

# Yinuo Jin

645 W. 59th Street, New York, NY  
yj2589@columbia.edu | yjin@colgate.edu  
Webpage : <https://yinuojin.github.io>  
Github : <https://github.com/YinuoJin>  
(917) 291-8421

## EDUCATION

**Columbia University** *Sep. 2019 - May. 2021 (Expected)*  
Bachelor of Engineering, Computer Science, GPA: 3.79/4.00

**Colgate University** *Sep. 2016 - May. 2019*  
Bachelor of Arts, Physics, GPA: 3.88/4.00

**Aquincum Institute of Technology** *Spring 2019*  
CS Study Abroad, affiliated with Budapest University of Technology and Economics (BME)

## RESEARCH EXPERIENCE

**Cell Image Segmentation - single-cell level spatial profiling** *May 2020 - Present*  
Computational Cancer Biology Lab, Columbia University  
*Supervisor: Dr. Elham Azizi*

- Implemented and trained an end-to-end Feature Pyramid Networks for cell nuclei image segmentation
- Refined segmentation results for adjacent, cluttered cells with incorporation of geometrical features, multi-level watershed and morphological post-processing steps
- Transferred the segmentation knowledge learnt from *in vitro* images to *in situ* tissue images

**Predicting intercellular Ligand-Receptor interactions** *May 2020 - Aug. 2020*  
Computational Cancer Biology Lab, Columbia University  
*Supervisor: Dr. Elham Azizi*

- Performed preprocessing, clustering & cell-type inference on multiple liver tissue scRNA-seq datasets
- Calculated pairwise cluster distance within and across datasets for cell-type imputation, predicted cluster-wise ligand-receptor interactions via NicheNet

**PICASSO: Copy Number Variation Inference** *Sep. 2019 - Present*  
Computational Biology Lab, Columbia University  
*Supervisors: Dr. Itsik Pe'er, Dr. Sitara Persad*

- Assisted developing PICASSO, an iterative HMM model to infer copy number variation regions (CNVs) and intra tumor heterogeneity from single-cell RNA-seq data
- Inferred cellular clonal structures and phylogenetic lineage using EM-clustering, determined the stopping criterion of tree iterations guided by BIC and hierarchical clustering
- Implemented large-scale, full genome length randomized simulations synthesized from real scRNA-seq samples for benchmark
- Optimized the running time over 60% by parallelizing the core computation steps of the framework

**NanoFusion: Oxford Nanopore Sequencing Analysis** *May 2019 - Aug. 2019*  
Computational Biomedicine Workgroup, BME  
*Supervisor: Dr. Péter Sárközy*

- Designed pipelines to compare performances and accuracy metrics of multiple Oxford Nanopore Sequencing basecallers on NA12878 Chr1 data
- Performed PHRED-score biased Multiple Sequence Alignments to basecalling reads, reduced soft-clipped regions and improved the mapping accuracy

**Seismic Data Analysis** *May 2017 - Sep. 2017*  
Geophysics Lab, Colgate University  
*Supervisor: Dr. Aubrey Adams*

- Retrieved and analyzed Earthquake events in New York state from 2016 to 2017 with Python and AWK, refined data on earthquake arrival time, epicentral locations & depth
- Characterized and distinguished shallow earthquakes from mining events in the Adirondacks region

|                      |   |   |
|----------------------|---|---|
| PROJECTS             | <b>Ancestral Sequence Reconstruction (ASR)</b><br>Course: Computational Genomics<br><i>Supervisor: Dr. Itsik Pe'er</i>  | <i>Jan. 2020 - May. 2020</i>  |
|                      | <ul style="list-style-type: none"> <li>· Implemented algorithms of among-site rate variations (ASRV) and site-specific constraints for marginal ASR, extending features to <i>TreeTime</i>, a python library for phylogenetic analysis</li> <li>· Evaluated reconstruction results on public mitochondrial datasets, improved the overall root state likelihood scores comparing against the baseline model</li> </ul>                                      |   |
|                      | <b>Gene-level Analysis of scRNA-seq data</b><br>Course: Machine Learning for Functional Genomics<br><i>Supervisor: Dr. David Knowles</i>  | <i>Oct. 2019 - Dec. 2019</i>  |
|                      | <ul style="list-style-type: none"> <li>· Predicted cell types of human SC-islets dataset with XGBoost model, extracted cell-type specific differential expression features with SHAP</li> <li>· Constructed gene co-expression networks from the trained model, evaluated the reliability of the networks with WGCNA</li> </ul>   |   |
|                      | <b>Phylogenetics tree with 18S rRNA sequences</b><br>Course: Computational Biology and Medicine<br><i>Supervisor: Dr. Péter Sárközy</i>   | <i>Mar. 2019 - May. 2019</i>  |
|                      | <ul style="list-style-type: none"> <li>· Refined a phylogenetic tree with progressive alignments &amp; hierarchical clustering on selected Vertebrata 18S rRNA sequences</li> <li>· Evaluated the results with NCBI taxonomy database &amp; MUSCLE package</li> </ul>   |   |
| MANUSCRIPTS          | <b>Yinuo Jin*</b> , Alexandre Toberoff*, and Elham Azizi. “Transfer learning framework for cell segmentation with incorporation of geometric features” Accepted to <i>NeurIPS 2020 Workshop on Learning Meaningful Representations of Life (LMRL)</i>   |   |
|                      | Aubreya Adams, <b>Yinuo Jin</b> , Monica Dimas, and Isabel Dove. “Leveraging USArray and Regional Networks to Characterize Natural and Mining Seismicity at the Intersection of the Grenville and Taconic Provinces” Submitted to <i>Seismological Research Letters: Eastern Section</i>  |   |
| RELEVANT COURSEWORKS | <b>Computer Science:</b> <ul style="list-style-type: none"> <li>· Computational Genomics</li> <li>· Computational Biology and Medicine</li> <li>· Artificial Intelligence</li> <li>· Analysis of Algorithms</li> </ul> <b>Statistics &amp; Biology:</b> <ul style="list-style-type: none"> <li>· Probability Theory</li> <li>· Statistical Inference (<i>Registered: Spring 2021</i>)</li> <li>· Principles &amp; Applications of Modern DNA-seq</li> </ul> | <ul style="list-style-type: none"> <li>· Machine Learning for Functional Genomics</li> <li>· Data Mining</li> <li>· Machine Learning</li> <li>· Advanced Programming</li> <li>· Statistical Machine Learning for Genomics</li> <li>· Genetics</li> <li>· Cells &amp; Human development</li> </ul> |
| HONORS & AWARDS      | Sigma Pi Sigma (SPS) Physics Honor Society<br>Charles A. Dana Scholars (top 5% seniors, juniors and sophomores)<br>Colgate University Douglas Rankin '53 Fellowships<br>Colgate University Hackett-Rathmell 1968 Memorial Fund  | <i>2020</i><br><i>2018</i><br><i>2017</i><br><i>2017</i>  |
| SKILLS               | <b>Languages:</b> Python, C++, Java, Bash, R<br><b>Framework/Tools:</b> Pytorch, GATK, Samtools, Minimap2   |   |