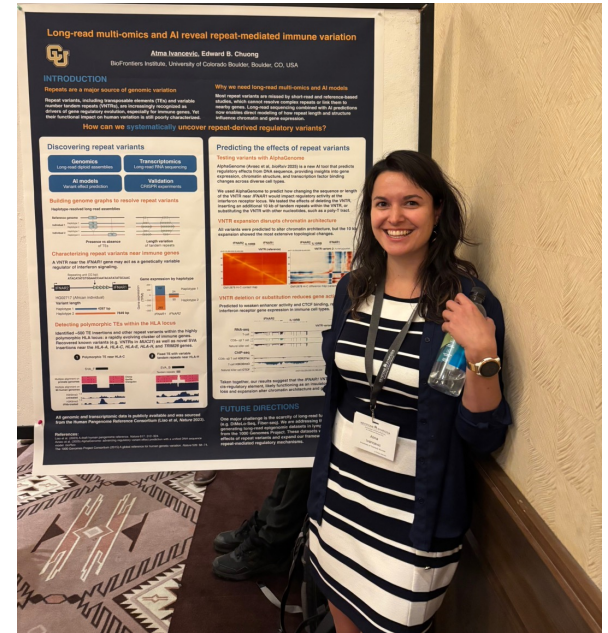


AI in Biology Recap (Part 1)

Eldorado Hotel, Santa Fe, New Mexico

15-19th Sept 2025



How is AI used in molecular biology?

- **Sequence-to-function prediction:** predicting gene expression, regulatory activity, and chromatin state from DNA sequence
- **Single-cell analysis:** identifying cell types, developmental trajectories, and spatial organization from transcriptomic and genomic data
- **Protein design and analysis:** generating novel protein sequences and predicting structure-function relationships using language models
- **Microscopy and imaging:** automated analysis of cellular morphology, organelle structure, and tissue architecture

Conference demographics

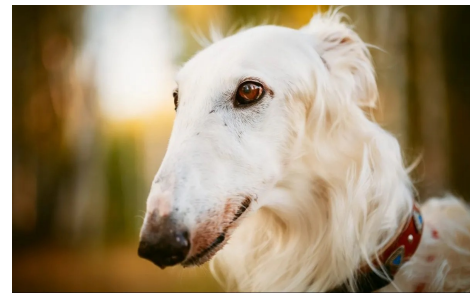
Tech-heavy audience with a strong industry presence

- Mostly computer scientists from industry or Ivy League universities
- Major representation from AI-focused biotech & tech companies (Amazon, Genentech, Calico Life Sciences, Recursion, NewLimit)
- Lack of domain knowledge in some instances

Predicting Gene Expression from DNA Sequence

(David Kelley, Calico Life Sciences)

- **Objective:** comprehensively annotate how every nucleotide influences cell-type and state-specific gene regulation
- **Borzoï** is a new model for predicting tissue-specific RNA-seq, eQTLs, sQTLs and paQTLs
- **Borzoï Prime** (like 3') extends to cell-type resolution by incorporating single cell atlases as pseudo-bulk profiles
- Transfer learning using **parameter-efficient fine-tuning (PEFT)** techniques enables modelling new datasets without retraining



Borzoï: [Linder et al, Nature Genetics \(2025\)](#)

Borzoï Prime: [Linder, Yuan & Kelley, bioRxiv \(2025\)](#)

PEFT: [Yuan, Linder & Kelley, bioRxiv \(2025\)](#)

Machine Learning for Single Cell & Spatial Transcriptomics

- **DR-GEM**: improves representation and annotations of rare cells in single cell data.
Christine Yeh, Stanford University
- **DREAMS**: preserves both local and global structure in dimensionality reduction.
Dmitry Kobak, Tübingen University
- **EHRformer**: uses spatial transcriptomics to identify CXCL9+/CXCL10+ macrophages that drive chronic rejection of lung transplants. *Lucy Luo, Northwestern University*
- **NIH malaria study**: uses foundation models and self-supervised learning to analyze skin inflammatory responses from histopathology slides of mosquito bite sites. *Julius Herzog, NIH Laboratory of Infectious Diseases*

DR-GEM: [Yeh et al, bioRxiv \(2025\)](#)

DREAMS: [Kury et al, arXiv \(2025\)](#)

EHRformer: [Luo et al, medRxiv \(2025\)](#)

Interesting talks summary

Deep Learning Prediction of The Chromatin Response to Transcription Factor Dosage from DNA Sequence (Sahin Naqvi, Boston Children's Hospital/Harvard Medical School): New PI, How does chromatin state respond to TF dosage? Used deep learning (ChromBPNet) to predict chromatin accessibility from motif features and validated experimentally (fine-tuned). [Naqvi et al., Cell Genomics \(2025\)](#)

Machine learning for single-cell spatial transcriptomics (Mingyao Li, University of Pennsylvania School of Medicine): Her lab is using AI in different ways. Multiple models for super-resolution tissue architecture(iStar), scaling up spatial transcriptomics to study large tissues (iSCALE), and select best regions to capture enough variation (S2Omics). [Coleman et al., Nature Methods \(2024\)](#).

Protein language models: strengths and limitations (Mona Singh, Princeton University): Presented on new protein LM trained with protein sequences for variant effect prediction and conservation. Main conclusion: Proteins with higher homologs available have more variants, and better predictions can be made from locality-aware pooling. [Hoang and Singh, Bioinformatics \(2025\)](#) : Bag-of-Mer (BoM) pooling: hierarchical attention pooling scheme designed to capture localized biological signals

Keynote Speaker: David Baker, University of Washington

Protein Design using Deep learning:

- Introduced the RF-diffusion model to generate small molecules from desired structure to output sequence
- **Uses of protein design:**
 - Antibody specificity generation
 - Binders to folded and unfolded proteins
 - Binding to TAU and target E3 ligases for protein degradation
 - Effector molecules to trigger conformational change
 - Drugs that only function in a desired place at a desired time
 - And many more...
- **Things to think about when designing proteins:**
 - Protein stability
 - Correct folding
 - Specificity

