Alex J. Lee

alex.lee@ucsf.edu alexj-lee 0000-0003-0001-2848 alexlee.netlify.app

Education

B.A. Biology (2014) — Vassar College

Ph.D Biological and Medical Informatics (09/2021-present) — UCSF

Summary

Developing state-of-the-art deep learning methods for computational biology across neuroanatomy, representation learning, and protein sequence modeling.

- **Discovering hidden brain organization using AI**: built CellTransformer, novel transformer architecture for multi-million-cell (10-40M) scale spatial transcriptomics datasets, enabling discovery of 1000+ anatomically realistic brain areas, including 10s/100s+ unannotated subregions.
- **Pioneering synthetic data approaches in protein design**: created BackboneRef, the first large-scale structure-based synthetic protein dataset (240,811 structures; 73M sequences; part of Microsoft Research's Dayhoff Atlas. BackboneRef data augmentation improves expression by 87.5% versus baselines (in revision at Cell Systems).

Technical skills

- Expert in PyTorch, HPC utilization (Azure/Amulet, SGE, Slurm), containerization (Docker, Singularity), and MLOps tooling (wandb, hydra) for scientific applications requiring high computational throughput and reproducibility.
- Experience with Git/Github: familiar with navigating large/complex code-bases ex. Dayhoff pLM training code.
- Familiar with distributed computing paradigms such as Dask, Ray, and serverless frameworks such as Modal/Coiled.

Employment and research experience

Abbasi lab (UCSF)

Graduate student, Biological and Medical Informatics PhD program — 09/2021-current
Thesis committee: Drs. Katherine Pollard & Ahmed Alaa

Scalable transformer-based representation learning for spatial transcriptomics to map the brain.

- Created state-of-the-art transformer pipeline (CellTransformer) for whole-brain spatial transcriptomics data (10-40M+ cells), enabling scalable, data-driven discovery of brain regions beyond existing atlases (1000s+ of regions, vs. 670 extant).
- **Improved on state-of-the-art by 784%** (spatially discreteness), with **13.4%** greater similarity to human annotations (Allen CCF).
- Led collaboration with Molecular Genetics group at Allen Institute for Brain Science; project resulted in **$338K** grant (1R01LM014619-01A1) and **first-author paper** in *Nature Communications*.
- Wrote modular testing software to reveal medical keywords that reliabily cause lowered accuracy in LLM medical Q&A (*TMLR*, **second auth.**).

Calico Life Sciences *Summer internship in Calico Applied ML group* — 06-2025 - 09/2025
Generative deep learning methods development for spatial genomics.

- Creation of a latent flow matching method for unmeasured gene imputation achieving state-of-the-art performance in cell type classification and accuracy. Further details unavailable due to NDA.

Microsoft Research New England

Summer intern in the BioML group 06/2024 - 09/2024
Drs. Ava Amini (primary) and Kevin Yang (secondary)
First demonstration that synthetic data augmentation for protein language modeling can directly and significantly improve wet-lab experimental results and generation novelty.

- Created BackboneRef, the largest-to-date synthetic protein dataset (240,811 *de novo* designed backbones, 2.4M predicted structures, 73M sequences; all open source), to augment training data for protein language models–utilized 100s of GPUs at once on Azure to run a pipeline using RFDiffusion, ProteinMPNN, Omegafold, and Foldseek.
- Used synthetic data augmentation to achieve 87.5% increase in expression success rate, 4% increase in pLDDT, and 50% increase in novelty of *de novo* designed proteins compared to baseline when using 170m-parameter state-space models.

Seeley lab (UCSF)

Research associate II {06/2018 - 09/2021}
Computational analysis and tool development to study the brain in health and during neurodegeneration.

- HPC (containerization, parallelization implementation) pipeline development for unified and optimal preprocessing of functional and structural MRI data, on UCSF Wynton (SGE) cluster, enabling rapid processsing at **1-5K+ scan scale** across thousands of nodes–pipeline resulted directly in **9 published papers** from venues like *Neuroimage*, *JAMA Neurology*, and *Brain*.
- Led, trained, and evaluated novel attention U-net model for segmenting neurons and microglia in large multiplexed immunofluorescence datasets of human cortical tissue [in collaboration with GE Life Sciences].
- Created computer vision pipeline for first-in-kind study comparing CTE biomarkers from PET imaging to postmortem pathology; one of the first demonstrations of low correlation of during-life and postmortem CTE biomarkers.

Nano Precision Medical Inc., now Vivani Medical Inc.

Research associate II {02/2015 - 05/2018}
Project leader–led tech. and analytical development for optimization of novel implantable drug delivery device using nanotstructured titanium for long term delivery of GLP-1 drugs.

- Project leader on key proof-of-concept development of sterile implantation protocols for *in vivo* (mouse) and *in vitro* demonstration of clinically meaningful drug delivery performance, supporting NPM-119 regulatory filings (IND); NPM-119 now entering first-in-human trials.
- Created proof-of-concept 6 and 12-month formulations for GLP1-RA using novel analytical chemistry approaches (HPLC/FPLC/MS) to identify critical degradation pathways.
- First employee designated to lead cross-departmental experimental collaboration integrating novel chemistry approaches with materials characterization tech-development,

Awards and grants

	award amount
Vassar College Undergraduate Research Summer Institute fellowship (2013)	$3000
NSF Graduate Research Fellowship Program Honorable Mention (2023)	
UCSF ARCS (Achievement Rewards for College Scientists) Scholarship (2023)	$17,917/yr
UCSF Graduate Research Mentorship Fellowship (2024)	$10,000/yr

UCSF Discovery Fellowship (2023-current) Tuition, $2k/yr stipend

Preprints / in preparation / in review

1. Yang, Kevin K.*, Sarah Alamdari*, **Alex J. Lee***, Kaeli Kaymak-Loveless, Samir Char, Garyk Brixi, Carles Domingo-Enrich, Chentong Wang, Suyue Lyu, Nicolo Fusi, Neil Tenenholtz, Ava P. Amini. (bioRxiv, 2025), p. 2025.07.21.665991 https://doi.org/10.1101/2025.07.21.665991 [* co-first; in revision at Cell Systems]

Full list of publications

1. Brown, Jesse A., **Alex J. Lee**, Kristen Fernhoff, Taylor Pistone, Lorenzo Pasquini, Amy B. Wise, et al., "Atrophy-driven functional network collapse in neurodegenerative disease" bioRxiv, 2023. https://doi.org/10.1101/2023.12.01.569654 (accepted at Nature Communications)
2. Flagan, Taru M., Chu, Stephanie A., Hakkinen, Suvi, Zhang, Liwen, et al. "Functional Connectivity Associations with Markers of Disease Progression in GRN Pathogenic Variant Carriers". Annals of Clinical and Translational Neurology. September 16, 2025. http://doi.org/10.1002/acn3.70170.
3. **Lee, Alex J.**, Shenqin Yao, Nicholas Lusk, Lydia Ng, Michael Kunst, Hongkui Zeng, Bosiljka Tasic, and Reza Abbasi-Asl. "Data-Driven Fine-Grained Region Discovery in the Mouse Brain with Transformers." Nature Communications, 2025. https://doi.org/10.1038/s41467-025-64259-4.
4. Rui, Patrick X., **Alex J. Lee**, Satvik Lolla, Vincent Wang, Russell Ro, Qiming Cui, and Reza Abbasi-Asl. "Assessing biomedical knowledge robustness in large language models by query-efficient sampling attacks". Transactions in Machine Learning Research (December 16, 2024). https://openreview.net/forum?id=pvol5JyVYB.
5. **Lee, Alex J.**, Robert Cahill, and Reza Abbasi-Asl. "Practical Considerations for Machine Learning-Enabled Discoveries in Spatial Transcriptomics." GEN Biotechnology 3, no. 3 (June 1, 2024): 130–35. https://doi.org/10.1089/genbio.2023.0050.
6. Cahill, Robert, Yu Wang, R. Patrick Xian, **Alex J. Lee**, Hongkui Zeng, Bin Yu, Bosiljka Tasic, and Reza Abbasi-Asl. "Unsupervised Pattern Identification in Spatial Gene Expression Atlas Reveals Mouse Brain Regions beyond Established Ontology." Proceedings of the National Academy of Sciences of the United States of America 121, no. 37 (September 10, 2024): e2319804121. https://doi.org/10.1073/pnas.2319804121.
7. Pasquini, Lorenzo, Felipe L. Pereira, Sahba Seddighi, Yi Zeng, Yongbin Wei, Ignacio Illan-Gala, Sarat C. Vatsavayai, et al. "Frontotemporal Lobar Degeneration Targets Brain Regions Linked to Expression of Recently Evolved Genes." Brain: A Journal of Neurology 147, no. 9 (September 3, 2024): 3032–47. https://doi.org/10.1093/brain/awae205.
8. Zhang, Liwen, Taru M. Flagan, Suvi Hakkinen, Stephanie A. Chu, Jesse A. Brown, **Alex J. Lee**, Lorenzo Pasquini, et al. "Network Connectivity Alterations across the MAPT Mutation Clinical Spectrum." Annals of Neurology 94, no. 4 (October 2023): 632–46. https://doi.org/10.1002/ana.26738.
9. Pasquini, Lorenzo, Susanna L. Fryer, Stuart J. Eisendrath, Zindel V. Segal, **Alex J. Lee**, Jesse A. Brown, Manish Saggar, and Daniel H. Mathalon. "Dysfunctional Cortical Gradient Topography in Treatment-Resistant Major Depressive Disorder." Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 8, no. 9 (September 2023): 928–39. https://doi.org/10.1016/j.bpsc.2022.10.009.
10. Pasquini, Lorenzo, Fatemeh Noohi, Christina R. Veziris, Eena L. Kosik, Sarah R. Holley, **Alex Lee**, Jesse A. Brown, et al. "Dynamic Autonomic Nervous System States Arise during Emotions and Manifest in Basal Physiology." Psychophysiology 60, no. 4 (April 2023): e14218. https://doi.org/10.1111/psyp.14218.
11. Chow, Tiffany E., Christina R. Veziris, Renaud La Joie, **Alex J. Lee**, Jesse A. Brown, Jennifer S. Yokoyama, Katherine P. Rankin, et al. "Increasing Empathic Concern Relates to Salience Network Hyperconnectivity in Cognitively Healthy Older Adults with Elevated Amyloid-β Burden." NeuroImage. Clinical 37, no. 103282 (2023): 103282. https://doi.org/10.1016/j.nicl.2022.103282.

12. Brown, Jesse A., **Alex J. Lee**, Lorenzo Pasquini, and William W. Seeley. "A Dynamic Gradient Architecture Generates Brain Activity States." NeuroImage 261, no. 119526 (November 1, 2022): 119526. https://doi.org/10.1016/j.neuroimage.2022.119526.

13. Mantyh, William G., Salvatore Spina, **Alex Lee**, Leonardo Iaccarino, David Soleimani-Meigooni, Elena Tsoy, Taylor J. Mellinger, et al. "Tau Positron Emission Tomographic Findings in a Former US Football Player with Pathologically Confirmed Chronic Traumatic Encephalopathy." JAMA Neurology 77, no. 4 (April 1, 2020): 517–21. https://doi.org/10.1001/jamaneurol.2019.4509.

14. Pasquini, Lorenzo, Gianina Toller, Adam Staffaroni, Jesse A. Brown, Jersey Deng, **Alex Lee**, Katarzyna Kurcyus, et al. "State and Trait Characteristics of Anterior Insula Time-Varying Functional Connectivity." NeuroImage 208, no. 116425 (March 2020): 116425. https://doi.org/10.1016/j.neuroimage.2019.116425.

Conference abstracts & posters

1. **Alex Jihun Lee**, Alma Dubuc, Shenqin Yao, Nicholas Lusk, Lydia Ng, Michael Kunst, Hongkui Zeng, Bosiljka Tasic, Reza Abbasi-Asl. "Multimillion cell self-supervised representation learning enables organ-scale tissue niche discovery". ICML Multi-modal Foundation Models and Large Language Models for Life Sciences workshop (2025).

2. **Alex Jihun Lee**, Sarah Alamdari, Chentong Wang, Reza Abbasi-Asl, Kevin Yang, Ava Amini. "Structure-based synthetic data augmentation for protein language models". ICLR Generative and Experimental Perspectives for Biomolecular Design (GEM) workshop (2025). https://openreview.net/forum?id=XAhcYXdqSJ. Selected for direct submission to Nature Biotechnology by editor Barbara Cheifet.

3. Chen, Yu., Winson Fu Zun Yang, Myrthe G. Rijpma, Jesse A. Brown, **Alex Lee**, Gianina Toller, Howard J. Rosen, Joel H. Kramer, Bruce L. Miller, Katherine P. Rankin. "The hierarchical organization of the salience, default, and executive networks in social inferences". Organization for Human Brain Mapping. (2024).

4. **Lee, A. J.**, Hongkui Zeng, Bosiljka Tasic, Reza Abbasi-Asl. "Masked autoencoders for spatial clustering in the mouse brain. Machine Learning in Computational Biology". (2023).

5. Brown, Jesse A., **Alex J. Lee**, Lorenzo Pasquini, Adit Friedberg, Gil D. Rabinovici, Joel H. Kramer, Maria Luisa Gorno Tempini, Howard J. Rosen, Bruce L. Miller, and William W. Seeley. "Local Neurodegeneration and Global Connectivity Adaptation across the FTD-AD Spectrum." Alzheimer's & Dementia 17, no. S6 (2021): e055308. https://doi.org/10.1002/alz.055308.

6. Flagan, Taru M., Stephanie A. Chu, Suvi Hakkinen, David McFall, Carolin Heller, Jonathan D Rohrer, Jesse A. Brown, et al. "Complement and NfL Associations with Brain Structure and Functional Connectivity Alterations in Presymptomatic and Symptomatic GRN Mutation Carriers." Alzheimer's & Dementia 17, no. S4 (2021): e050737. https://doi.org/10.1002/alz.050737.

7. Zhang, Liwen, Taru M. Flagan, Stephanie A. Chu, Suvi Hakkinen, Jesse A. Brown, **Alex J. Lee**, Lorenzo Pasquini, et al. "Presymptomatic and Symptomatic MAPT Mutation Carriers Feature Functional Connectivity Alterations." Alzheimer's & Dementia 17, no. S4 (2021): e054128. https://doi.org/10.1002/alz.054128.

Presentations

Invited speaker, ML for Protein Engineering (ML4PE)

Community service

Reviewer for:

- NeurIPS 2025 Workshop on Multimodal Foundation Models and Large Language Models for Life Sciences
- ICML 2025 Workshop on Multimodal Models and Large Language Models for Life Sciences

- AAAI 2026 Workshop on Neuro for AI & AI for Neuro

Patents

1. (PCT) Jesse A. Brown; Alex J. Lee. 2024. SYSTEMS AND METHODS FOR AUTOMATED NEUROIMAGING SIGNAL PROCESSING, WO2024259443 - SYSTEMS AND METHODS FOR AUTOMATED NEUROIMAGING SIGNAL PROCESSING.

References

Reza Abbasi-Asl
Associate Professor, Neurology at UCSF
reza.abbasiasl@ucsf.edu

Ava Amini
Principle Researcher at Microsoft Research New England
ava.amini@microsoft.com

Kevin Yang
Senior Principal Researcher at Microsoft Research New England
yang.kevin@microsoft.com

Bosiljka Tasic
Directory of Molecular Genetics at Allen Institute for Brain Science
bosiljkat@alleninstitute.org

Jesse Brown
Founder, Radiata [former Assistant Prof. Neurology at UCSF]
jesse.brown@radiata.ai