Updating immune cell deconvolution for the spatial genomics era

NanoString Technologies 530 Fairview Avenue North, Seattle, WA 98109

Summary

A spatially-resolved gene expression dataset cannot be fully understood without a census of the cell types present in each profiled region. Here we introduce the NanoString Quantitative Single Cell Deconvolution (qSCD) algorithm for using gene expression data to quantify mixed cell types. We validate the algorithm in tumor tissues, and we demonstrate its use in a NSCLC tumor profiled with the GeoMx® Cancer Transcriptome Atlas RNA panel.

GeoMx DSP reagents and workflow

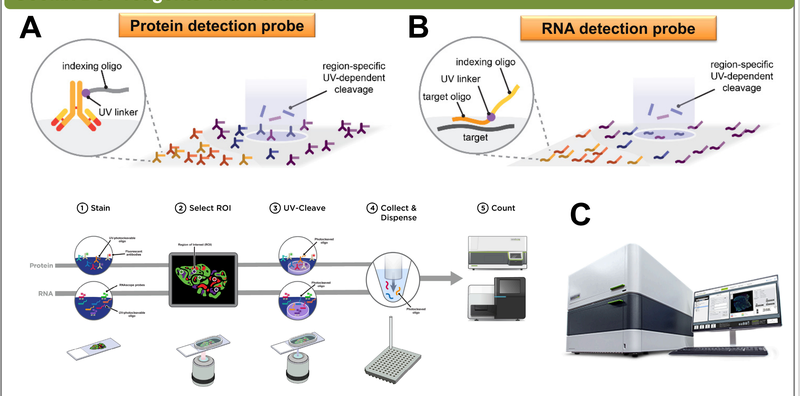
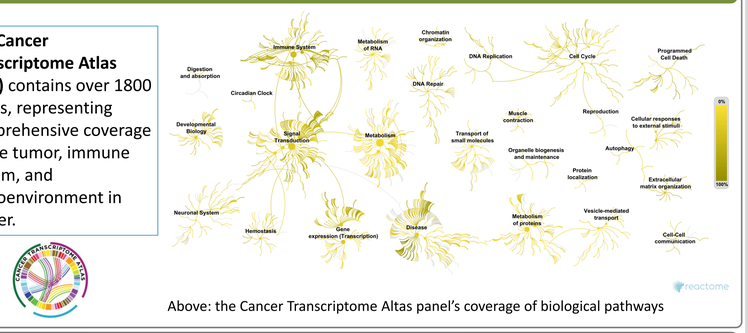


Fig 1. The GeoMx Digital Spatial Profiler (GeoMx DSP) is based on barcoding technology that enables spatially resolved, digital characterization of proteins or RNA in a highly multiplexed (over 2,000-plex) assay. The oligonucleotide tags cleaved from discrete regions are quantitated by NGS or nCounter, and counts are mapped back to tissue location, yielding a spatially-resolved digital profile of analyte abundance. Using a UV-cleavable linker, epitope-specific antibodies (A) or in-situ-hybridization probes (B) are conjugated with unique DNA-oligo tags. GeoMx DSP shapes and illuminates UV lights over user-defined tissue subregions of interest only to cleave & collect DNA-oligo tags and records xy coordinates of the subregions (C). Cleaved tags from each ROI are collected and counted using nCounter or a NGS sequencer.

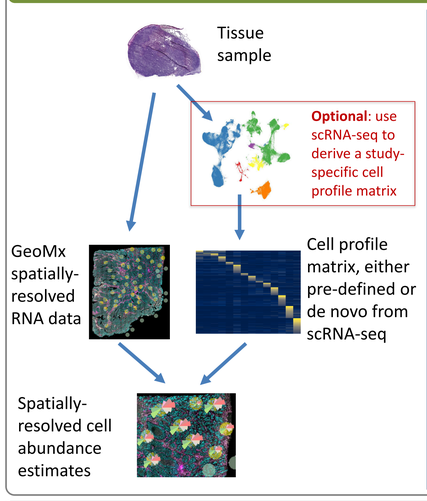
Spatial RNA profiling with the Cancer Transcriptome Atlas

The **Cancer Transcriptome Atlas (CTA)** contains over 1800 genes, representing comprehensive coverage of the tumor, immune system, and microenvironment in cancer.



Above: the Cancer Transcriptome Altas panel's coverage of biological pathways

Quantitative Single Cell Deconvolution (qSCD) algorithm



Algorithm details

Key development: log-linear loss function

- Many deconvolution methods (NNLS, v-SVR) are extensions of linear regression, which can be written as an optimization problem:

$$\hat{\beta} = argmin_{\beta} \|y - X\beta\|_2$$

where **y** = observed gene expression,
X = immune cell profile matrix,
β = cell abundances

- But error in gene expression is best modelled on the log-scale. We therefore adapt the above to minimize the squared log-scale residuals:

$$\hat{\beta} = argmin_{\beta} \|log(y) - log(X\beta)\|_2$$

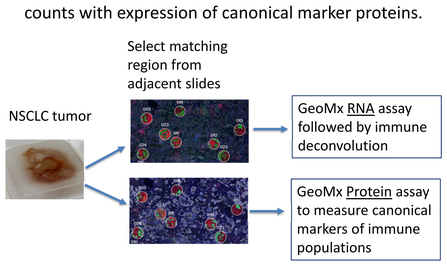
- We also use a non-negativity constraint on β – there can't be less than zero of a cell type.

Other algorithm features

- We compute a p-value for each cell type, enabling "detected" vs. "undetected" calls.
- The GeoMx platform counts the nuclei in each region profiled. This lets us estimate the number of each cell type, not just the proportion.
- We explicitly model background counts to enable estimation in small, low-signal regions.

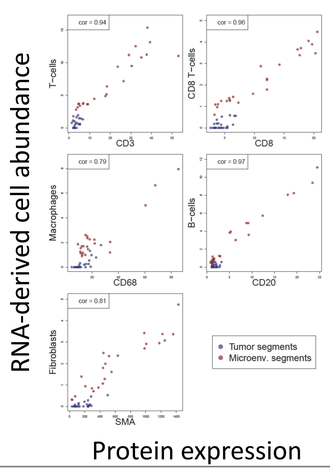
Validation in situ of RNA deconvolution algorithm vs. protein expression

Approach: Gather GeoMx RNA and protein data from consecutive slides from the same tumor. In corresponding regions, compare RNA-based cell counts with expression of canonical marker proteins.

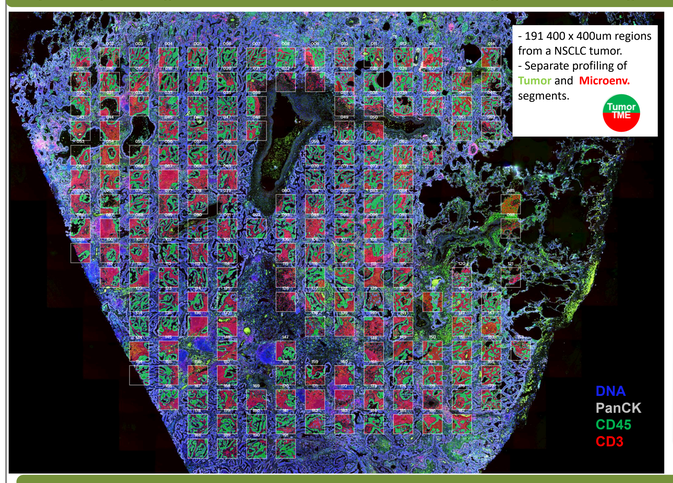


Protein panel

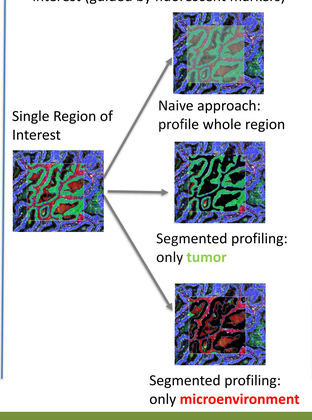
Core Cell Profiling		Immune Cell Typing		Immune Activation Status	
Rb IgG	CD3	CD45RO	FAPalpha	CD127	PD-L2
Ms IgG1	CD4	FOXP3	CD14	CD25	CD40
Ms IgG2a	CD45	CD34	CD163	CD80	CD44
PD-1	CD56	CD66b		ICOS	CD27
Histone H3	CD8	IO Drug Target		Pan-Tumor	
CD68	CTLA4	4-1BB	ARG1	MART1	Her2/ErbB2
HLA-DR	GZMB	LAG3	B7-H3	NY-ESO-1	PTEN
S6	GAPDH	OX40L	IDO1	S100B	ER alpha
Ki-67	PD-L1	Tim-3	STING	Bcl-2	PR
B2M	PanCK	VISTA	GITR		
CD11c	SMA				
CD20	Fibronectin				



Separate profiling of Tumor and Microenvironment segments from regions of a NSCLC tumor

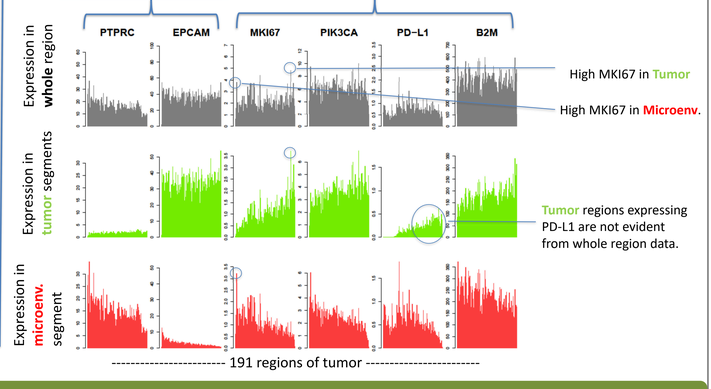


Segmented profiling: measure gene expression only within cell types of interest (guided by fluorescent markers)



Confirming accuracy: Tumor/ Immune markers are highly specific to appropriate segments

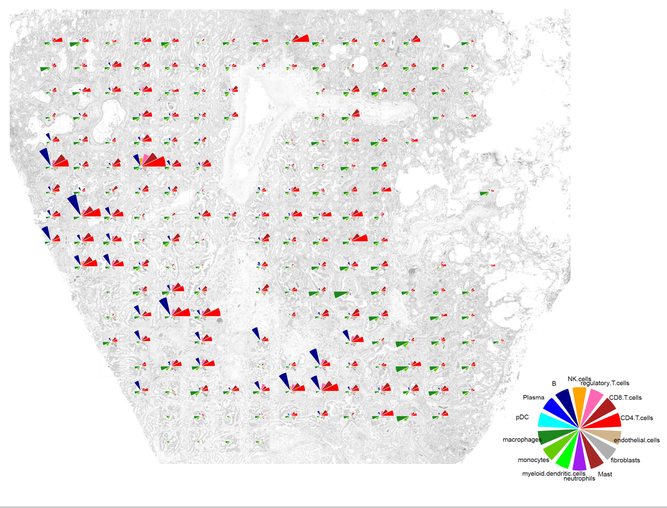
Importance of segmented profiling: Canonical driver and immuno-oncology genes are expressed by both Tumor and Microenvironment. Only with segmented profiling can we attribute their expression either cell type.



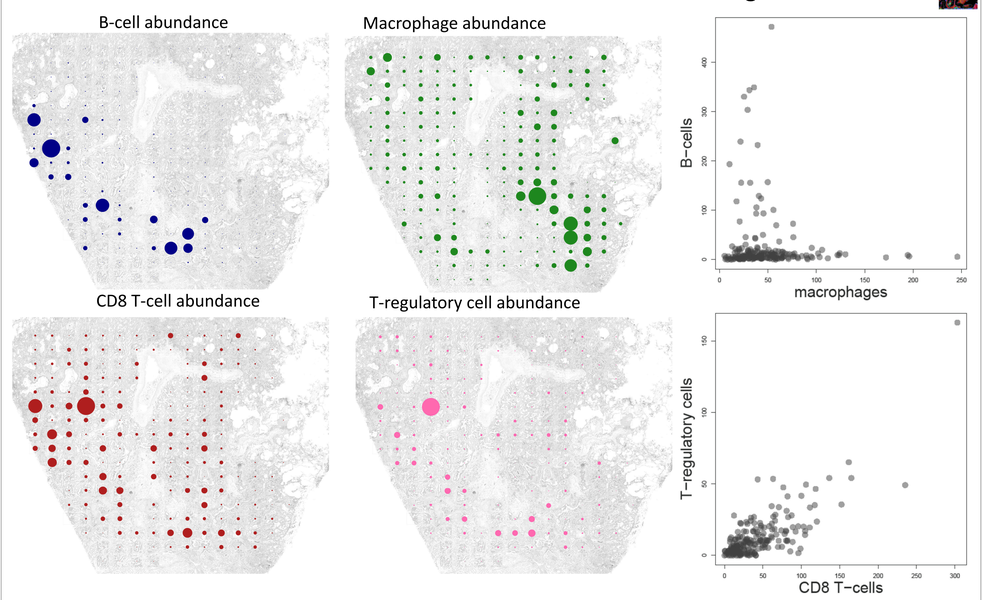
Spatially-resolved immune cell deconvolution in microenvironment segments

Results: Abundance of immune cells in Microenvironment

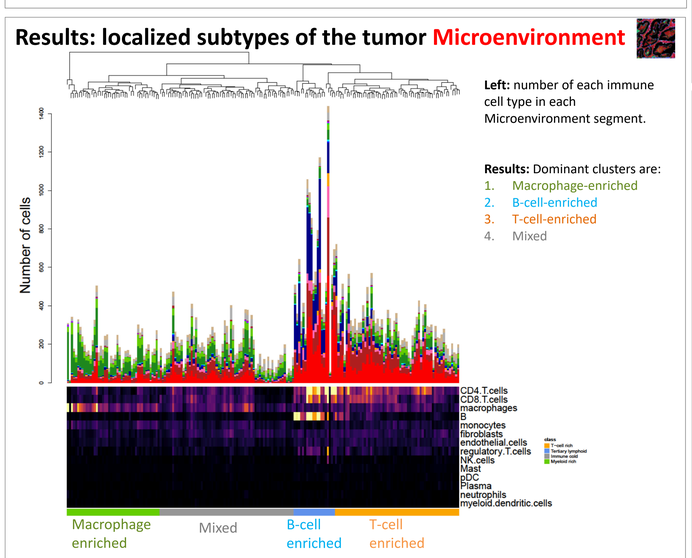
Below: Estimated counts of 14 cell types in microenvironment segments from 191 regions.



Results: How do immune cells co-localize across Microenvironment segments?



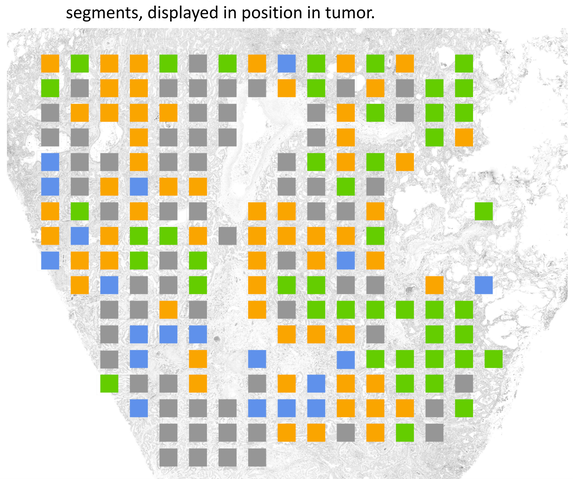
Results: localized subtypes of the tumor Microenvironment



Left: number of each immune cell type in each Microenvironment segment.

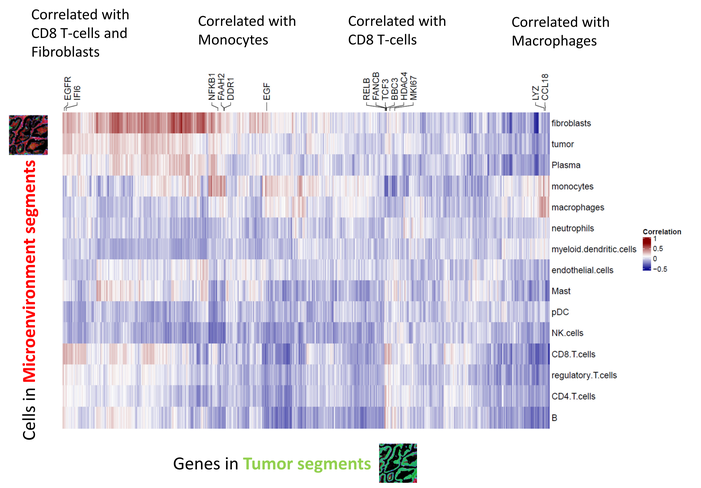
Results: Dominant clusters are:
1. Macrophage-enriched
2. B-cell-enriched
3. T-cell-enriched
4. Mixed

Below: cluster assignments for Microenvironment segments, displayed in position in tumor.

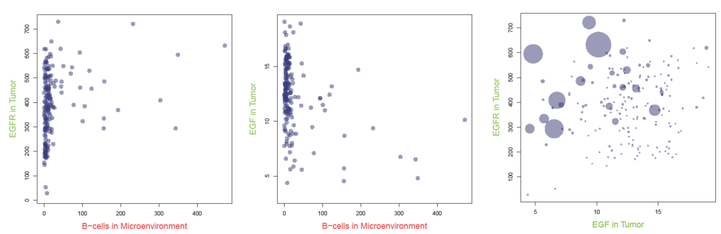


Results: How does Tumor gene expression interact with the neighboring Microenvironment?

Correlation between immune cells in Microenvironment and genes in Adjacent Tumor.



Relationship between B-cells in Microenvironment and selected genes in Adjacent Tumor



Conclusion

- **In GeoMx RNA datasets, the NanoString Quantitative Single Cell Deconvolution (qSCD) algorithm accurately quantifies immune cell populations in minute regions of FFPE tumors.**
- **Cell mixture deconvolution enables new insights from spatially-resolved RNA data.**

Researchers interested in participating in NanoString's technology access program for its Digital Spatial Profiling technology should contact us at TAP@nanostring.com.

Patrick Danaher, Sarah Church, Chris Merritt, Dan Zollinger, Jason Reeves, Youngmi Kim, Joseph Beechem*
NanoString Technologies, Inc.
*Correspondence: jbeechem@nanostring.com

AGBT, 02-23-2020 ~02-26-2020