Dig Vijay Kumar Yarlagadda

digvijayky@gmail.com | digvijayky.com

Education

Doctor of Philosophy in Computational Biology and Medicine 12/2026
Tri-Institutional Program at Cornell University, Sloan Kettering Institute, & Rockefeller University
Advisors: Christina Leslie and Joan Massagué; Rotations: Dana Pe'er and Quaid Morris
Dissertation: Representation Learning for Modeling the Spatial Architecture of Cancer

Master of Science in Computer Science 12/2017
University of Missouri-Kansas City; GPA: 4.0/4.0; Advisor: Praveen Rao
Thesis: One-Shot Learning Model for Cancer Diagnosis from Histopathological Images

Bachelor of Technology in Electronics and Communication Engineering 09/2015
Sree Vidyanikethan Engineering College, India; first class with distinction
Senior project: Efficient image compression using predictive vector quantization

Research Interests

My research interests lie at the interface of artificial intelligence and biology. I develop multimodal representation learning methods, 3D generative models, and vision models for spatial and single-cell transcriptomics, spanning model pretraining, post-training alignment, reinforcement learning and agentic AI infrastructure for scientific discovery. I combine these approaches with wet-lab experiments and clinical specimen analyses to study cellular programs and tissue interactions that sustain residual disease.

Publications

** indicates equal contribution.*

Yarlagadda, D., & Leslie, C. **Nicheverse: modeling spatial architecture using discrete representation learning**. *Manuscript in preparation.* nicheverse.org

Yarlagadda, D., Wang, Z., Jiang, H., Vuong, L., López-Sanmiguel, A., Yang, C.-Y., Kotecha, R.R., Chen, Y.-B., Hakimi, A.A., Leslie, C., & Massagué, J. 2026. **Developmental reversion underlies resistance to immune checkpoint blockade in kidney cancer**. *bioRxiv. Under revision at Cancer Cell.*

Hussain, Z.*, Yarlagadda, D.*, Li, Q., Tezcan, N., ..., Deborde, S., Leslie, C., & Sherman, M.H. 2026. **Invasion status stratifies the composition of the human pancreatic cancer perineural niche**. *bioRxiv. Under review at Cancer Discovery.*

Yarlagadda, D., Massagué, J., & Leslie, C. 2023. **Discrete representation learning for modeling imaging-based spatial transcriptomics data**. *Workshop on Bioimage Computing, International Conference on Computer Vision (ICCV).* **Oral**.

Yarlagadda, D., Rao, P., Rao, D., & Tawfik, O. 2019. **A system for one-shot learning of cervical cancer cell classification in histopathology images**. *SPIE Medical Imaging.*

Yarlagadda, D., & Leslie, C. 2026. **CellSplat: End-to-end 3D Gaussian splatting for cellular decomposition of spatial transcriptomics**. *Workshop on End-to-End 3D Learning, Conference on Computer Vision and Pattern Recognition (CVPR).*

Yarlagadda, D., & Leslie, C. 2026. **Spatially-aware cell type annotation using hierarchical vector-quantized variational autoencoders**. *Workshop on Computer Vision for Science, Conference on Computer Vision and Pattern Recognition (CVPR).*

Schüffler, P.*, Yarlagadda, D.*, …, & Fuchs, T. 2021. **Overcoming an annotation hurdle: Digitizing pen annotations from whole slide images**. *Journal of Pathology Informatics*, 12.

Ho, D., Yarlagadda, D., …, & Fuchs, T. 2021. **Deep multi-magnification networks for multi-class breast cancer image segmentation**. *Computerized Medical Imaging and Graphics*, 88.

Lopez Barron, D.E., Yarlagadda, D., Rao, P., Tawfik, O., & Rao, D. 2018. **Scalable storage of whole slide images and fast retrieval of tiles using Apache Spark**. *SPIE Medical Imaging.*

Lee, J., …, Yarlagadda, D., …, & Massagué, J. 2024. **TGF-β and RAS jointly unmask primed enhancers to drive metastasis**. *Cell.*

Chan, J., …, Yarlagadda, D., …, Pe'er, D., & Rudin, C. 2021. **Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer**. *Cancer Cell.*

Xie, C., …, Yarlagadda, D., …, & Fuchs, T. 2020. **Beyond classification: Whole-slide tissue histopathology analysis by end-to-end part learning**. *Medical Imaging with Deep Learning*, *Proceedings of Machine Learning Research (PMLR)*, 121:843-856. **Oral**.

Schüffler, P.J., Geneslaw, L., Yarlagadda, D., …, & Fuchs, T. 2021. **Integrated digital pathology at scale: A solution for clinical diagnostics and cancer research at a large academic medical center**. *Journal of the American Medical Informatics Association*, 28(9):1874-1884.

D'Alfonso, T.M., …, Yarlagadda, D., …, & Tan, L.K. 2021. **Multi-magnification-based machine learning as an ancillary tool for the pathologic assessment of shaved margins for breast carcinoma lumpectomy specimens**. *Modern Pathology.*

Schüffler, P., …, Yarlagadda, D., …, & Hameed, M. 2022. **Efficient visualization of whole slide images in web-based viewers for digital pathology**. *Archives of Pathology & Laboratory Medicine*, 146:1273-1280.

Kim, D., …, Yarlagadda, D., …, & Hanna, M. 2020. **(Re)Defining the high-power field for digital pathology**. *Journal of Pathology Informatics*, 11.

Godet, I., …, Yarlagadda, D., …, & Massagué, J. **Metabolic reprogramming and sensitivity of dormant cancer cells to RAS inhibitors in lung adenocarcinoma**. *In preparation.*

Spina, E., …, Yarlagadda, D., …, & Massagué, J. **A fibrotic niche sustains micrometastases in lung adenocarcinoma**. *In preparation.*

Software

- **Nicheverse** (nicheverse.org): hierarchical VQ-VAE that turns imaging-based spatial transcriptomics (Xenium, MERFISH, CosMx, seqFISH) into a shared vocabulary of cell states and tissue niches. Ships an MCP server and a Claude Agent Skill so Claude Code can run inference and annotation as tools. The annotation harness grounds each label in the code's own marker evidence and PubMed citations, applies deterministic verification gates, and runs an adversarial refuter model before a label is accepted.

Research Experience

Graduate student, Leslie Lab & Massagué Lab — Memorial Sloan Kettering Cancer Center
Advisors: Christina Leslie, Joan Massagué — *Jul 2020 - Dec 2026*

- **Nicheverse:** a foundation model pretrained on 57.1 million human cell and spot profiles across 140+ datasets. The pretraining objective combines negative-binomial reconstruction with a detection hurdle and codebook commitment penalty, enabling zero-shot transfer across ten spatial platforms without fine-tuning. Benchmarked against ten comparator methods on 23 annotated reference tasks. I built a post-training verification pipeline for model-generated annotations: a reward-like scoring function checks each proposed label against deterministic evidence gates (marker enrichment, citation resolution), and an adversarial refuter with the opposite optimization objective attempts to overturn each label under the same constraints, analogous to debate-based alignment and reward modeling in RLHF. Labels that fail are rejected or routed to human review.
- **Developmental reversion in clear cell renal cell carcinoma:** spatial and single-cell profiling revealed that cancer cells occupy a continuous axis of embryonic nephrogenesis and that immune checkpoint blockade enriches for developmentally committed states. Nicheverse identified relative $CD8^+$ T-cell exclusion in regions enriched for these states. Immunocompetent mouse models recapitulated the enrichment of nephrogenic persister cells after checkpoint blockade.
- **CellSplat:** performs transcript-to-nucleus assignment in imaging-based spatial transcriptomics by modeling nuclei as anisotropic 3D Gaussians. The geometry of these Gaussians is optimized directly against the volumetric DAPI signal and 3D transcript coordinates.
- **Perineural invasion in pancreatic ductal adenocarcinoma:** developed the imaging-based spatial transcriptomics framework and analyzed invaded and noninvaded nerve neighborhoods, resolving differences in malignant, stromal and immune organization.

Summer Intern, Troyanskaya Lab, Center for Computational Biology — Flatiron Institute, NY
Advisor: Olga Troyanskaya — *May 2024 - Aug 2024*

- Pretrained vision-language foundation models aligning light-sheet microscopy representations with bulk and single-cell genomic embeddings across embryonic development, using contrastive and reconstruction objectives for cross-modal transfer.

Machine Learning Engineer, Fuchs Lab — Memorial Sloan Kettering Cancer Center
Advisor: Thomas J. Fuchs — *Sep 2018 - Jun 2020*

- Developed C++/CUDA kernels for gigapixel whole-slide image tiling, streaming I/O and on-the-fly augmentation, and multi-GPU distributed training infrastructure for self-supervised and weakly supervised pretraining across more than 25 cancer types.
- Patented inventions and code licensed to Paige.AI; Paige Prostate received FDA De Novo authorization in 2021.

Graduate Research Assistant, Rao Lab — UMKC, Kansas City, MO
Advisor: Praveen Rao — *May 2016 - Dec 2017*

- Developed memory-augmented one-shot learning models for cervical cell classification with limited labels.

- Built an Apache Spark-based distributed in-memory platform for whole-slide image storage and tile retrieval.

Professional Experience

AgShift Santa Clara, CA

Deep Learning Engineer *Sep 2017 - Aug 2018*

- Developed convolutional neural networks for produce grading from multispectral and natural images; external advisor: Dilip Krishnan.

Strayos St. Louis, MO

Computer Vision & Machine Learning Engineer *Mar 2017 - Aug 2017*

- Developed 3D reconstruction, point-cloud processing and structure-from-motion pipelines for drone imagery.

Patents

- Machine learning model for analyzing pathology data from metastatic sites, U.S. Patent 11,449,994 (licensed to Paige.AI)
- Detection of annotated regions of interest in images, U.S. Patents 11,682,186 and 12,511,861 (licensed to Paige.AI)
- Scalable and efficient whole slide image reader for deep learning and viewer applications, U.S. Patent 12,038,571 (WO2020227401A1) (licensed to Paige.AI)
- Detection of labels in biomedical images, U.S. Patent 12,555,672 (WO2022235564A1) (licensed to Paige.AI)
- End-to-end part learning for image analysis, U.S. Patent Application No. 18/008,133 (licensed to Paige.AI)
- Digital pathology records database management, U.S. Patent Application No. 17/911,093 (licensed to Paige.AI)
- System and method for assessing food quality and shelf life prediction using deep learning techniques, U.S. Provisional Patent Application No. 62/686,826

Talks

Discrete representation learning for modeling imaging-based spatial transcriptomics data

NCI Junior Investigator Meeting, Bethesda, MD **(Invited Talk)** *2024*

WEHI Spatial Technology Symposium, Melbourne **(Invited Talk)** *2025*

Spatial Biology Summit, Stanford University **(Poster)** *2024*

ICCV Bioimage Computing Workshop, Paris, France **(Oral)** *2023*

Developmental reversion underlies resistance to immune checkpoint blockade in kidney cancer

Cancer Biology and Genetics Program Seminar, MSKCC **(Invited Talk)** *2025*

Gerry Metastasis Center Metastasis Club, MSKCC **(Invited Talk)** *2025*

Ritesh Kotecha Lab, Memorial Sloan Kettering Cancer Center **(Invited Talk)** *2025*

NCI Junior Investigator Meeting, Bethesda, MD **(Invited Talk)** *2023*

Mentoring Experience

Research Mentor, Memorial Sloan Kettering Cancer Center

- Joel Pepper (Computer Science, PhD)
- Leah Morales (Computational Biology, PhD)
- Shenni Liang (Computational Biology, PhD)
- Thahmina Ali (Computational Biology, PhD)
- Hang Yin (Immunology, PhD)
- Haskeerat Minhas (Computational Biology, MS)
- Kyle Kuo (Computer Science, undergraduate)

Service

Reviewer, Neural Information Processing Systems (NeurIPS)	*2019, 2020, 2021, 2023, 2025*
Reviewer, International Conference on Learning Representations (ICLR)	*2022, 2024, 2025, 2026*
Reviewer, International Conference on Machine Learning (ICML)	*2024, 2026*
Reviewer, Conference on Computer Vision and Pattern Recognition (CVPR)	*2026*
Reviewer, Conference on Health, Inference, and Learning (CHIL)	*2022*
Reviewer, Computers in Biology and Medicine	
Reviewer, PLOS ONE	
Organizer, AI for Overcoming Global Disparities in Cancer Care Workshop, ICLR	*2020*

Skills

Computational: Python, C++, CUDA, PyTorch, JAX, TensorFlow, scikit-learn, ScanPy, OpenCV, Point Cloud Library, Eigen, PyVista, Datashader, Apache Spark, Linux, Git

Methods: pretraining at scale (self-supervised, weakly supervised, contrastive objectives; C++/CUDA kernels for multi-GPU distributed training); post-training alignment and verification (evidence-gated reward signals, adversarial refutation, debate-style critique, harness engineering); reinforcement learning methods (reward design from experimental outcomes, online learning loops with iterative feedback); VQ-VAE and sparse autoencoders; 3D Gaussian splatting and neural rendering; vision-language models; graph neural networks

AI agents and tooling: Claude Code, MCP, Agent Skills, harness engineering for scientific verification

Experimental: mouse models of cancer metastasis; cell culture; functional assays; reporter development; imaging-based spatial transcriptomics; single-cell sequencing; multiplex fluorescence in situ hybridization; microfluidic systems; immunofluorescence and immunohistochemistry; light-sheet microscopy