured regions of immune infiltration.

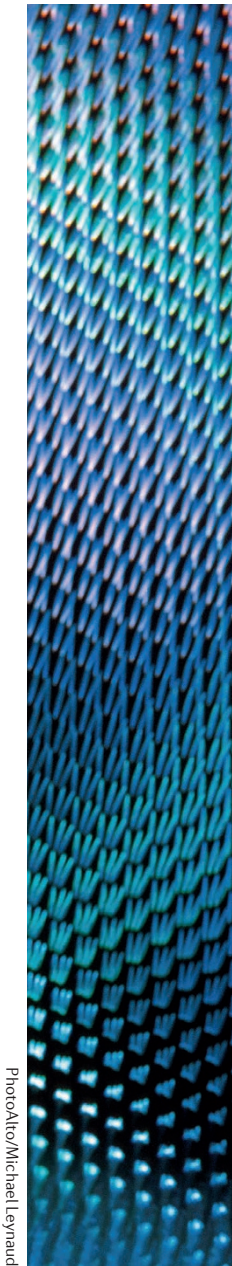
For smFISH data sets, as transcripts are imaged in situ, single-cell resolution (and even subcellular resolution) is available, but label multiplexing typically limits analyses to hundreds of transcripts rather than transcriptome-wide. Both teams applied their methods to smFISH data of 249 genes in mouse hippocampus sections. SpatialDE identified 32 spatially structured genes, whereas trendsceek found a median of 54 spatially structured genes across 15 hippocampal regions.

Overall, SpatialDE and trendsceek can thus be applied to distinct types of spatial gene expression data. It will be interesting to apply them to additional tissue types, including side-by-side comparisons of the tools to examine reasons for, and any implications of, differences in numbers and identities of the genes identified. Finally, the authors note that the tools are extendible to additional data types, including spatial expression data in 3D or through time series.

Darren J. Burgess

**ORIGINAL ARTICLES** Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* <https://doi.org/10.1038/nmeth.4636> (2018) | Edsgård, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. *Nat. Methods* <https://doi.org/10.1038/nmeth.4634> (2018)

**FURTHER READING** Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015)



PhotoAlto/Michael Leynaud

“ adapted statistical methodologies from physical science fields ”