#638 Spatial Single-Cell Whole Transcriptome Imaging in Tissue and Fixed Cells using CosMx Spatial Molecular Imaging

Joseph Beechem, Rustem Khafizov, Tushar Rane, Evelyn Metzger, Claire Williams, Hiromi Sato, Michael Rhodes, Ashley Heck, Rachel Liu, Kim Young, Max Walter, Ryan Garrison, John Hamann, Andrew Grootsky, Joseph Phan, Erin Piazza, Margaret Hoang, Dwayne Dunaway

NanoString Technologies
530 Fairview Avenue North, Seattle, WA 98109

Abstract

Abstract: The CosMx™ Human Whole Transcriptome Imaging Panel (WTx) has been developed at 18,985-plex and covers the complete protein coding transcriptome in humans. Detailed titration studies revealed that high-sensitivity whole transcriptome imaging can be obtained using 2-hybridization tiles per protein coding gene, resulting in a single-cell (and subcellular) hybridization panel comprising over 37,990 imaging barcodes. The imaging barcodes are encoded at 156-bits (4 on-cycles and 35 dark-cycles per code) and each imaging probe is Hamming Distance = 4 from each other, resulting in a very low false-code detection. Advancing single-cell imaging to the whole transcriptome level opens up a single unified approach to accomplishing essentially all single-cell transcriptome experimentation (both imaging and non-imaging). Depending upon the sample type investigated (from fixed-cells to entire tissue sections), the transcripts-per-cell and genes-per-cell of SMI whole transcriptome imaging approaches that obtained by even the highest-resolution single-cell RNA-seq, with maximums of up to 10,000 transcripts per cell and over 3,000 genes per cell fully imaged. In order to better segment the entire transcriptome into single cells, SMI has been extended to allow high-plex protein imaging (up to ~75-plex) on exactly the same slide that was whole transcriptome imaged. This allows advanced data-driven cell-segmentation software to be developed using high-resolution images of the plasma membrane, nuclear compartment, and cytoplasm. In this manner, the whole transcriptome can be accurately placed into the single cells (of origin). Data will be presented spanning the entire range of single-cell experimentation, from cells-in-culture to dissociated tissue samples to multiple 1X1 cm² FFPE tissue sections. Tertiary analysis of whole transcriptome data will include cell typing, ligand-receptor detection, pathway enumeration, and neighborhood analysis. The high-dimensional whole transcriptome plus multi-omic protein data is streamed directly to the cloud-based AtoMx™ Spatial Informatics Platform.

Introduction to CosMx platform

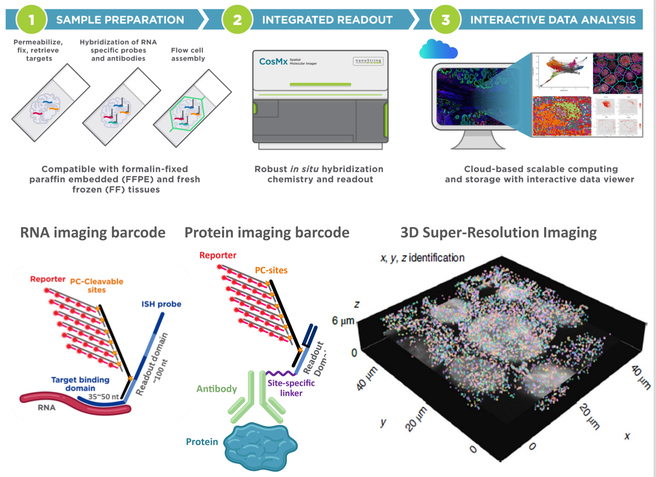


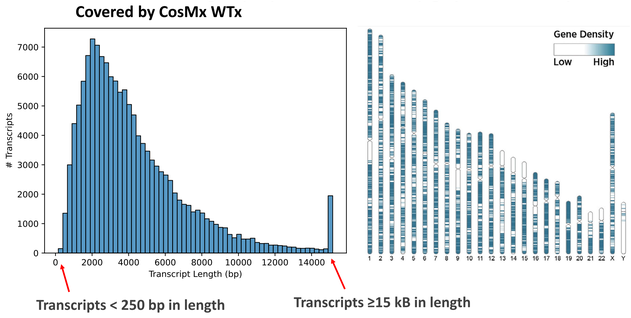
Figure 1: Description of CosMx Technology. Detection of RNA relies on barcode readout on the SMI instrument via several rounds of reporter binding and fluorescence imaging utilizing universal SMI readout reagents. The analysis workflow involves 3D primary image processing to identify and register reporter spots and decoding of reporter spots to RNA transcripts with registered X, Y, Z spatial locations.

CosMx Chemistry: how we image the whole transcriptome

Hybridization Probe Design:

We designed two probes to each protein-coding gene in the human transcriptome as defined by the HUGO Gene Nomenclature Committee (HGNC) and with available sequence in NCBI RefSeq. The RefSeq transcriptome used was based on the T2T genome build, but all probes were validated to also hit hg38-based transcripts for each gene. Probes were designed by our Gauntlet 2 (G2) probe design architecture and considered thermodynamic and structural parameters, probe-probe binding and homology, specificity to the intended gene, and coverage of known transcripts. Due to sequence similarity, some targets hit multiple genes, and fully redundant targets were removed from the design. Further, we removed all mitochondrially-encoded targets and the top 10 most highly expressed genes as measured by the Cancer Genome Atlas (ACTB, ACTG1, EEF11A1, EEF2, FTL, GAPDH, PSAP, RPL3, TPT1, UBC; 50 negative control targets also included). The only genes excluded from design beyond that were those without sequence available in RefSeq, none failed on the basis of sequence composition or length.

Distribution of Lengths (left) and Chromosomes (right) for All Transcripts Covered by CosMx WTx



Imaging Barcode Design:

The barcoding scheme is designed to assign a unique barcode to each target transcript from a set of 156-bit barcodes (four color reporters in each readout round over 39 readout rounds), with Hamming distance 4 (HD4) and Hamming weight 4 (HW4). Every barcode is separated by an HD of at least four from all other barcodes to maximally suppress RNA decoding error. Every barcode has a constant HW4, in which each target is 'on' in four rounds and 'off' in 35 rounds.

Assignment of barcodes was done algorithmically, with consideration given to the expression of each target across a broad range of scRNA-seq datasets. Assignment prioritized higher expressed targets and aimed at color balance in each chemistry cycle. Lower expressed targets were assigned randomly to the remainder of the barcodes.

Establishing a sensitivity and specificity baseline for CosMx WTx using deep RNA-seq of FFPE cell lines from Cancer Cell Line Encyclopedia (CCLE)

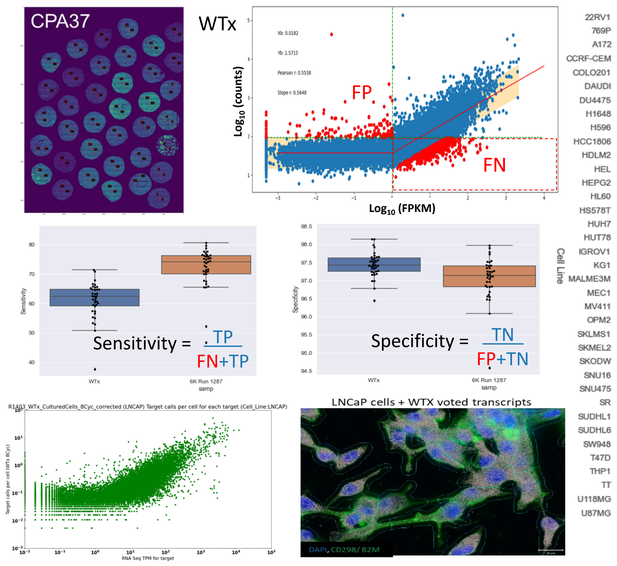


Figure 3: (upper) Example 37-line cell-pellet array (CPA) and CosMx WTx imaging counts (y-axis) vs RNA-seq transcriptome (x-axis). False positives and False negatives in red, and break-point ~sensitivity of CosMx (~1 FPKM). **(middle)** Full cell-line sensitivity and specificity across 37 CCLE cell lines. **(lower)** Example of CosMx WTx directly on fixed-cells in culture (lnCAP) compared with RNA-seq, with over 8,000 transcripts imaged per single cell detected.

Winning the "Space Race": CosMx SMI gives the world its first whole transcriptome, single-cell spatial atlas

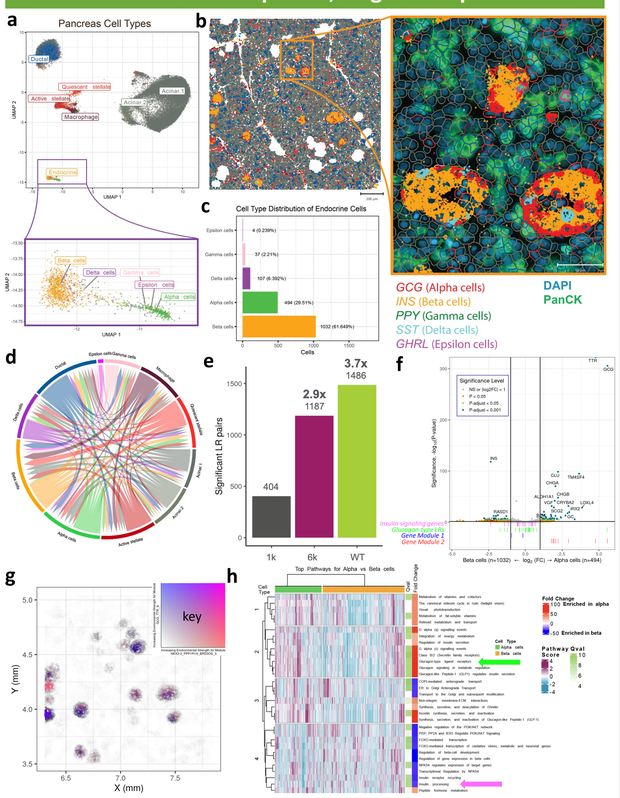


Figure 4: Whole transcriptome spatial analysis with CosMx SMI using the publicly-available Pancreas Data. **(a)** UMAP of expression space showing distinct cell types. **Inset:** magnification of endocrine cell type. **(b)** About 47,000 cells were measured on tissue. Shown here is half (9 FOVs) of the public data (18 FOVs). **Inset:** Magnified image showing spatial details of three islets of Langerhans -- micro-organs that play a key role in regulating blood sugar levels. Cell segmentation boarders denote cell types found in UMAP. **(c)** Cell type proportions among endocrine cells match expected frequencies. **(d)** Ligand-Receptor (LR) interactions between pancreas cells. **(e)** The number of significant LR interactions using CellTalkDB based on the panel content. **(f)** Differential expression volcano plot between alpha and beta cells. Red and blue rugs show genes that were identified in two modules shown in (g). Magenta rug: gene positions for insulin signaling pathway (enriched in beta cells). Green rug: gene positions for Glucagon-type ligand receptors (enriched in alpha cells) **(g)** Spatial gene co-expression analysis with InSituCor identifies the spatial environment around a 3-gene module containing *NKX2.2 and PPP1R1A* (red module) and the 8-gene module containing *GCG, TTR*, and others (blue module). **(h)** Heatmap of top differential enriched pathways between alpha and beta cells. Colored arrows indicate pathways highlighted in panel (f).

Sneak Peak: New Algorithmic Enhancements for CosMx SMI

	Current Decoding	Enhanced Decoding
Mean Transcripts per Cell	**726**	**1214**
Mean Unique Genes per Cell	**339**	**479**
Signal to Noise	**4.235**	**5.79**

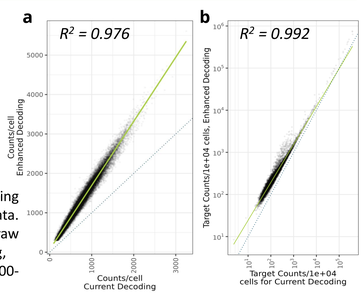


Figure 5: We are continuously improving our algorithms that decode CosMx data. In pancreas WTx, we examined raw transcript counts between our existing, current decoding algorithm (up to 6000-plex) and our new, up to 19000-plex algorithm. The results show higher SNR, greater counts per cell, detection of more unique genes per cell (table). Moreover, cells have about 60% more transcripts (**a;** y = 68 + 1.6x) and genes get an 80% boost in counts (**b;** y = -16 + 1.8x) compared to existing decoding.

Robust best-in-class cell segmentation using both protein and RNA information

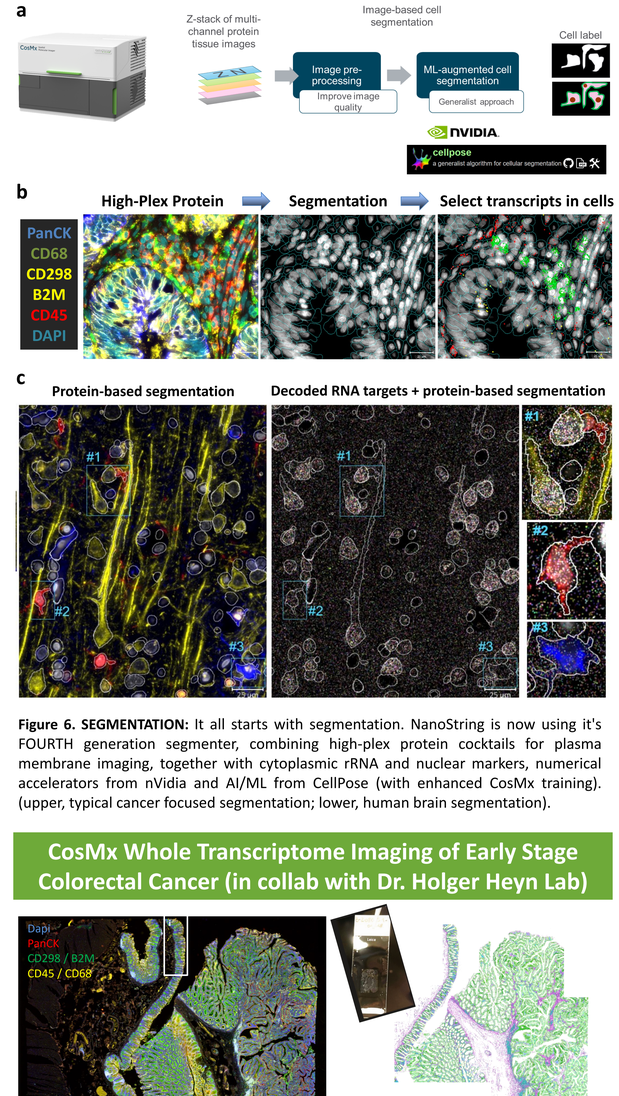


Figure 6. SEGMENTATION: It all starts with segmentation. NanoString is now using it's FOURTH generation segmenter, combining high-plex protein cocktails for plasma membrane imaging, together with cytoplasmic rRNA and nuclear markers, numerical accelerators from nVidia and AI/ML from CellPose (with enhanced CosMx training). (upper, typical cancer focused segmentation; lower, human brain segmentation).

CosMx Whole Transcriptome Imaging of Early Stage Colorectal Cancer (in collab with Dr. Holger Heyn Lab)

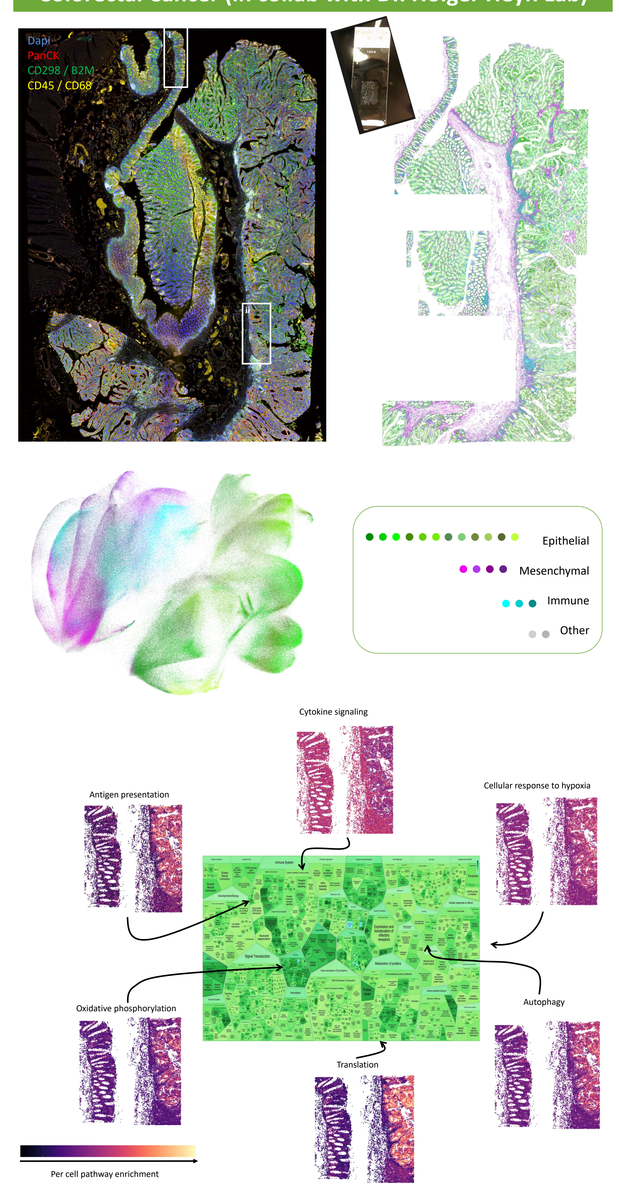


Figure 7: Early colorectal cancer tissue sample (Tubullovillous Adenoma). Total single cells measured = 1.4 Million cells and 1.4 Billion transcripts. (**Upper:** Cell typing spatial map, **Middle:** Various UMAP projections, **Lower:** When imaging ~19,000 genes, you can simply "spatially-map" the entire Reactome onto the tissue morphology. Shown above are 6 (of the over 400) Reactome maps that are directly imaged by CosMx WTx. Say a fond "good-bye" to chasing-down small numbers of marker-genes to infer the biology that is taking place, explore the entire Reactome in completeness.

CosMx Whole Transcriptome Imaging of Early Human Development (in collab with Dr. Bernadette de Bakker *et al*)

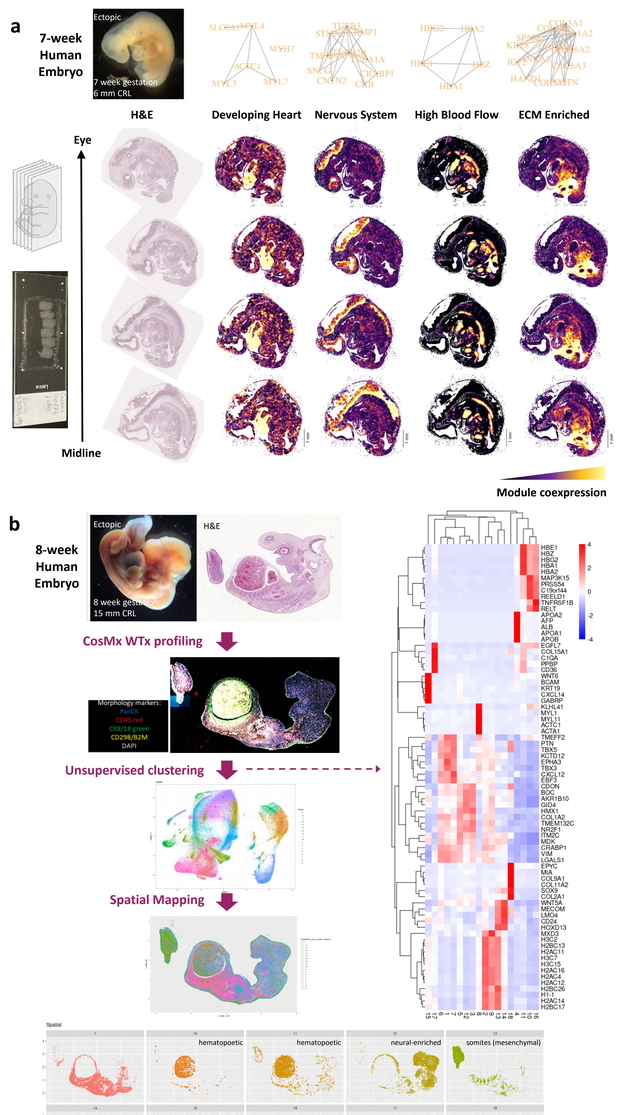


Figure 9: Spatial profiling of ectopic human embryos using CosMx WTx. **(a)** 7-week human embryo dissected from fallopian tube with crown-rump-length (CRL) of 6 mm. Four 5 µm thick sagittal sections were profiled in CosMx. Corresponding H&E serial sections shown with four of 57 developmental biology modules identified with CosMx WTx profiling and InSituCor analysis (Danaher *et al.*, bioRxiv 2023). **(b)** CosMx WTx workflow on one 8-week ectopic human embryo, 5 µm sagittal section. Unsupervised Leiden clustering performed on CosMx WTx resulted in 18 Leiden clusters (shown as UMAP, gene expression heatmap, spatial mapping onto embryo). For example, see skin cluster 15 with epithelial markers, basal cell modulators, fibroblasts, and keratins. Embryo study in collaboration with BS de Bakker, MJB van den Hoff *et al.*, Biobank, Amsterdam UMC. Dutch Fetal.

Conclusions

CosMx Whole Transcriptome Assay (WTx) directly images 37,872 imaging barcodes representing 18,936 protein-coding RNA transcripts.

- Sensitivity ~62% and 97.5% Specificity as measured against Cancer Cell Line Encyclopedia FFPE Cell-Pellet Array. Fixed-cells can yield over 8000 transcripts per cell (Figure 3)
- Example data set using healthy pancreas FFPE sample placed into public domain. Rare cell types down to 0.008% easily quantified, along with over 1,400 Ligand-Receptor interactions and key gene networks between alpha and beta cells (Figure 4)
- Early Detection of Colon Cancer in collaboration with Dr. Holger Heyn generated CosMx WTx data on 1.44 Million Cells detecting over 1.4 Billion Transcripts. Spatial mapping of entire reactome pathways (over 400 of them) accomplished (Figure 7)
- A "fourth-generation" cell-segmenter has been developed, with the ability to follow plasma membranes of single cells for multiple millimeters (Fig 6)
- 3D Whole Transcriptome imaging in a developing 7-week human embryo accomplished (Figure 9)

Scan here to download or learn more