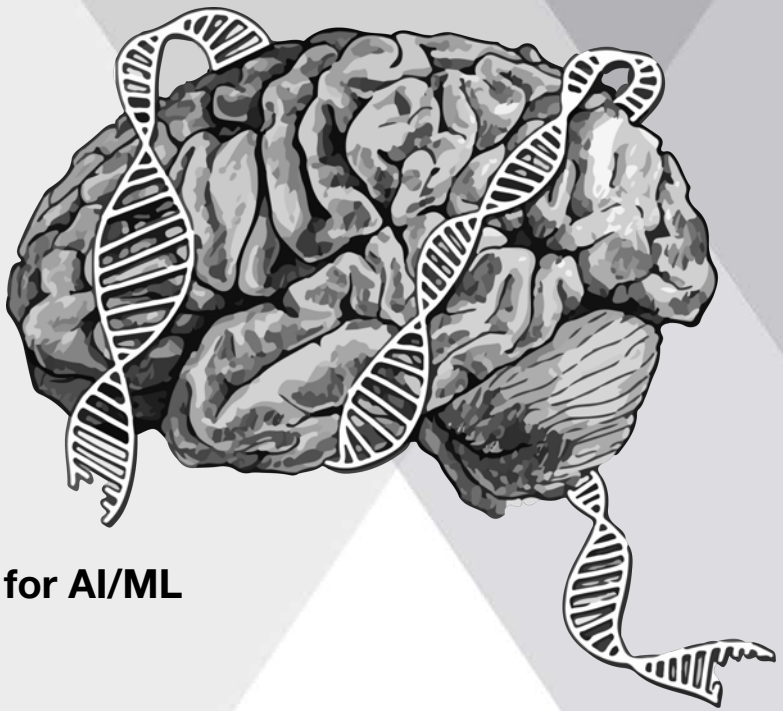


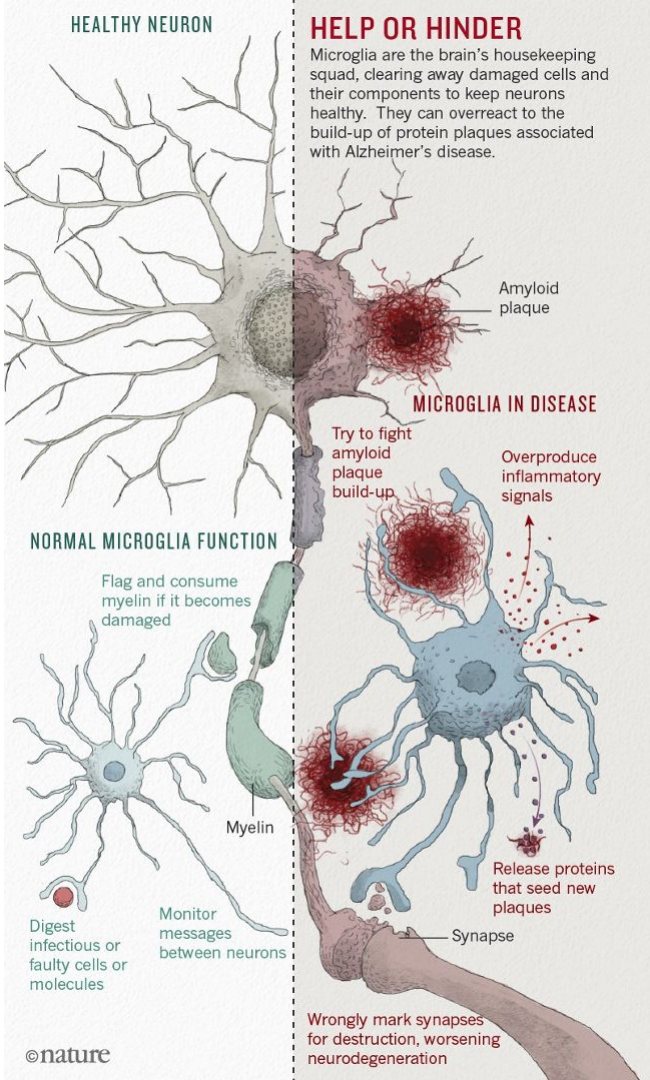
Multi-omic Human Brain Immune Cell (HBIC) Resources for AI/ML Applications

Donghoon Lee, Ph.D.
Assistant Professor
Department of Psychiatry & Genetics and Genomic Sciences
Icahn School of Medicine at Mount Sinai, NY

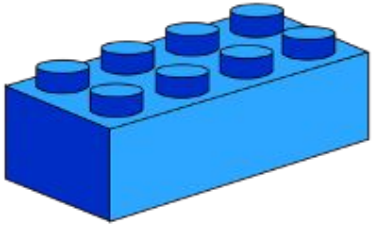
Icahn School of Medicine at Mount Sinai

Human Brain Immune Cell (HBIC) Resource

- ➔ Microglia and myeloid origin cells, collectively known as **human brain immune cells** (HBICs), are implicated in pathogenesis of various neurodegenerative disease, including Alzheimer's Disease (AD).
- ➔ Common and rare AD risk loci affect genes that are preferentially or selectively expressed in HBICs.
- ➔ To characterize disease mechanism, we generated one of the largest single-cell **HBIC** dataset to date to capture the full spectrum of molecular signatures.
- ➔ We propose this multi-omic single-cell molecular profiles will be critical input for AI/ML applications to accelerate research about the etiopathogenetic mechanisms of immune cells on AD.



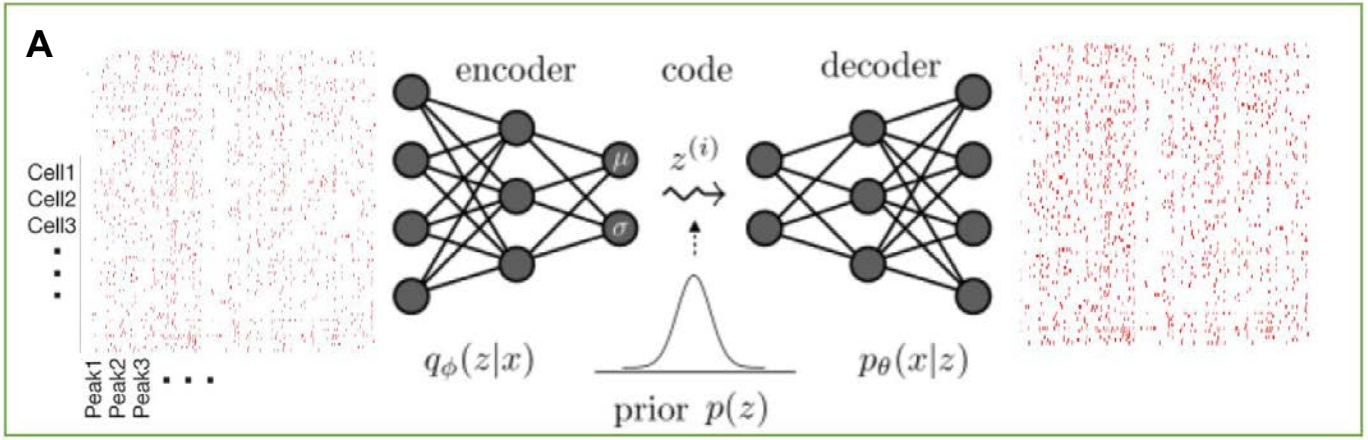
Building block 1: variational autoencoder

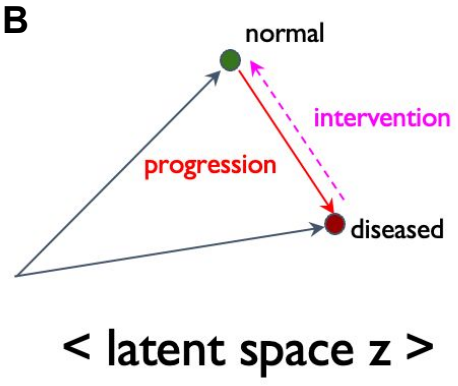


encoder maps input x into latent space z

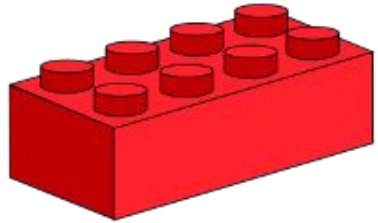
decoder infer x from latent space z using prior

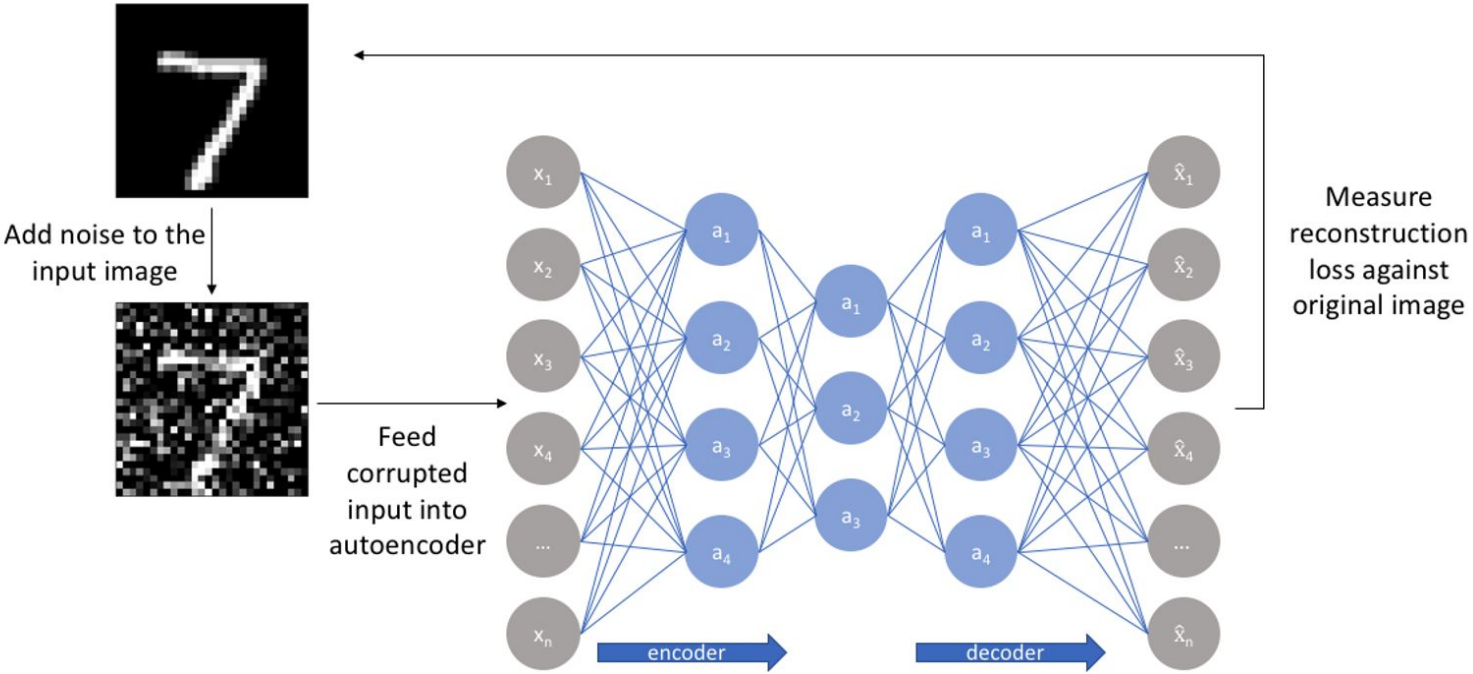
variational inference allow us to estimate prob. density of latent space z into a “continuous” manifold



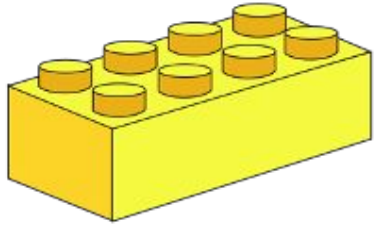


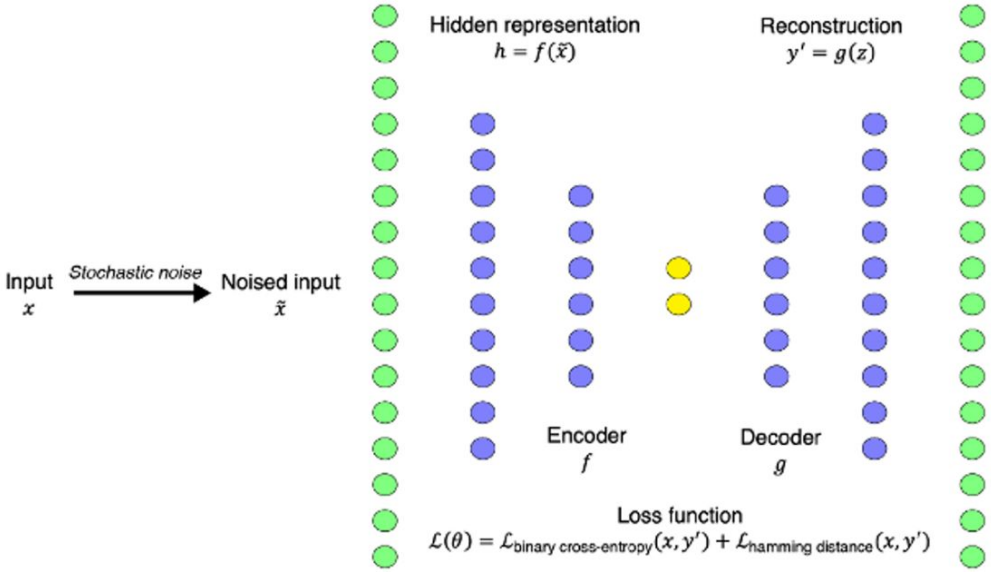
Building block 2: denoising autoencoder





Building block 3: modified loss function to reward sparsity





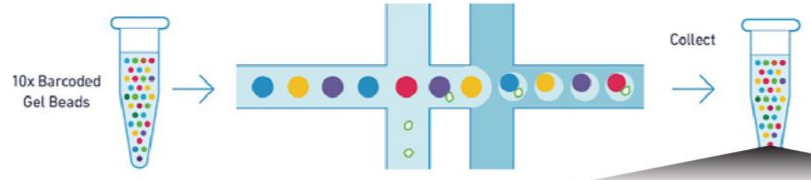
Encoder $z = q_\theta(z|\tilde{x})$

Decoder $\hat{y} = p_\emptyset(\tilde{x}|z)$

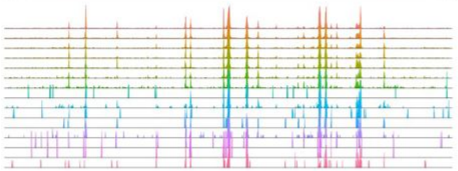
$$\mathcal{L}_{\theta,\emptyset} = -\mathrm{E}_{q(z|x)}[\log p_\emptyset(\tilde{x}|z)] - KL[q(z|\tilde{x})||p(z)] + \mathcal{L}_{Hamming}$$

Continuous approx. of hamming distance

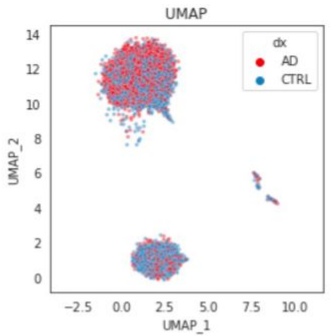
$$\mathcal{L}_{Hamming} = \sum y(1-\hat{y}) + (1-y)\hat{y}$$

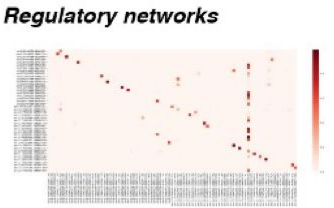


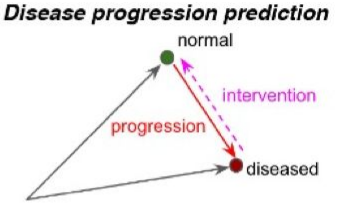
Autopsy/Biopsy HBICs from PFC
scRNA-seq (N=82)
scATAC-seq (N=38)

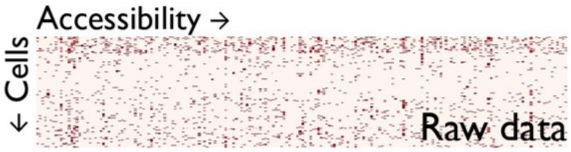


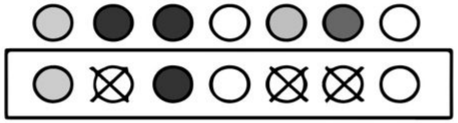
Trained on 124k obs (cells)





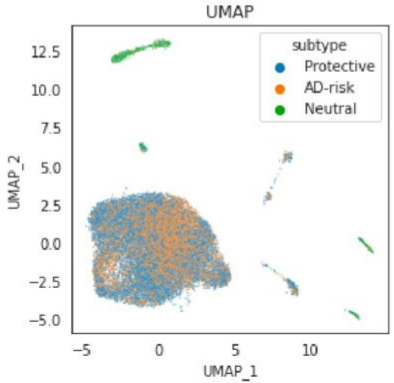






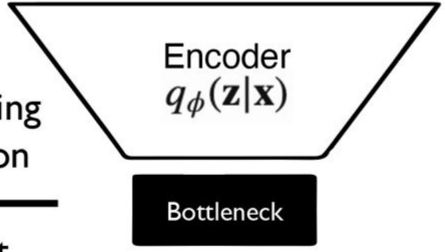
$\mathbf{x}$

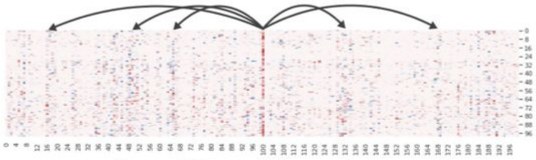
$\mathbf{x}'$ AD-risk signature subtyping



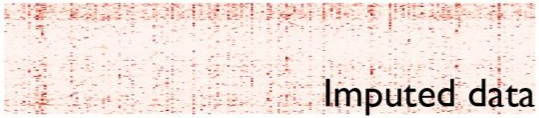
Clustering based on

Latent Embedding z

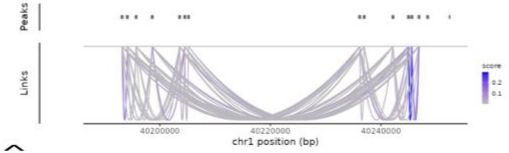




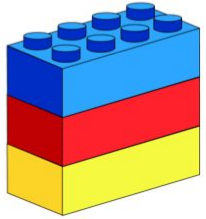
In-sillico Knockout



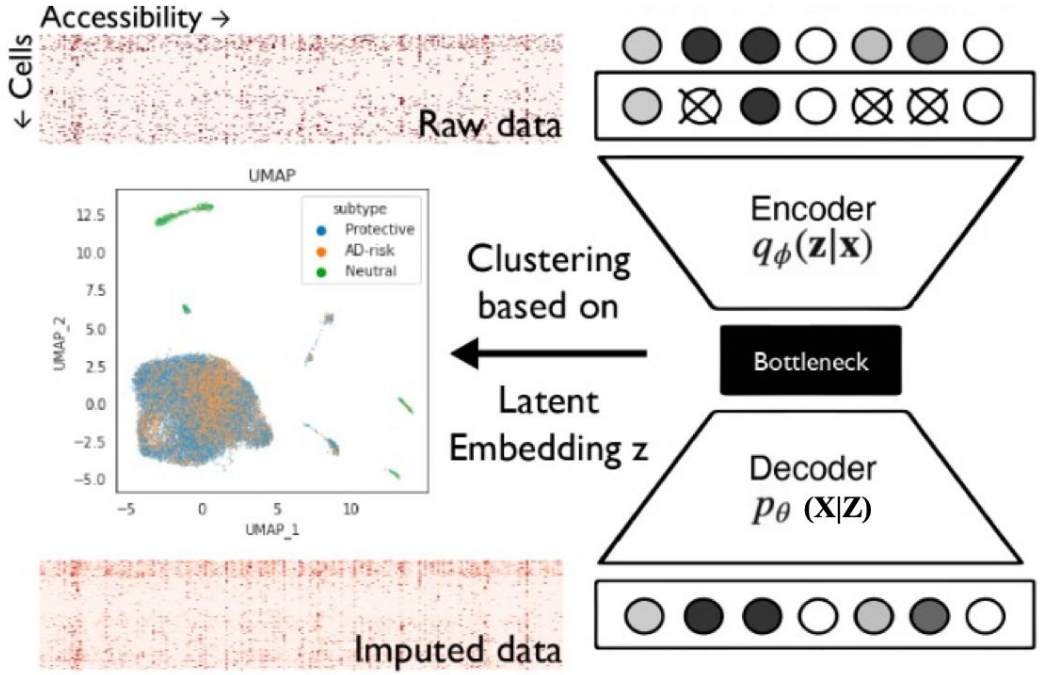
$\hat{\mathbf{x}}$



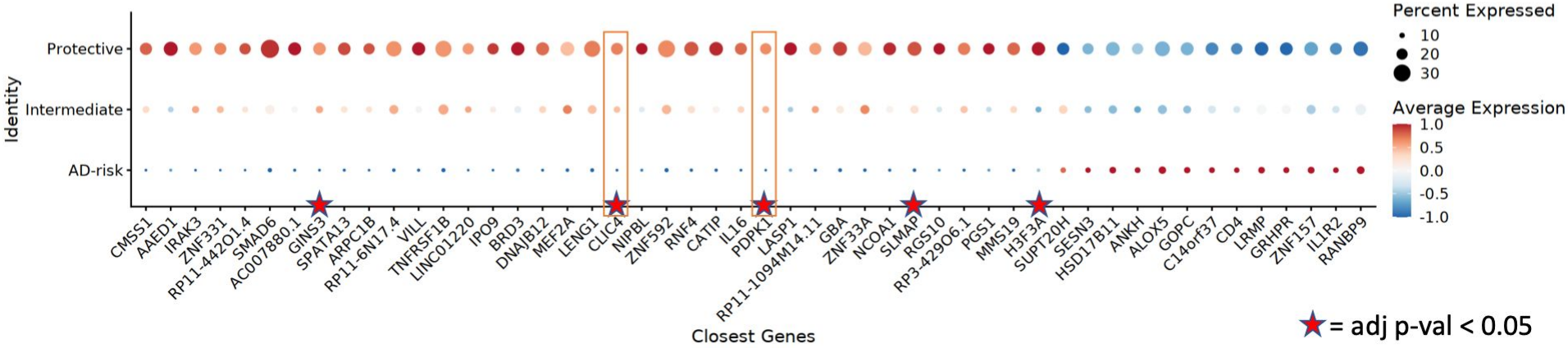
Co-accessible Network



Example 1. AD-risk subtyping

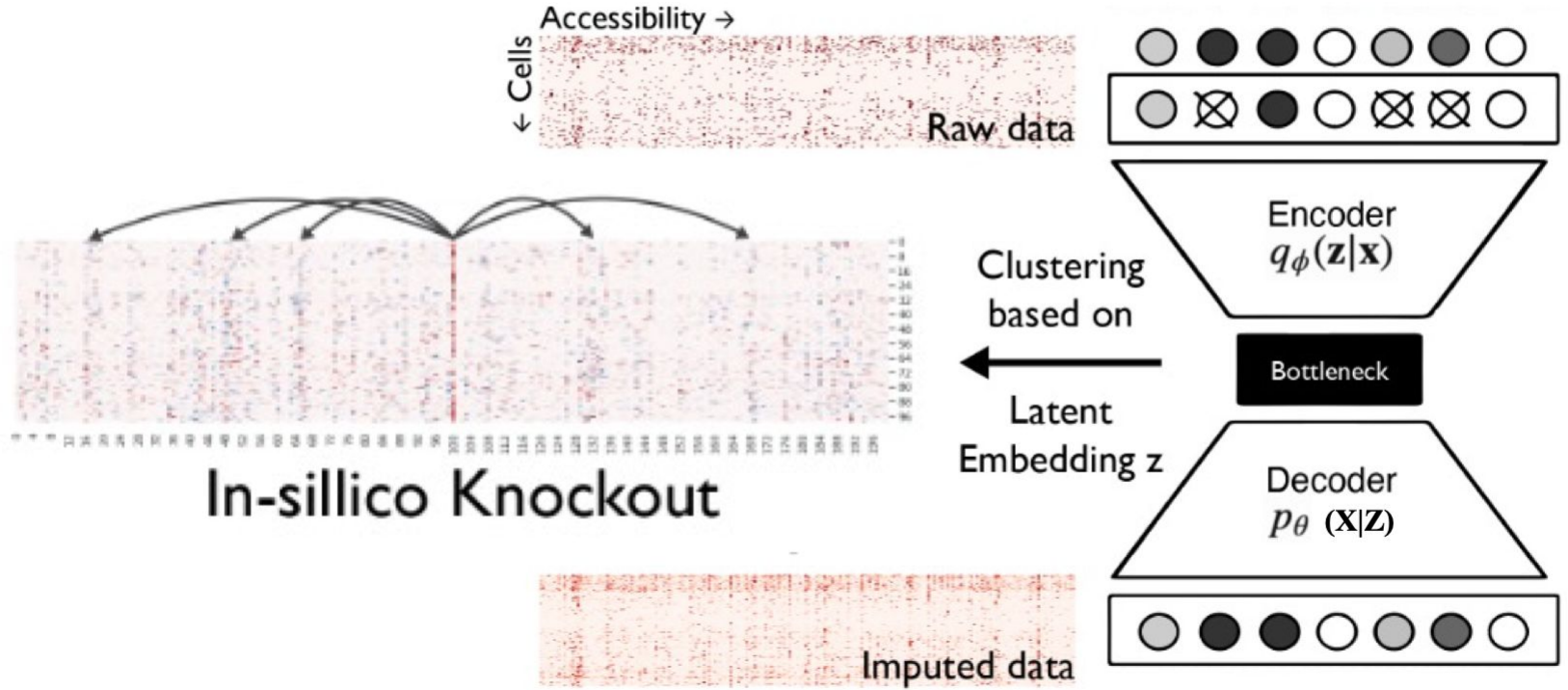


Find differentially accessible peaks across AD-risk score

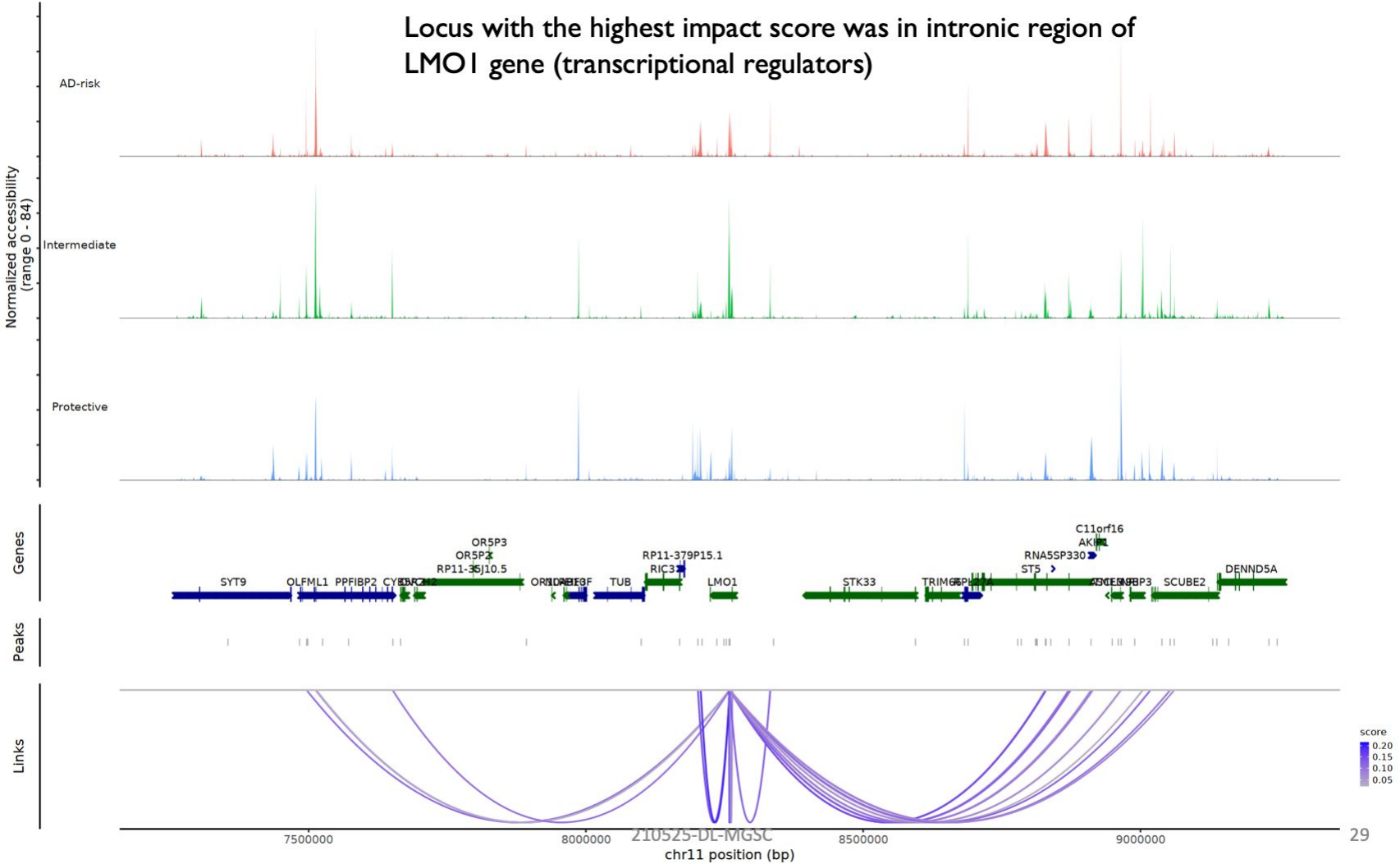


- **CLIC4** (chloride channel protein) primes signal for cytokine IL-1β and in inflammasome NLRP3 activation. Dysregulation of the NLRP3 inflammasome is recognized as the common feature of chronic inflammatory diseases such AD
- **PDPK1** (phosphoinositide-dependent protein kinase), when PDK1 suppresses the activity of ADAM-17, α-secretase, toxic Aβ fragments accumulate, further activating PDK1

Example 2. Building cis-co-accessible networks



Locus with the highest impact score was in intronic region of LMO1 gene (transcriptional regulators)



MSSM
Panos Roussos (PI)
Donghoon Lee

Funding
R01AG065582
R01AG067025
R01AG050986

Thank You