

Education

Maharaja Agrasen Institute of Technology

Delhi, India

B.Tech, Information Technology,

Graduation Date: July 2023

Coursework: Data Structures and Algorithms, Computer Architecture, Software Engineering, Discrete Mathematics, Database Management Systems, Computer Networks, Operating Systems, Object-Oriented Programming, Theory of Computation, Compiler Design, Calculus, Linear Algebra, Probability and Statistics

Publication

Publications

- **Yadav A**, Alvarez K, Yip KY, Ruppini E, Kumsta C, Sinha S. Mapping structural aging across human tissues reveals tissue-specific trajectories, coordinated deterioration, and genetic determinants (PathStAR). 2025. Preprint. Under review at *Nature*. Available from: <https://doi.org/10.1101/2025.09.29.679316>
- **Yadav A**, Alvarez K, Adeleye A, Wang YX, Sinha S. TLPath predicts telomere length in human tissues from histopathology images. 2025. Preprint. Forthcoming at *Cell Reports Methods*. Available from: <https://doi.org/10.1101/2025.03.04.641489>
- Sinha RK, **Yadav A**, Sinha S. PathQC: Determining molecular and physical integrity of tissues from histopathological slides. 2025. Preprint. Available from: <https://doi.org/10.1101/2025.09.29.679347>
- Birk S, Bonafonte-Pardàs I, Feriz AM, Boxall A, Agirre E, Memi F, Maguza A, **Yadav A**, Armingol E, Fan R, Castelo-Branco G, Theis FJ, Bayraktar OA, Talavera-López C, Lotfollahi M. Quantitative characterization of cell niches in spatially resolved omics data. *Nat Genet*. 2025;57:897–909. doi:10.1038/s41588-025-02120-6
- Choudhury M, Liu L, **Yadav A**, Chapman O, Ahmadi Z, Younis R, Sharma C, Goel N, Sridhar S, Kenkre R, Dutta A, Wang S, Shulman E, Rahman Dhruva S, Hoang D-T, Tharp K, Paul M, Malicki D, Yip KY, Ruppini E, Chavez L, Sinha S. ecPath detects ecDNA in tumors from histopathology images. 2024. Preprint. Available from: <https://doi.org/10.1101/2024.11.13.623494>

Research Experience

Sanford Burnham Prebys

San Diego, CA

Research Assistant

Oct 2024 – Present

- **PathStAR** (first author; preprint): Developed a large-scale framework mapping structural aging across ~25,000 GTEx histology slides, integrating morphology with transcriptomic, methylation, and germline data.
- **TLPath** (first author; preprint): Built deep learning models that predict telomere length directly from H&E images, demonstrating that tissue architecture encodes telomere-associated aging signals.
- **ecPath** (co-author; preprint): Designed and trained transcriptomics-guided multimodal models to detect extrachromosomal DNA (ecDNA) from routine pathology slides.
- Leading development of the first **mouse histopathology foundation model** using large-scale self-supervised learning to extend digital pathology beyond human tissues.
- Conducting a systematic **benchmarking of digital pathology based vision–language foundation models** (PLIP, Conch, TITAN, BiomedCLIP) to evaluate alignment quality, interpretability, and biological fidelity.

Google Summer of Code (DeepChem)

Remote

Open Source Fellow

May 2024 – Sept 2024

- Developed an end-to-end computational pipeline for **drug target identification** using UniProt features and ML-based prioritization.
- Predicted protein structure and druggability to validate and rank potential targets across diseases.

Wellcome Sanger Institute (Lotfollahi Lab)

Remote

Research Assistant

June 2024 – July 2024

- Contributed to the revision of **NicheCompass** (Nature Genetics, 2025), retraining alternative model variants and improving robustness and scalability.
- Assisted in building a **self-supervised foundation model for spatial transcriptomics** using I-JEPA, extending representation learning beyond histology.

Indian Institute of Technology Delhi (Gupta Lab)

Delhi, India

Research Assistant

Dec 2023 – May 2024

- Developed a cell-type classification pipeline using autoencoders, graph coarsening, and GNNs; accepted at **ISBI 2024**.
- Built ML models to classify SARS-CoV-2 variants from Raman spectra with 94% and 89% accuracy.

Ersilia Open Source Initiative

Remote

Machine Learning Intern

Dec 2022 – Mar 2023

- Built **ChemSampler**, a generative modeling library for molecular design using seed-conditioned sampling.
- Implemented an easy-to-use API for molecule generation and integrated external chemical databases (ChEMBL, PubChem).

Google Summer of Code

Remote

Open Source Developer

May 2022 – Sept 2022

- Designed rule-based NLP models (NLTK) to identify complex biological patterns in genomic datasets.
- Built a pipeline to extract features such as gene deletions and allele substitutions from ~1M genotype records across global PomBase labs.

Conference Presentations & Posters

- **Yadav, A.**, et al. *Mapping Structural Aging Across Human Tissues (PathStAR)*. Selected for **oral presentation** at the **AI for Healthy Aging and Longevity Workshop**, AAAI 2026.
- **Yadav, A.**, et al. *Structural Aging in Human Ovarian Tissue: Early Insights from PathStAR*. Poster at the **Reproductive Aging Conference (RAC)**, San Jose, CA, May 2025.
- **Yadav, A.**, et al. *TLPPath: Leveraging Digital Pathology Foundation Models for Telomere Length Prediction*. Poster at the **Women in Computer Vision (WiCV) Workshop**, CVPR 2025.
- Ekta Srivastava, **Anamika Yadav**, Manoj Kumar, Ishaan Gupta, Sandeep Kumar. *Optimized Graph Coarsening for Improved Single-Cell RNA-seq Analysis*. Short Paper at **ISBI 2024**.

Skills

- Programming Infrastructure: Python (PyTorch, NumPy, pandas, scikit-learn), Bash, Git, Linux, Slurm HPC workflows, GPU-based, large-scale model training and data processing
- Machine Learning & Representation Learning: Vision Transformers (ViT, DINO, I-JEPA), CLIP-style vision-language pretraining, multimodal contrastive/self-supervised learning, embedding alignment, interpretability and probing
- Computational Pathology: Whole-slide image preprocessing (OpenSlide), patch-level feature extraction, global k-means morphological clustering, spatial co-occurrence and graph-based tissue modeling, digital pathology foundation model benchmarking
- Multi-omics Integration: Bulk RNA-seq differential expression (DESeq2, limma), methylation and pathway enrichment analysis (GSEA, MSigDB), integrative modeling of histology with transcriptomic and genomic characteristics.