#2331 InSituCor: Uncovering gene co-regulation networks in spatial transcriptomics data

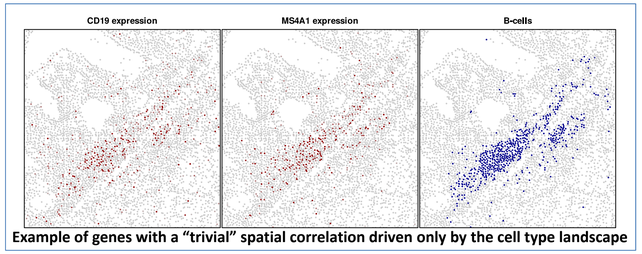
Patrick Danaher, Lidan Wu, Dan McGuire, Michael Patrick, David Kroeppler, Haiyan Zhai, Joachim Schmid, Joseph M. Beechem

nanoString Technologies
530 Fairview Avenue North, Seattle, WA 98109

Abstract

Spatially correlated gene sets reveal intricate and interesting biology: When several genes tend to be expressed in the same tissue regions, it suggests causality, either through gene-gene interactions or some latent variable influencing all the genes. Exploring spatial correlations can be a productive hypothesis-generating exercise.

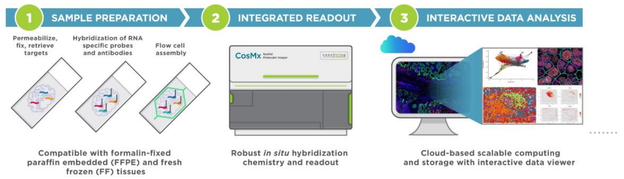
Most spatial correlations are trivial: Tissues' cell type landscapes induce spatial correlation between genes expressed by the same cell type. E.g. CD19 and MS4A1 (CD20) will be spatially correlated by virtue of being B-cell markers. Results like this crowd out more interesting findings.

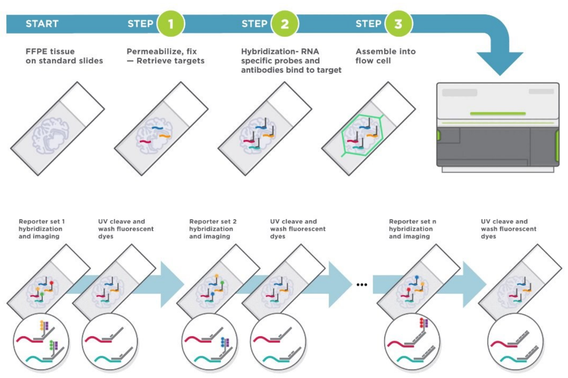


Example of genes with a "trivial" spatial correlation driven only by the cell type landscape

The InSituCor algorithm discards trivial correlations, exposing gene sets worth attention: InSituCor measures spatial correlation unexplained by the cell type landscape. Its findings are thus depleted of uninteresting marker gene results and enriched with causal biology.

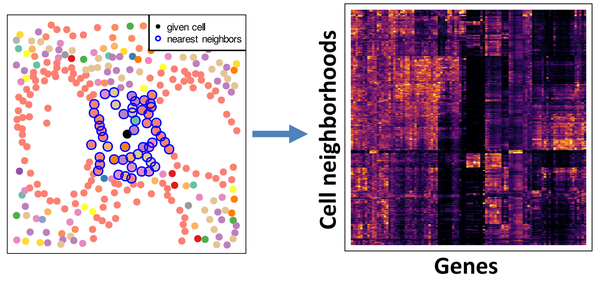
Methods: Single-cell imaging performed with the CosMx™ Spatial Molecular Imager (SMI)





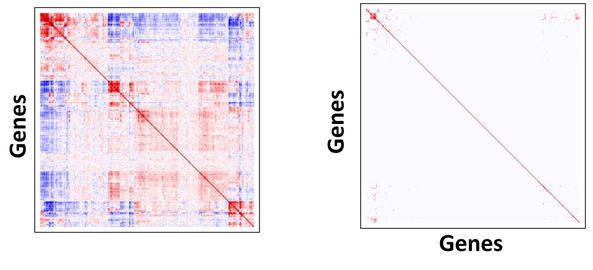
InSituCor algorithm: Isolating non-trivial spatial correlations

Step 1: record the gene expression profile around each cell's neighborhood.



Naïve spatial correlations: take gene-gene correlations in the neighborhood expression matrix.

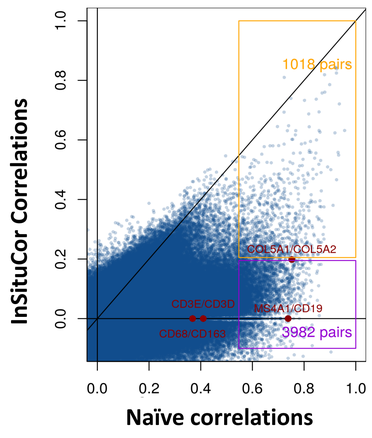
InSituCor: take gene-gene correlations conditional on cell type.



InSituCor yields a more selective & interesting set of correlations.

Right: gene-gene correlations under Naïve and InSituCor approaches.

Key results:
- InSituCor discards ~80% of naïve correlations.
- Correlations between marker genes are successfully negated.



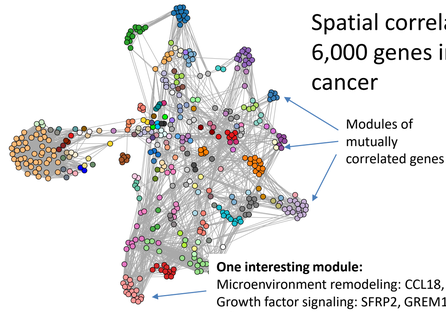
Resources

R package:
https://github.com/Nanostring-Biostats/InSituCor

Preprint:
https://www.biorxiv.org/content/10.1101/2023.09.19.558514v1

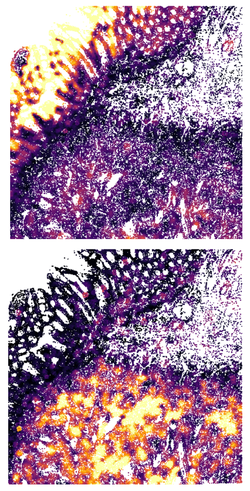
Workflow: spatial correlations as a vehicle for hypothesis generation

Step 1: cluster correlated genes into modules

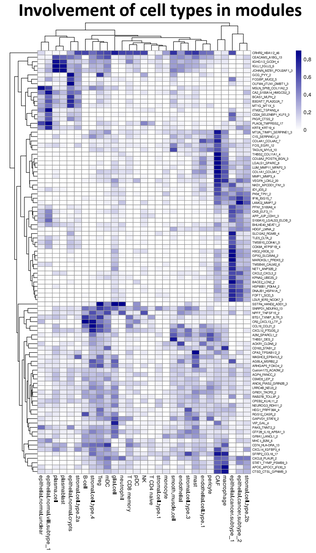


Step 2: use gene content, spatial trends and cell type involvement to prioritize interesting modules.

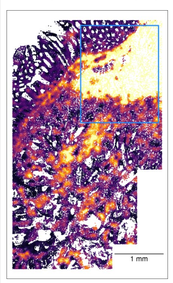
Activity levels of selected modules



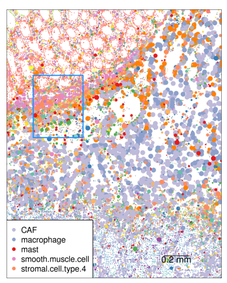
Involvement of cell types in modules



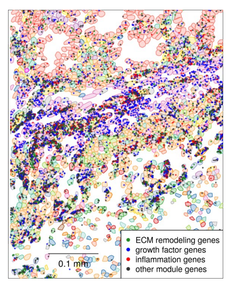
Step 3: Deeply explore the modules of greatest interest.



Activity of microenvironment remodeling module in cellular neighborhoods



Activity of microenvironment remodeling module in single cells



Transcripts from microenvironment remodeling module genes

Screen for pan-cancer spatially correlated gene modules

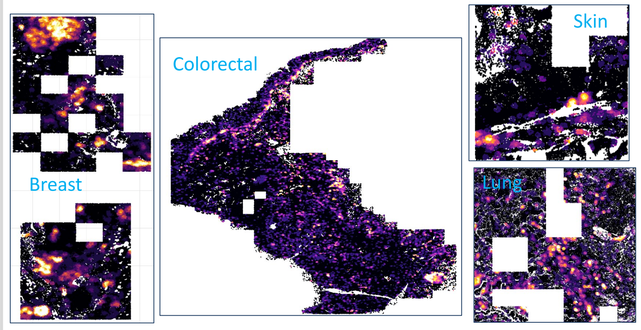
Motivation: by isolating spatial correlations that appear consistently across multiple cancer types, can we discover fundamental behaviors of the tumor microenvironment?

Approach: Run InSituCor separately for cancer samples run with the CosMx Human 6k Discovery Panel: skin, colon, kidney, breast, lung. Search for consistent spatial correlations.

Results: below: selected pan-cancer modules

Primary cell types	Top markers	functions
Mostly stromal & smooth muscle cells	MYL2-OGN-2	Myosin & osteoglycin; cytoskeleton & ECM components
	NR2F2-SYNPO2-2	Marker for clus 3 & 7; Smooth-muscle-cell hormone receptor/transcription factor & regulate actinin/filamin binding activity (cytoskeleton)
	ELP6-FLNC-2	tRNA processing & muscle-specific filamin (cytoskeleton)
	FGL2-KCNMA1-2	Mucosal secreted protein & smooth-muscle-cell Ca-activated BK channel.
	MYL9-MYH11-30	Giant cluster of 30 genes, 3 myosin, 3 cytoskeleton (actin), 2 muscle related cytoskeleton binding proteins
	CELSR2-CER1-2	Cadherin (CELSR2) & cytokine (CER1, BMP antagonist)
	ERCC8-ADRA2B-68	Giant cluster of 68 genes; various functions, including housekeeping activities like actin (ADRA2B), ubiquitin ligase (ERCC8), etc.
	NDRG4-PLPPR1-2	Heart muscle (NDRG4) & oligodendrocytes (PLPPR1)
	LAMB4-DGKB-SLC2A13-3	3 genes, laminin and 2 metabolic proteins
	NR2F2-SYNPO2-2	Marker for clus 3 & 7; Smooth-muscle-cell hormone receptor/transcription factor & regulate actinin/filamin binding activity (cytoskeleton)
CAF, endothelial	IL12A-LIFR-2	Cytokine for NK/T cells (IL12A) & adipose-enriched Cytokine receptor (LIFR)
	DOCK9-EDN1-2	Regulate cadherin binding activity & endothelial cell-secreted peptide
	NAPSA-LRRK2-6	6 genes, diverse functions: protein processing (NAPSA, LRRK2), water transport (AQP4), lipid metabolism (LPCAT1) , neuronal (NRGN) & immune (SFTPD) signaling modulation.
	MFAP4-TIMP3-RARRES2-3	3 genes enriched in adipose tissues, ECM binding protein & proteinase inhibitor, adipocyte secreted chemotactic protein (RARRES2)
	RAMP3-VWF-9	9 genes, blood/plasma/vascular related
	COL4A4-GREM1-2	Collagen & smooth muscle-enriched cytokine (BMP antagonist)
	ADAM12-CTSK-2	Marker for clus 10 & 15; ECM proteinases in bone & skeletal muscle regeneration
	EMP1-SPON2-2	Epithelial receptor & Smooth-muscle-cell adhesion protein
	MMP3-CXCL5-5	5 genes; 2 ECM proteinase, 3 immune regulation (chemokine, cytokine, hematopoietic growth factor)
	COL3A1-COL1A2-50	Giant cluster of 50 genes, markers for both cluster 10 & 15; 12 collagen & fibronectin, 3 complement, 6 ECM proteinase or inhibitors, more ECM components (glycan).
Fibroblast, Macrophage, Malignancy	ADAM12-CTSK-2	Marker for clus 10 & 15; ECM proteinases in bone & skeletal muscle regeneration
	ETS1-NOX4-2	NK/T-cell transcription factor & NADPH oxidase
	TPSAB1/2-CTSG-5	5 genes; metabolic enzymes, plasma proteins, markers for mast cells (TPSAB1/2) & neutrophils (CTSG)
	AGRP-AP2S1-IDS-3	3 genes; neuropeptide (AGRP), endocytosis (AP2S1), lysosomal enzyme for glycans (IDS)
	COL3A1-COL1A2-50	Giant cluster of 50 genes; ECM components & remodeling, some immune
Malignancy, Fibroblast, Endothelial	P3H2-PDE10A-2	Hydroxylase enzyme involved in collagen assembly (P3H2) & cAMP/cGMP hydrolysis (PDE10A)
	OSMR-VCAM1-2	Cytokine receptor (OSMR) & vascular adhesion (VCAM1)
	ACAT1-PLIN2-2	Energy metabolism & lipid storage.
	LRRC41-LOX-9	9 genes, house keeping activities like ubiquitin ligase (LRRC41), metabolism (LOX, ABCC3/4)
	COL4A1-COL4A2-17	Giant cluster of 17 genes, 3 collagen & laminin, 2 myosin & integrin, 3+ endothelial markers (CD34, KDR, PODXL)

Spatial distribution of a chemokine module (CCL19, CCL21)



Scan here to download or learn more