Virtual cell foundation model reveals genotype specific fibroblast-macrophage crosstalk in the non-small cell lung cancer microenvironment

Yubin Xie, Hikari Sorensen, Yu Phoebe Guo, Daniel Millman, Jillian Astarita, Eshed Margalit, Dexter Antonio, Maxime Dhainaut, Lucas Cavalcante, Eric Siefkas, Hargita Kaplan, Rodney Collins, Francis Fernandez, Jake Schmidt, Michela Meister, Abhishaike Mahajan, Joy Tea, Carl Ebeling, Anastasia Mavropoulos, Keith Mitchell, Ryan Huang, Rob Schiemann, Aaron Salm, Dulce Ovando, Harry Winters, Nicole Snell, Shafique Virani, Ron Alfa, Lacey Padron, Emily Corse, Jacob Rinaldi, **Jérémie Decalf** *, Daniel Bear *

* equal contribution I All authors are affiliated to NOETIK Inc., South San Francisco, CA

Abstract

We developed OCTO-vc, a spatial virtual cell foundation model, trained on CosMx spatial transcriptomics across over 40 million cells from 1,399 non-small cell lung cancer (NSCLC) tumor resections. OCTO-vc allows the simulation of gene expression in “virtual cells” placed within spatially resolved TMEs and unlocks the ability to conduct in silico, cell-type specific experiments. To identify local cell-cell interactions between macrophages and tumor cells or fibroblasts, we designed virtual “clonal neighborhood” simulations, in which a virtual macrophage is surrounded by digital replicates of real tumor cells or fibroblasts sampled from NSCLC patients’ TME. We simulated 500 tumor or fibroblast clonal neighborhoods for each patient, leading to inferred gene expression on a total of 873,000 virtual macrophages. This allowed the identification of tumor and fibroblast-mediated transcriptional programs spatially associated with macrophage state, and unveil spatial crosstalk enriched in subsets of NSCLC patients.

1. Generating multimodal data for training a foundation model of human tumor spatial biology

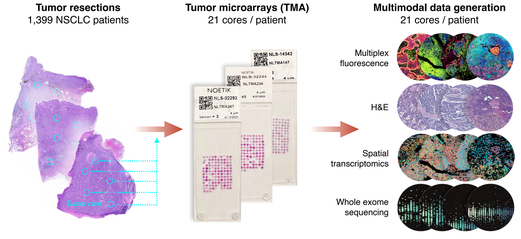


Figure 1. NOETIK’s multimodal data generation. Formalin-fixed, paraffin-embedded (FFPE) tumor blocks obtained from 1,399 non-small cell lung cancer (NSCLC) patients were used to generate Tissue Microarrays (TMAs). Multimodal spatial data were generated using three platforms: 16-channel multiplex Immunofluorescence (mIF), 1000-plex transcriptomics (Cosmx) and hematoxylin-eosin (H&E), in addition to whole exome sequencing. These multimodal data were used to train vision transformer models for learning biological relationships, and perform zero-shot inferences.

For more details on NOETIK’s platform :

Poster by Padron L et al, SITC 2024

2. OCTO-vc: A virtual cell foundation model trained on human tumor data

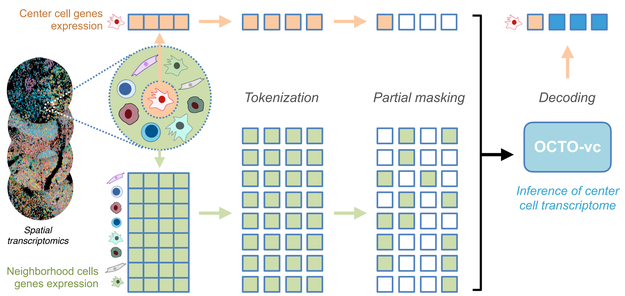


Figure 2. OCTO-vc model architecture. OCTO-vc is a transformer-based model trained to predict gene expression in tissue using spatial transcriptomics. For each training example, a central cell and its neighboring cells are used as input. Gene expression from both the center cell and neighboring cells are partially masked. The model learns to reconstruct the center cell’s gene expression based on its spatial context. This allows OCTO-vc to capture how local cellular environments shape gene expression, similar to how language models learn meaning from surrounding words.

For more informations on OCTO-vc

3. Clonal neighborhoods: An *in silico* experiment to model discrete cell-cell interactions within the NSCLC tumor microenvironment

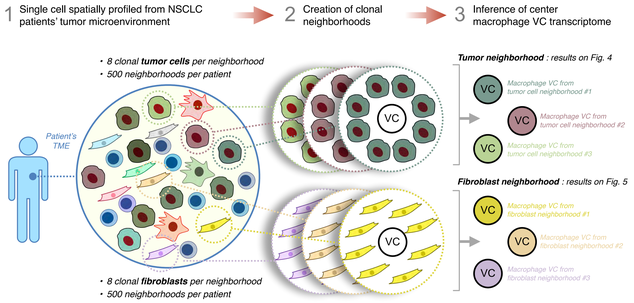


Figure 3. Clonal neighborhood experimental design. Spatially profiled single cells from patient samples are used to construct clonal neighborhoods for virtual cell (VC) experiments. Individual cells (tumor cells or fibroblasts) are sampled from the patients TME and placed as eight identical neighbors surrounding a ***macrophage VC*** to simulate controlled cellular neighborhoods at scale. 500 tumor or fibroblast clonal neighborhoods were simulated for each patient, leading to inferred gene expression on a total of 873,000 macrophages VC.

4. Tumor cells clonal neighborhoods identify tumor programs spatially associated with pro/anti-tumorigenic macrophages

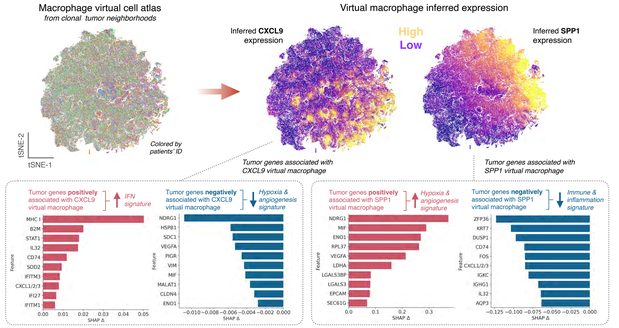


Figure 4. Clonal tumor neighborhoods influence virtual macrophage transcriptome. tSNE plots show virtual macrophages colored by patient (left) and by expression of CXCL9 and SPP1 (middle and right), two hallmark of macrophages states in tumors *[1]*. SHAP analysis from an XGBoost model reveals top clonal tumor-derived genes associated with each virtual macrophage program. Tumor cells promoting CXCL9 macrophages are enriched for antigen presentation and interferon response genes (MHC I, STAT1, IFITM1/3), indicating an inflamed, immune-activated tumors state. SPP1+ macrophages are linked to hypoxic and angiogenesis tumor cells program (VEGFA, NDRG1, LDHA).

5. Fibroblasts clonal neighborhoods identify fibroblasts programs spatially associated with pro/anti-tumorigenic macrophages

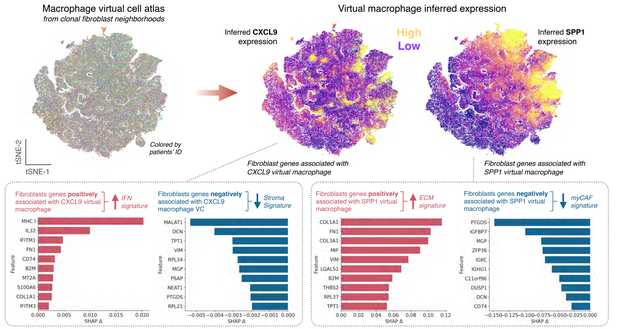


Figure 5. Clonal fibroblast neighborhoods influence macrophage-VC transcriptome. tSNE plots display virtual macrophages colored by patient identity (left) and expression of CXCL9 and SPP1 [1]. SHAP analysis from an XGBoost model identifies top fibroblast-derived genes driving variation in virtual macrophage states. CXCL9 macrophages associate with immune signaling genes (MHC I, IL32, CD74), pointing to an inflammatory fibroblast influence. SPP1+ macrophages are linked to extracellular matrix genes and fibrosis (e.g., COL1A1, COL3A1), suggesting pro-tumor fibroblast program.

6. ECM producing fibroblasts are enriched in tumors from KRAS/STK11 mutant NSCLC

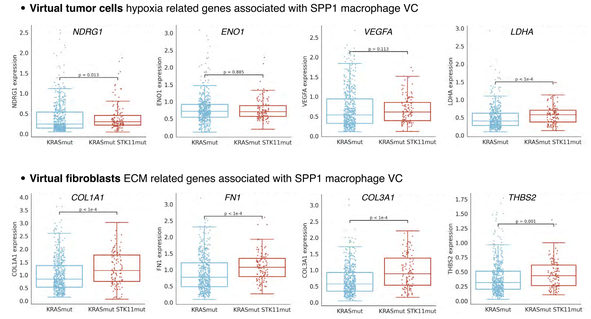
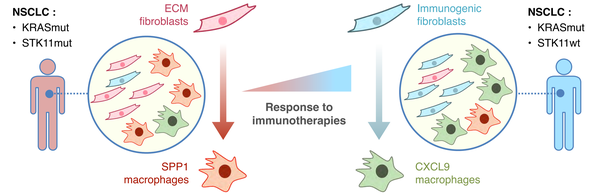


Figure 6. ECM related genes are elevated in virtual fibroblasts from KRAS$_{mut}$ STK11$_{mut}$ NSCLC tumors. Virtual tumor (top) and fibroblasts (bottom) gene expression across KRAS$_{mut}$ and KRAS$_{mut}$ STK11$_{mut}$ tumor cores was assessed using our OCTO-vc model. Inferred gene expression was averaged as the core level and compared between tumors genotypes. Genes from tumor cells and fibroblasts positively associated with SPP1 macrophage VC (see Fig.4 & 5) are displayed. Mann-Whitney test was used for statistical analysis.

7. SPP1-macrophage to fibroblast cross-talk may underlie immunotherapy resistance of KRAS/STK11 mutant NSCLC



Conclusion :

- Clonal neighborhoods provide a powerful tool to decipher discrete cell-cell interactions from human tumors.
- Conducting clonal neighborhoods experiments leveraging our foundation OCTO-vc model, we uncovered distinct gene tumor cells and fibroblasts programs that polarize macrophages towards immunogenic (CXCL9) or immunosuppressive (SPP1) states.
- ECM-producing fibroblast associated with SPP1 macrophages are enriched in KRASmut STK11mut NSCLC, suggesting a genotype-specific immunosuppressive microenvironment that may account for resistance to immunotherapies *[2], [3]*.
- Our approach identifies cellular crosstalks shaping the tumor microenvironment and provides insights for discovering actionable targets to improve immunotherapy in subpopulations of NSCLC patients

References *: [1] Bill, R. et al. Science (2023), [2] Skoulidis, F. et al. Cancer Discov. (2018), [3] Cords, L. et al. Cancer Cell (2024)*