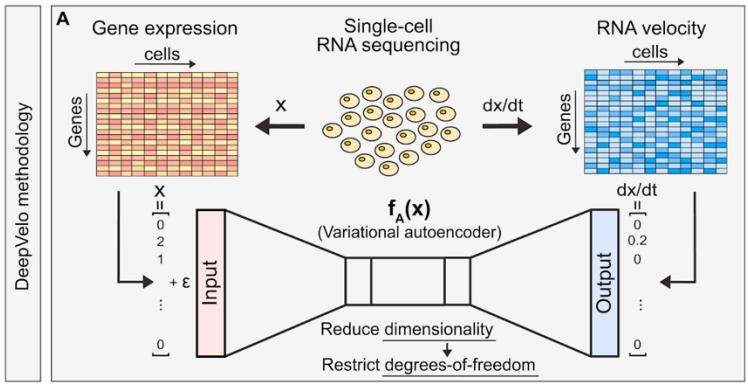
Gerstein lab experience in analyzing chromatin changes and regulatory network rewiring over developmental time-courses



Ordinary Differential Equation (ODE)-based framework for transcriptomic dynamics.

We recently developed *DeepVelo*[1] (**Fig. 1**), a neural network–based framework for modeling the full dynamics underlying scRNA-seq experiments, which typically only capture a static snapshot of gene expression. *DeepVelo* embeds a variational autoencoder with ODEs to capture the continuous transcriptional dynamics within individual cells, allowing for the analysis of complex nonlinear gene interactions in a regulatory cascade. Importantly, by assembling cells from different developmental stages, *DeepVelo* can accurately predict cell states further into the future. We initially tested *DeepVelo* on mouse pancreatic endocrinogenesis, illustrating the robustness and general validity of the approach. Then, we applied it to decipher the gene expression dynamics within the developing mouse brain, specifically in the dentate gyrus and neocortex. The samples represent scRNA-seq experiments from different tissues, technical platforms, and developmental time scales. This work showcased *DeepVelo*'s ability to deconvolve gene co-expression networks on two additional data sources (mouse gastrulation, developing human forebrain) and benchmarked it against linear and state-of-the-art vector-field learning approaches on out-of-sample velocity prediction accuracy.

Machine learning for predicting GRNs from single-cell omics datasets.

We have developed a set of machine learning approaches and bioinformatics tools to model and analyze large-scale multi-modal data for understanding functional genomics and gene regulation. The Gerstein lab led a capstone project in PsychENCODE that integrated and analyzed multi-modal datasets for human brain tissue (dlPFC) from 1,866 individuals[2]. We have also integrated scRNA-seq data for 18,025 cells from PsychENCODE and 14,012 from outside sources. Using these multiomics data (ChIP-seq, RNA-seq, and Hi-C), we identified ~79,000 brain-active enhancers and ~2.5M expression quantitative trait loci (eQTLs).

Network rewiring.

We have leveraged our extensive experience in large-scale integrative analysis across various consortia, such as ENCODE, modENCODE, 1000 Genomes, KBase, and BrainSpan, to construct comprehensive regulatory networks for humans and model organisms based on the ENCODE and modENCODE datasets. These networks, integrating TF-gene, TF-miRNA, and miRNA-gene interactions, reveal intricate statistical patterns and distinct binding preferences, especially in humans, where a notable propensity for distal binding is observed, likely due to the expansive intergenic spaces in the human genome. Furthermore, through our novel cross-species multi-layer network framework *OrthoClust*[3], we have analyzed co-expression networks in an integrated manner, utilizing orthology relationships between species from RNA-seq data generated by ENCODE and modENCODE consortia. This approach has uncovered conserved modules across humans, worms, and flies, crucial for understanding developmental processes. Additionally, we introduced a framework to quantify differences between networks. By comparing matching networks across organisms, we have established a consistent ordering of rewiring rates among various network types. This work highlights the dynamics of regulatory networks and provides valuable insights into the evolutionary aspects of network rewiring.

REFERENCES

1. Chen Z, King WC, Hwang A, Gerstein M, Zhang J. DeepVelo: Single-cell transcriptomic deep velocity field learning with neural ordinary differential equations. *Sci Adv*. 2022;8(48):eabq3745. doi:10.1126/sciadv.abq3745

2. Wang D, Liu S, Warrell J, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science*. 2018;362(6420). doi:10.1126/science.aat8464

3. Yan KK, Wang D, Rozowsky J, Zheng H, Cheng C, Gerstein M. OrthoClust: an orthology-based network framework for clustering data across multiple species. *Genome Biol*. 2014;15(8):R100. doi:10.1186/gb-2014-15-8-r100