Multi-modal data analysis

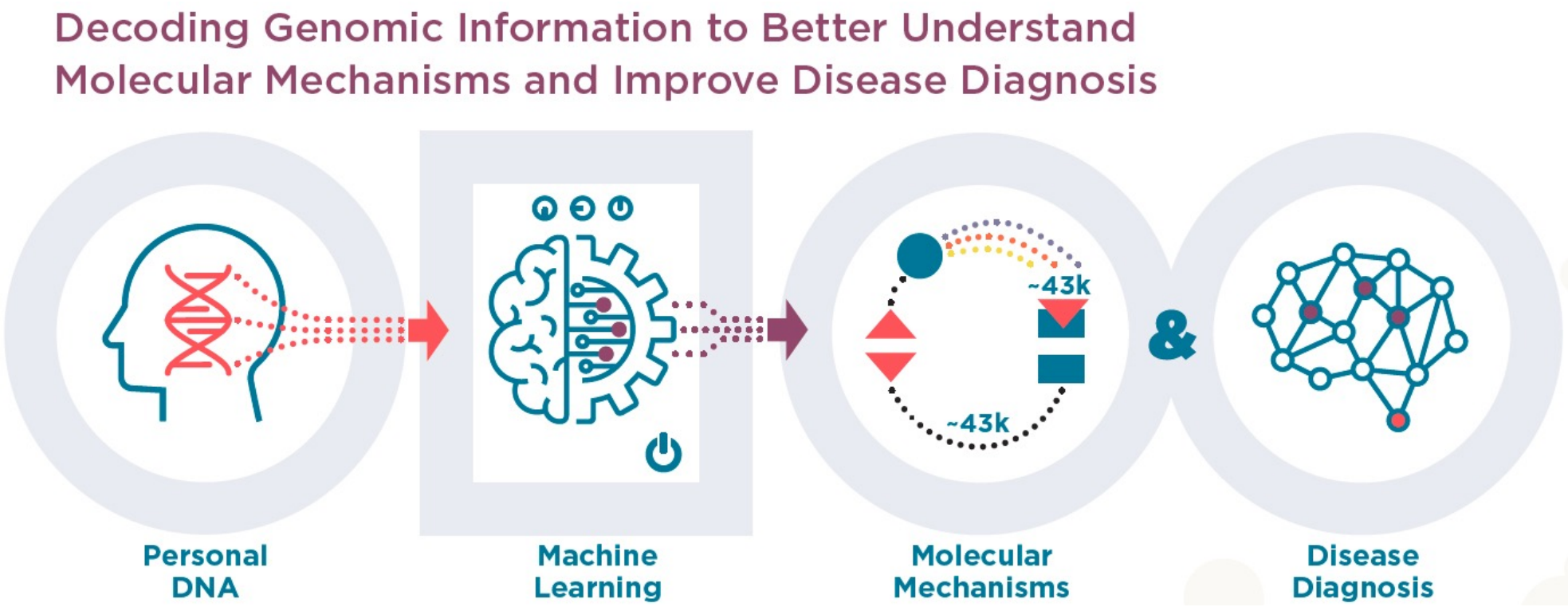
- with an application to brain functional genomics

Daifeng Wang, Ph.D.
Assistant Professor
Department of Biostatistics and Medical Informatics
Department of Computer Sciences (affiliate)
Investigator, Waisman Center
University of Wisconsin – Madison
daifeng.wang@wisc.edu
https://daifengwanglab.org/

BMI/CS 776
Spring 2022

Machine learning to open "black box" from genotype to phenotype

- Goal
  - *Advance biological knowledge on genomics in brain diseases*
- Approach
  - *Machine learning, Bioinformatics, Computational Biology*



https://www.waisman.wisc.edu/2020/01/07/new-researcher-uses-machine-learning-to-decode-genomic-information/

Outline

- Background
  - Functional genomics & multi-omics
  - Gene regulation

- Multi-view learning for multi-modal data
  - Multi-view Empirical Risk Minimization (MV-ERM)
  - Manifold alignment

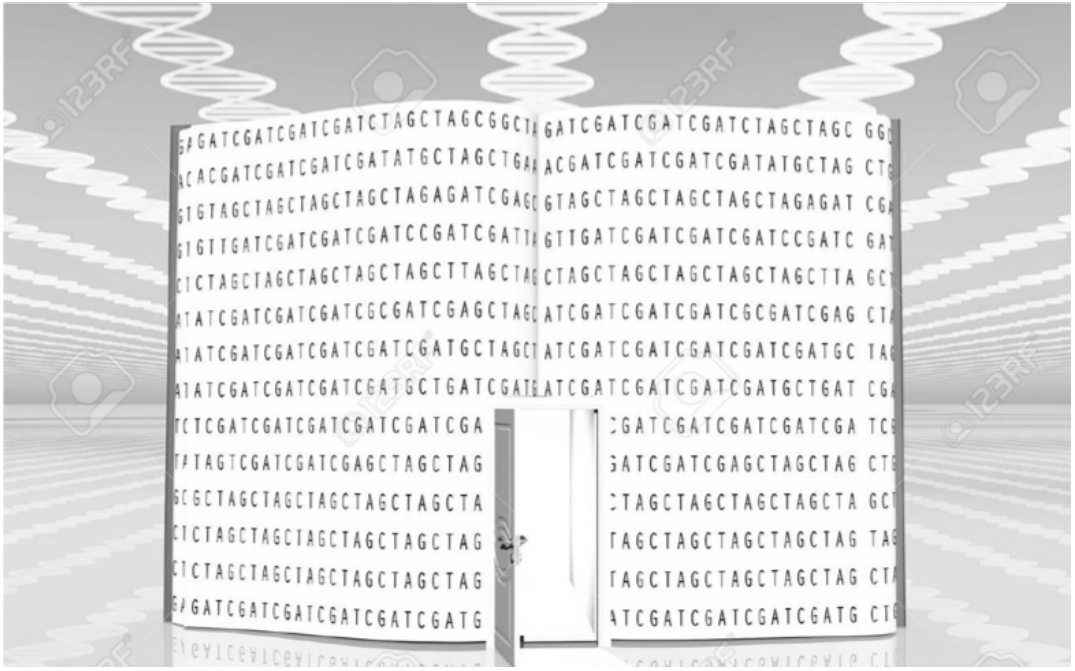
- App 1: *single-cell multi-modal data integration*
  - Predicting neuronal electrophysiology from gene expression
- App 2: *deep manifold-regularized classification*
  - Predicting cortical layers of neurons from multi-modalities
- App 3: *biologically interpretable neural network modeling*
  - Disease prediction and functional prioritization

Your genome is your genetic code book

Book	Genome
Chapters	Chromosomes
Sentences	Genes
Words	Elements
Letters	Bases

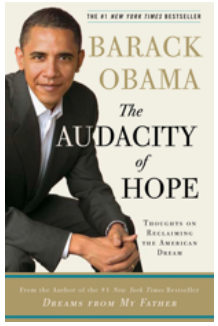
Human

- 46 chromosomes
- ~ 20,000 – 25,000 genes
- ~ Millions elements
- 4 unique bases (A, T, C, G), ~3 billion in total



https://goo.gl/images/vMaz4T

How to read your genetic code book?



Chapter One

Republicans and Democrats

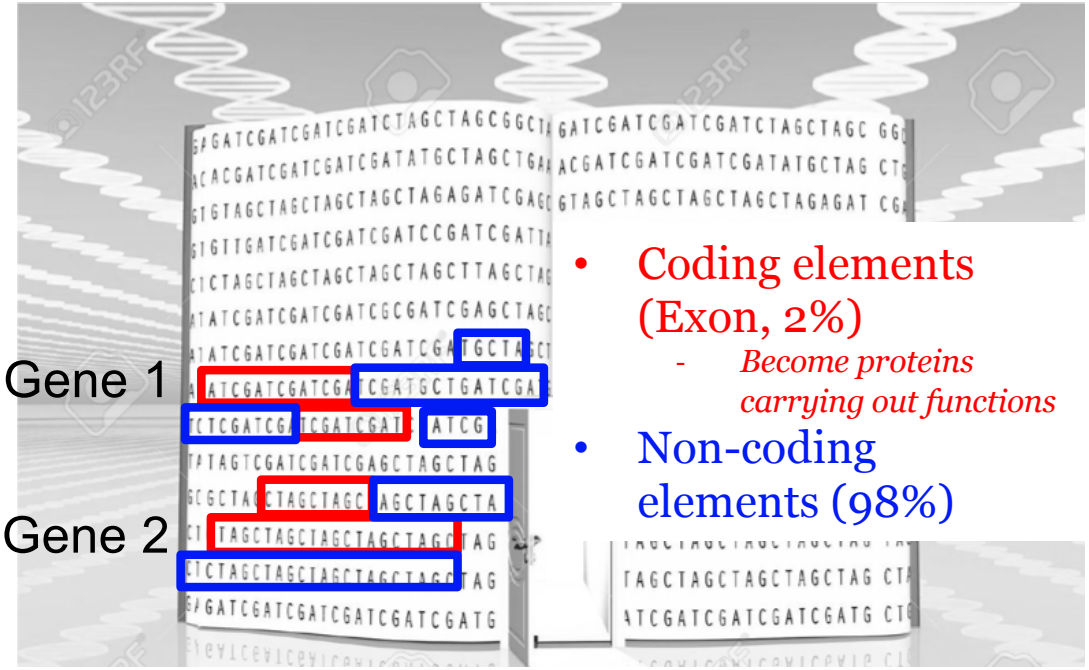
Book	Genome
Chapters	Chromosomes
Sentences	**Genes**
Words	**Elements**
Letters	Bases

"On most days, I enter the Capitol through the basement. A small subway train carries me from the Hart Building, where …"

- Key words
- Non-key words

Overhead, the ceiling forms a creamy white oval, with an American eagle etched in its center. Above the visitors' gallery, the busts of the nation's first twenty vice presidents sit in solemn repose.

And in gentle steps, one hundred mahogany desks rise from the well of the Senate in four horseshoe-shaped rows. Some of these desks date back to 1819, and atop each desk is a tidy receptacle for inkwells and quills. Open the drawer of any desk, and you will find within the names of the senators who once used it—Taft and Long, Stennis and Kennedy—scratched or penned in the senator's own hand. Sometimes, standing there in

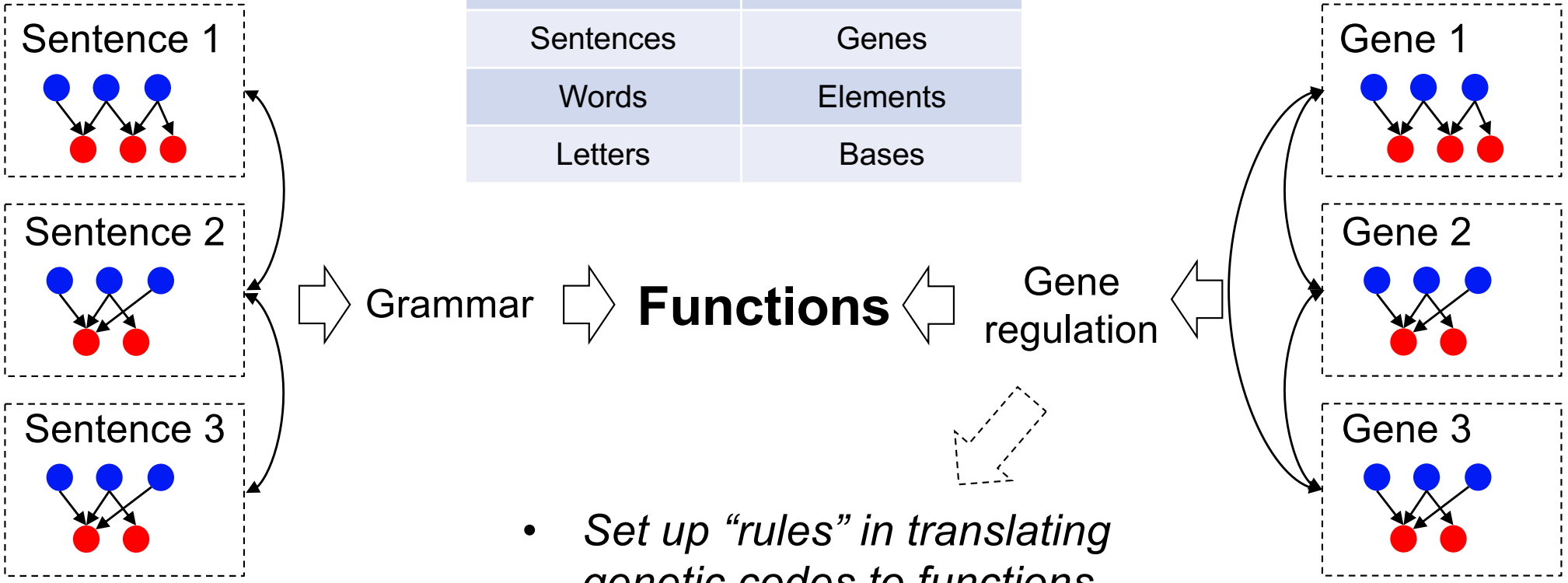


- Coding elements (Exon, 2%)
  - *Become proteins carrying out functions*
- Non-coding elements (98%)

https://goo.gl/images/vMaz4T

Grammar for book is clear but not for genome

Book	Genome
Chapters	Chromosomes
Sentences	Genes
Words	Elements
Letters	Bases

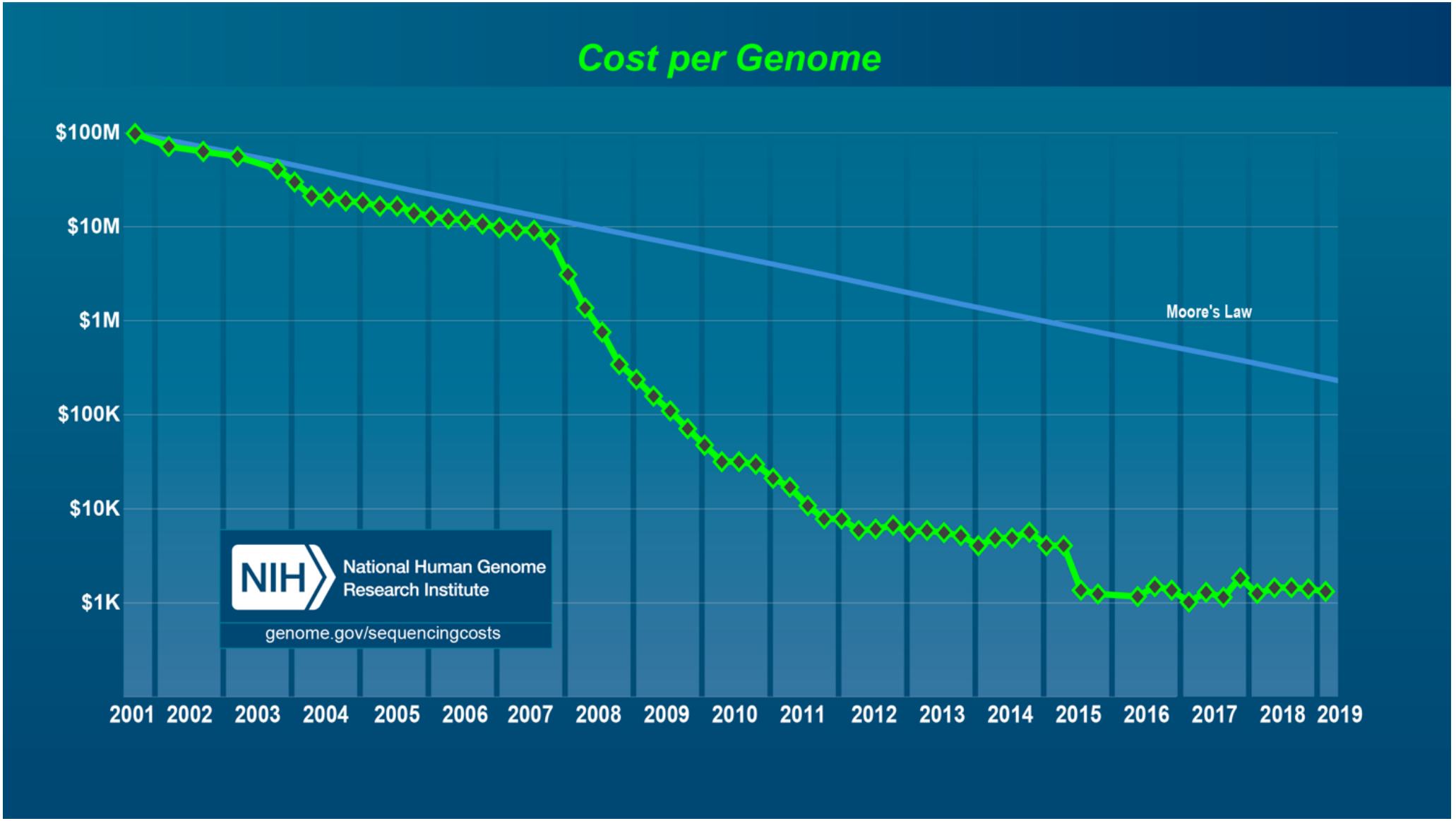


- Key words
- Non-key words

- *Set up "rules" in translating genetic codes to functions*
- *Broken rules -> Abnormal functions*
- *Unclear*

- Coding elements
- Non-coding elements

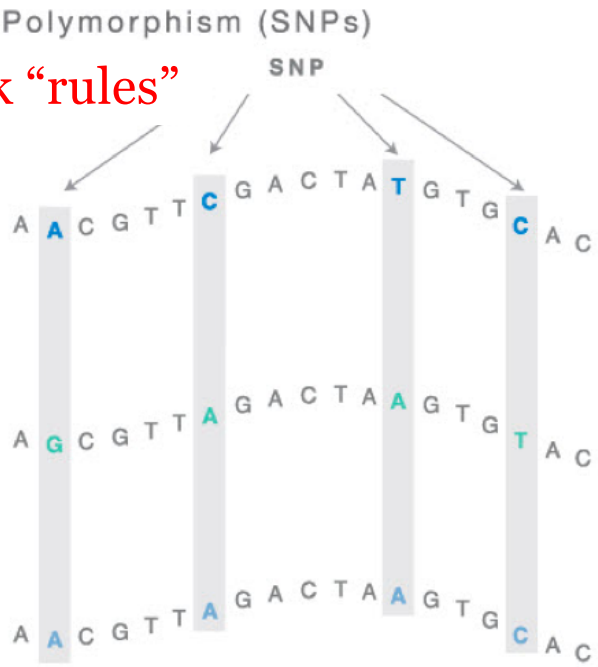
Low sequencing cost enables reading our whole genome



Differences across our genomes: DNA variants

Single Nucleotide Polymorphisms (SNPs) normally happen ~1% on individual human genome.

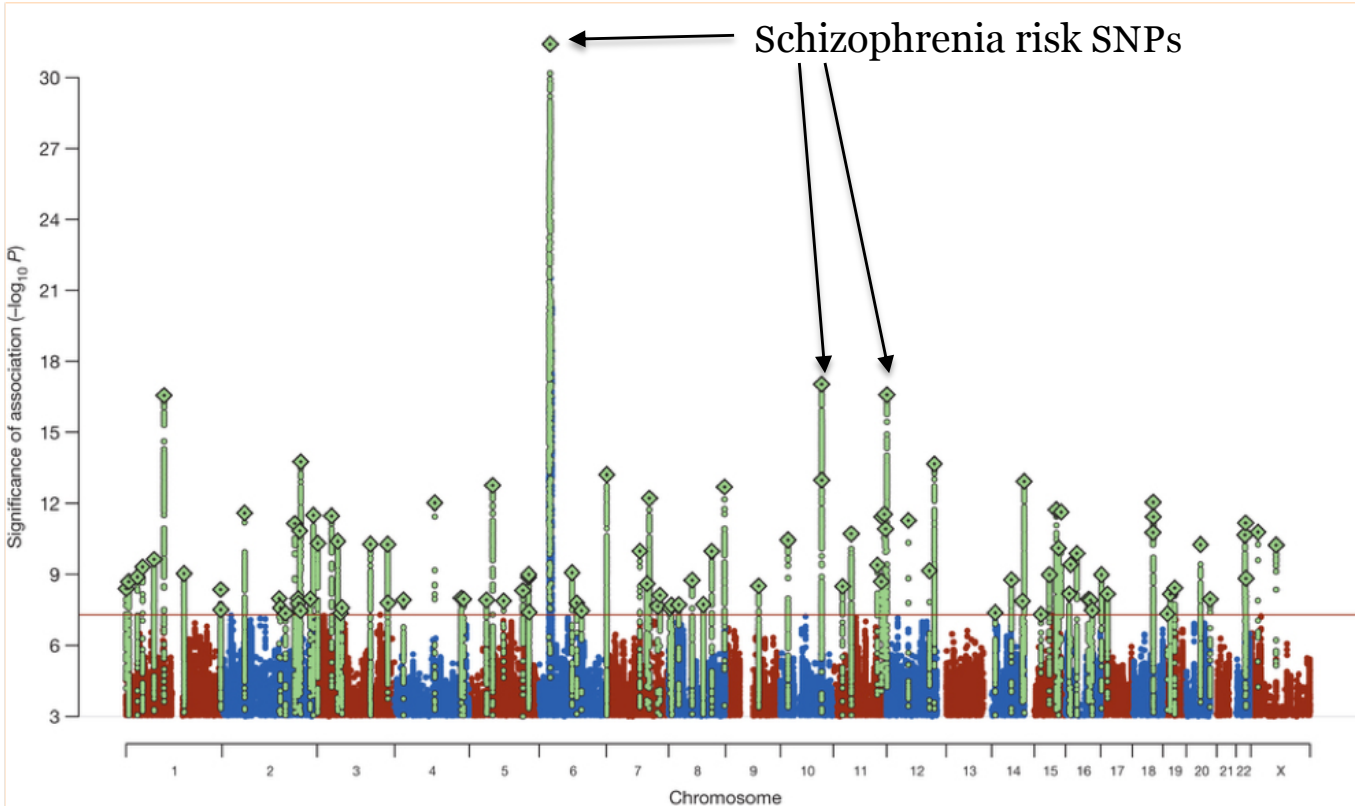
Most SNPs are harmless but some break "rules"



Talking Glossary of Genetic Terms
NATIONAL HUMAN GENOME RESEARCH INSTITUTE
NATIONAL INSTITUTES OF HEALTH | genome.gov
Illustration by Darryl Leja, NHGRI

Genome-Wide Association Study (GWAS) identifies disease associated genetic variants

36,989 schizophrenia cases and 113,075 controls
in Psychiatric Genomics Consortium

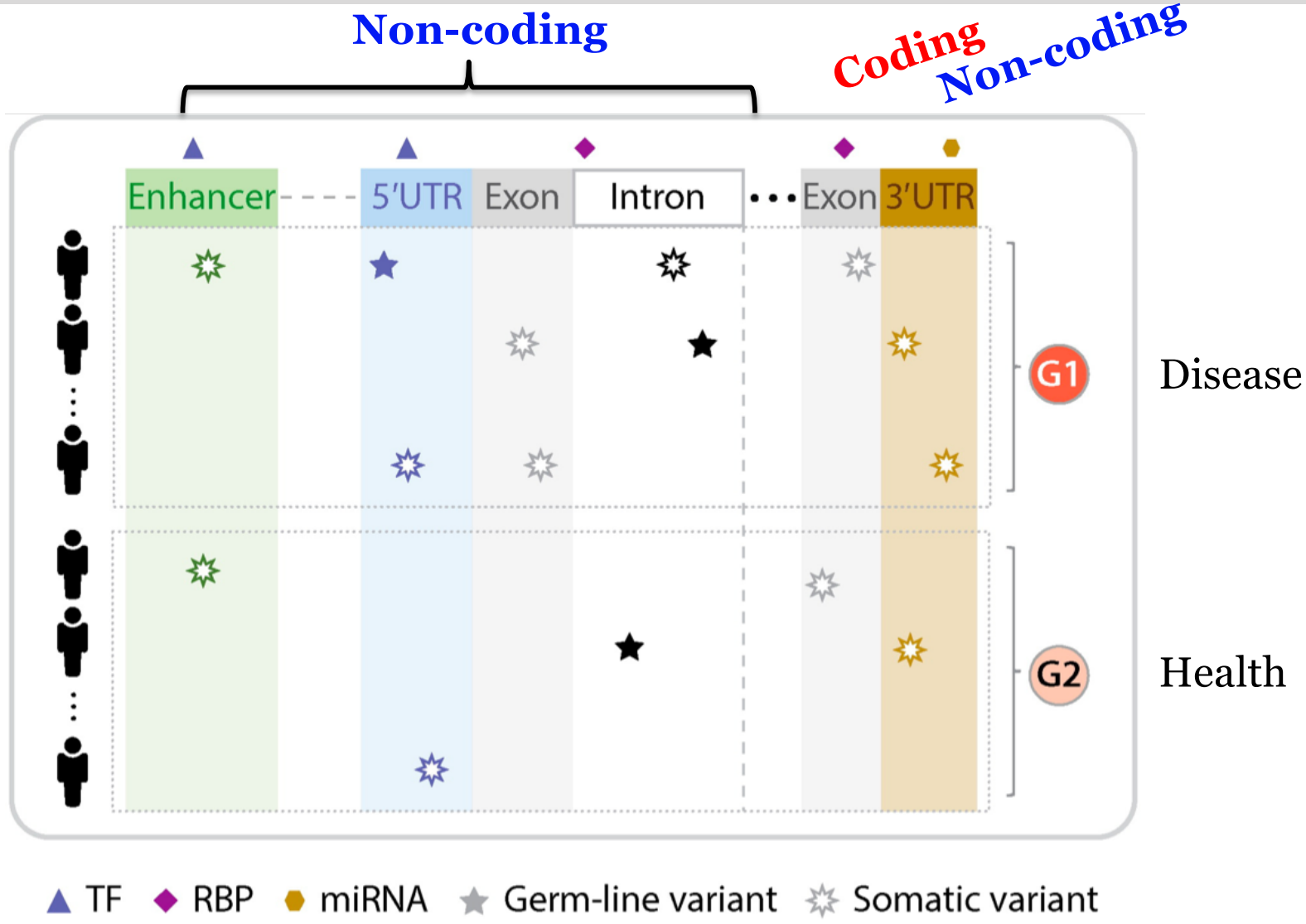


$P=5*10^{-8}$

Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nature (2014)

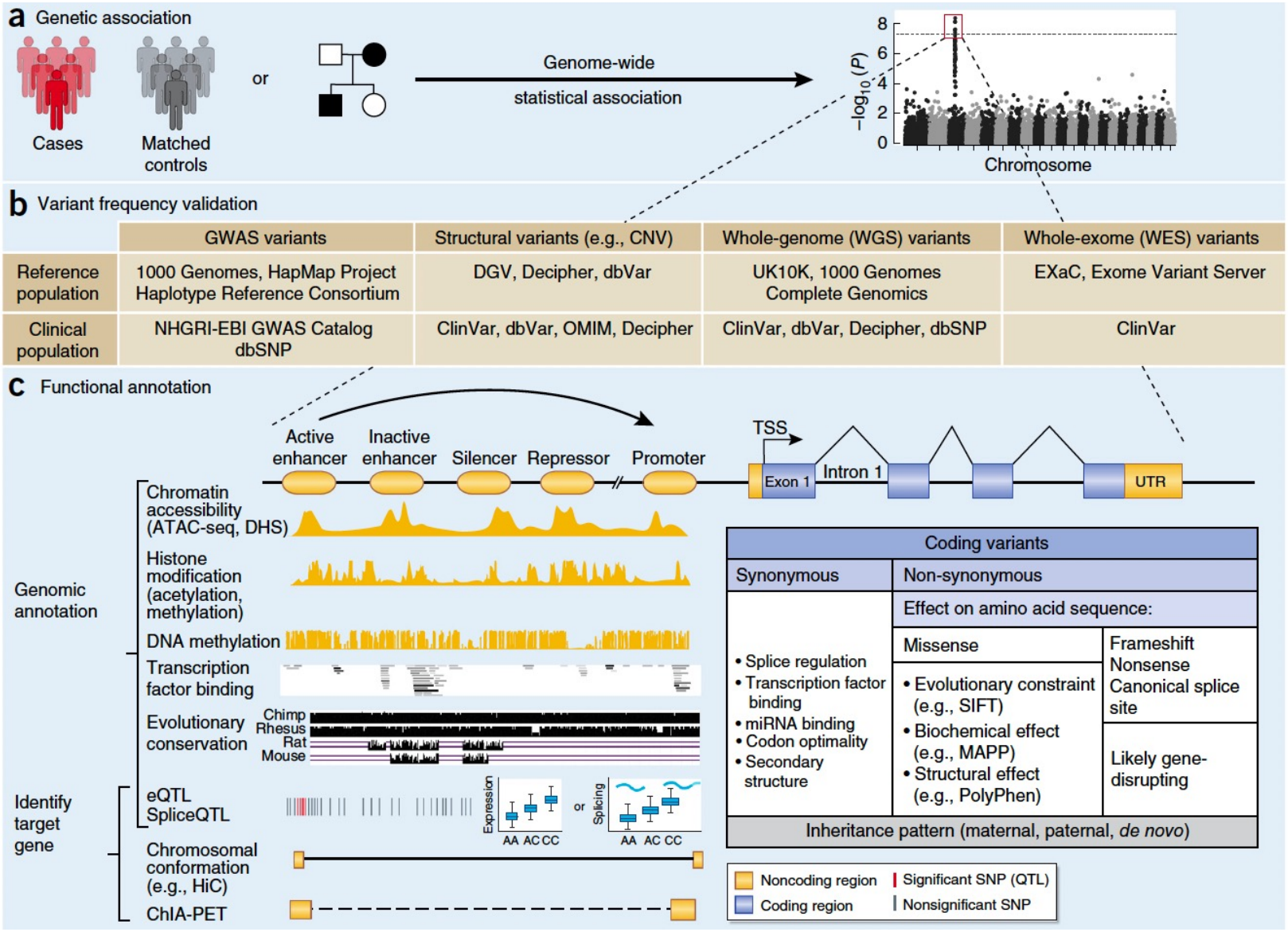
However, genotype-phenotype association can't tell mechanisms

How to link non-coding disease SNPs to genes?



McGillivray P, et al. 2018.
Annu. Rev. Biomed. Data Sci. 1:153–80

Complex mechanisms from genotype to phenotype



Disease-associated genomic variants

How do variants function?

Gandal et al., Nature Neuroscience, 2016

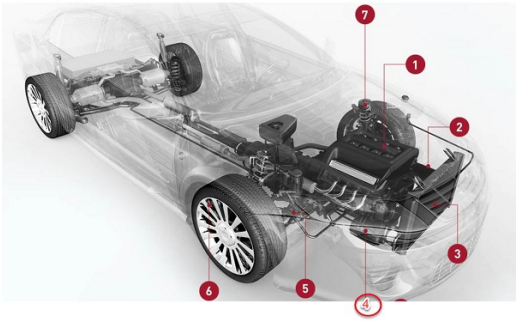
Hierarchical understanding from genotype to phenotype

Functional Elements
- variants
- genes
- regulatory regions

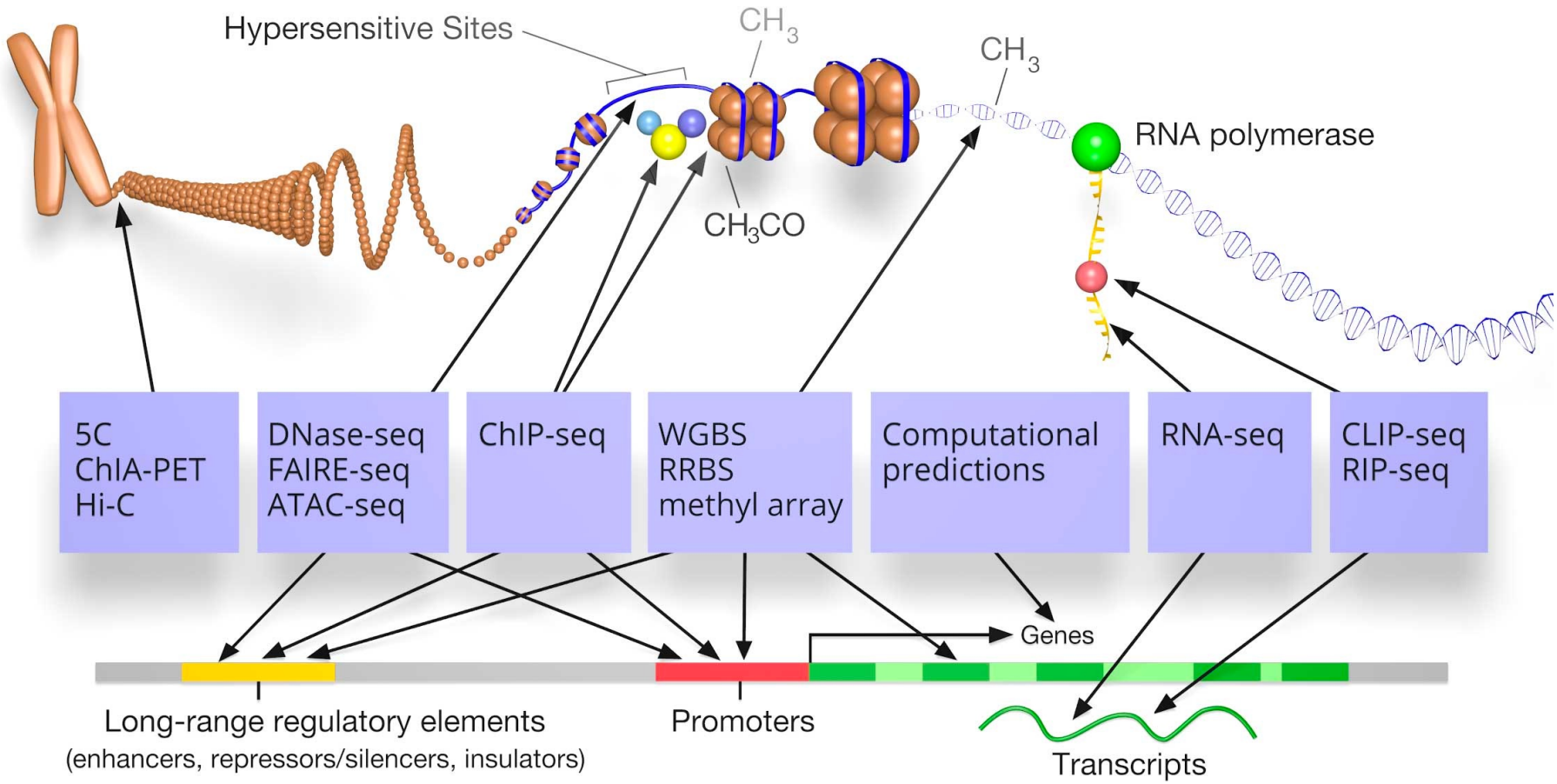
Interactions
- gene regulation
- chromatin interaction
- TF binding

Mechanisms
- pathways
- circuits
- processes

Prediction & Prioritization
- disease variants & genes
- networks
- cell types



Multi-omics data for various functional elements

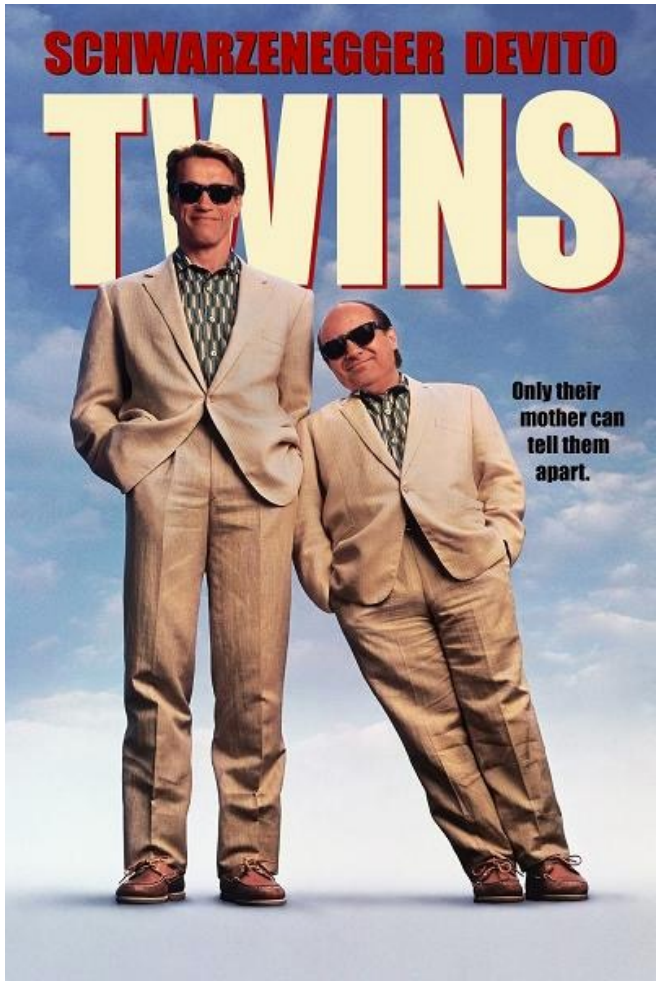


ENCODE

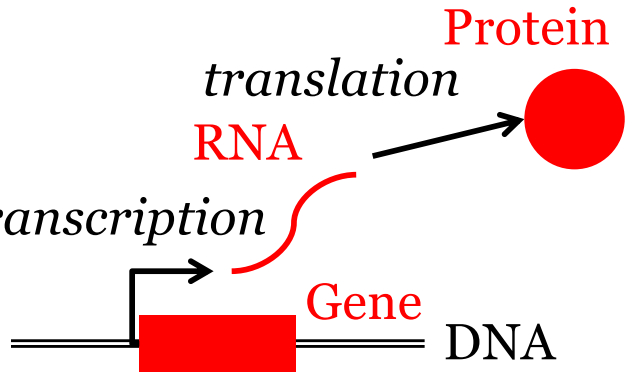
Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

Gene expression and regulation

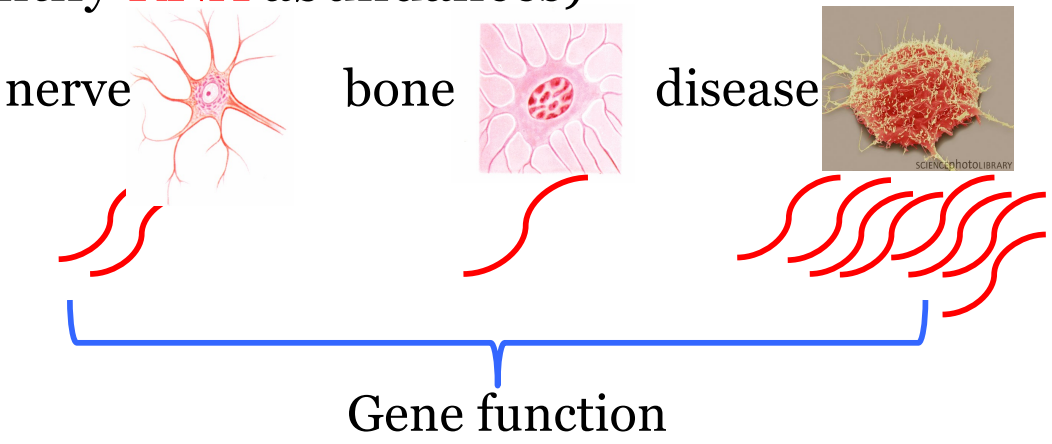
Identical DNA but different gene expression



Central dogma

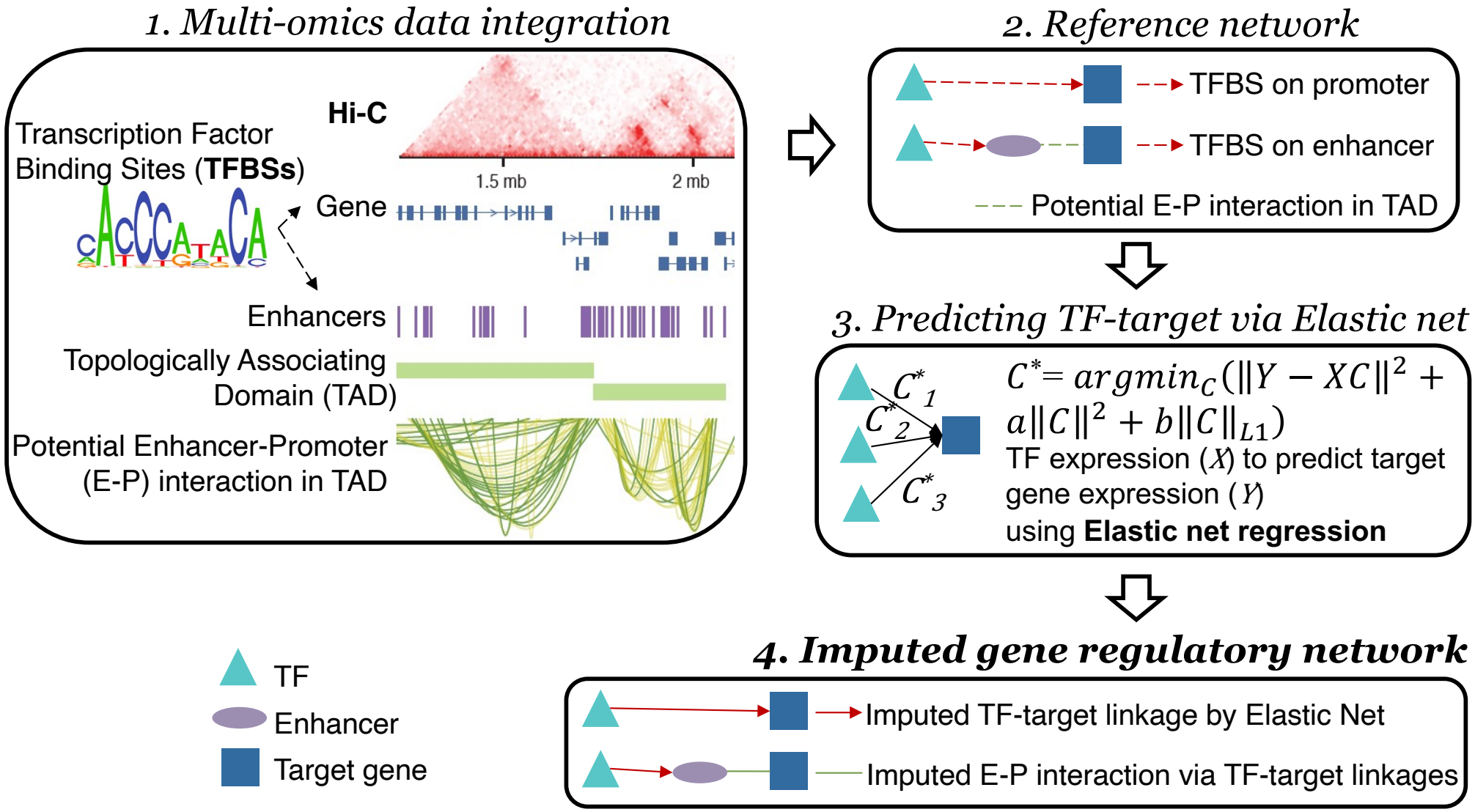


Gene expression levels (e.g., values to quantify RNA abundances)



Gene regulation: mechanisms controlling gene expression levels

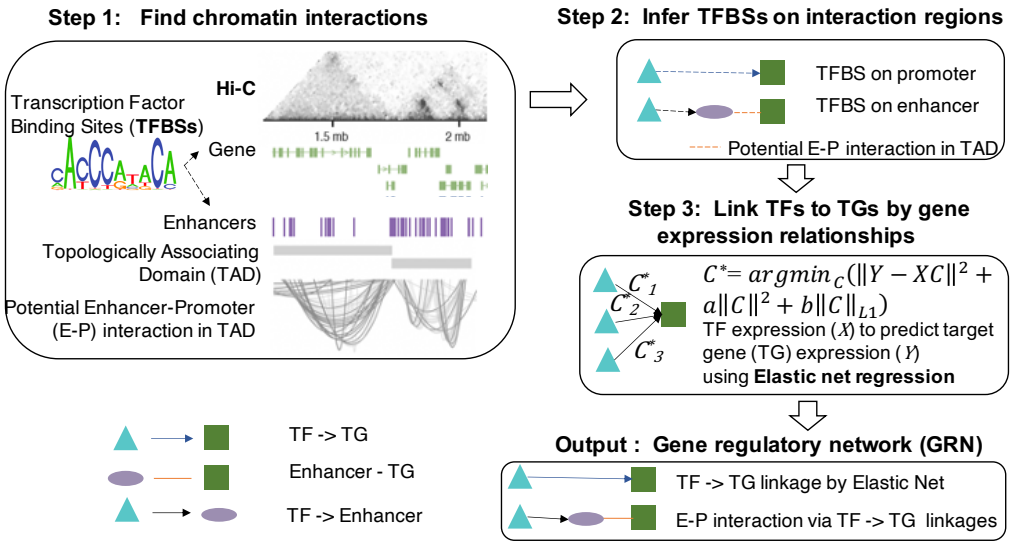
Gene regulatory networks linking functional elements from multi-omics



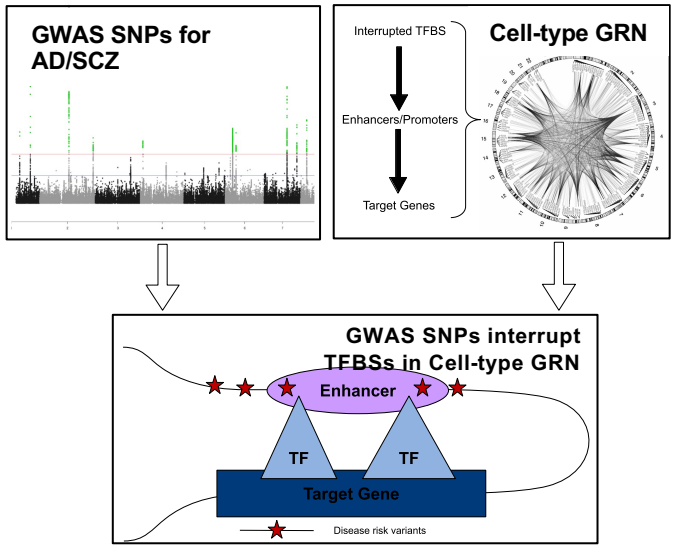
Jin et al, Genome Medicine, 2021

Gene regulatory networks linking GWAS SNPs to disease genes

1. Predicting cell-type gene regulatory networks (GRNs) via multi-omics



2. Identifying cell-type disease genes via GWAS and cell-type GRNs



Output: Cell-type disease genes, networks and functions

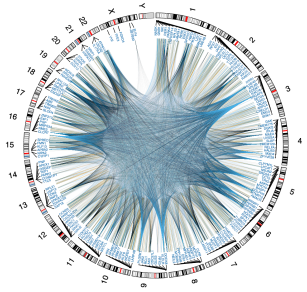
SNP -> TF -> enhancer -> cell-type disease gene -> function

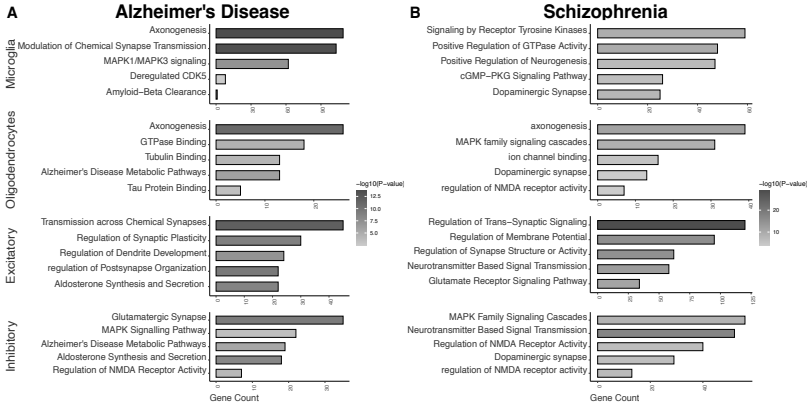
SNP -> TF -> cell-type disease gene -> function

3. Microglia GRN, cell-type disease gene functions

Microglia

4523 TGs, 176 TFs, 12927 enhancers

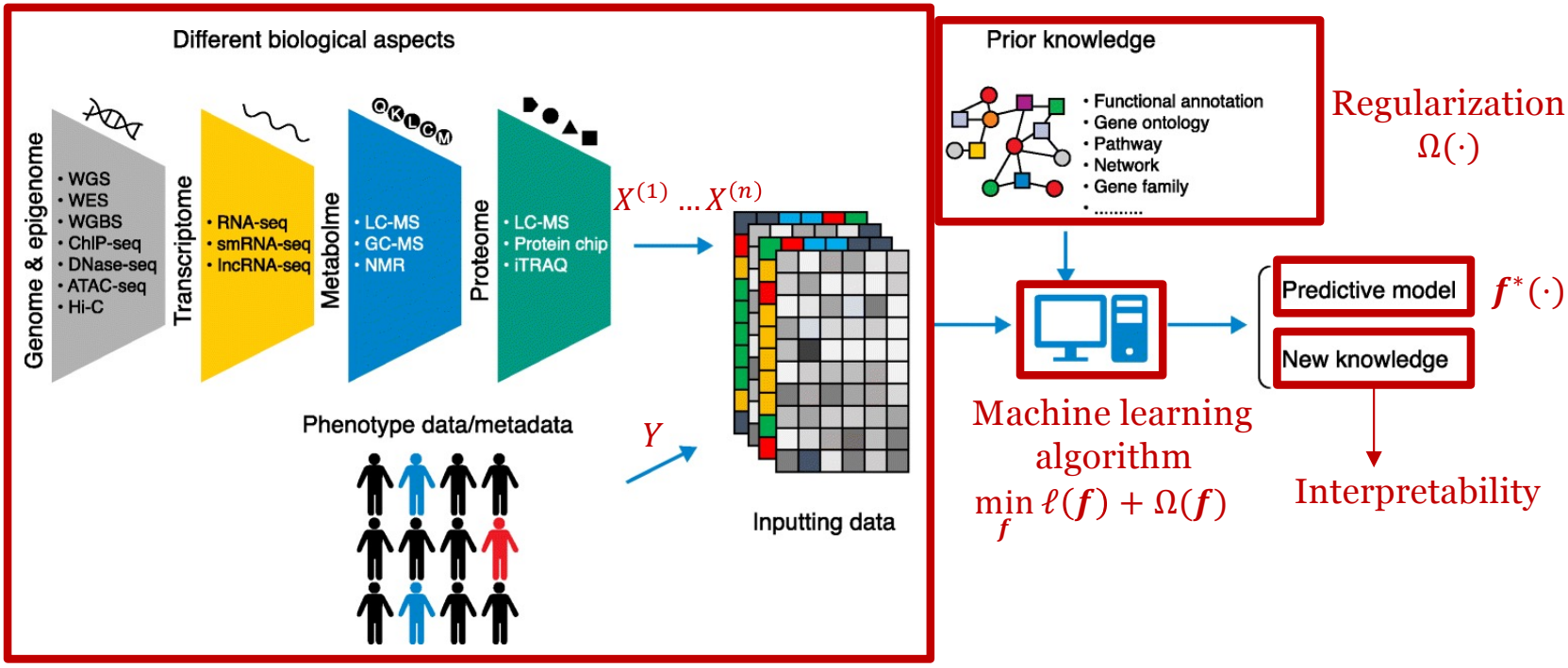




Ting Jin (Ph.D. student, Biomedical Data Science)

Jin et al, Genome Medicine, 2021

Machine learning for multi-omics analysis, prediction and interpretation



Xu. Genome Bio. 2019

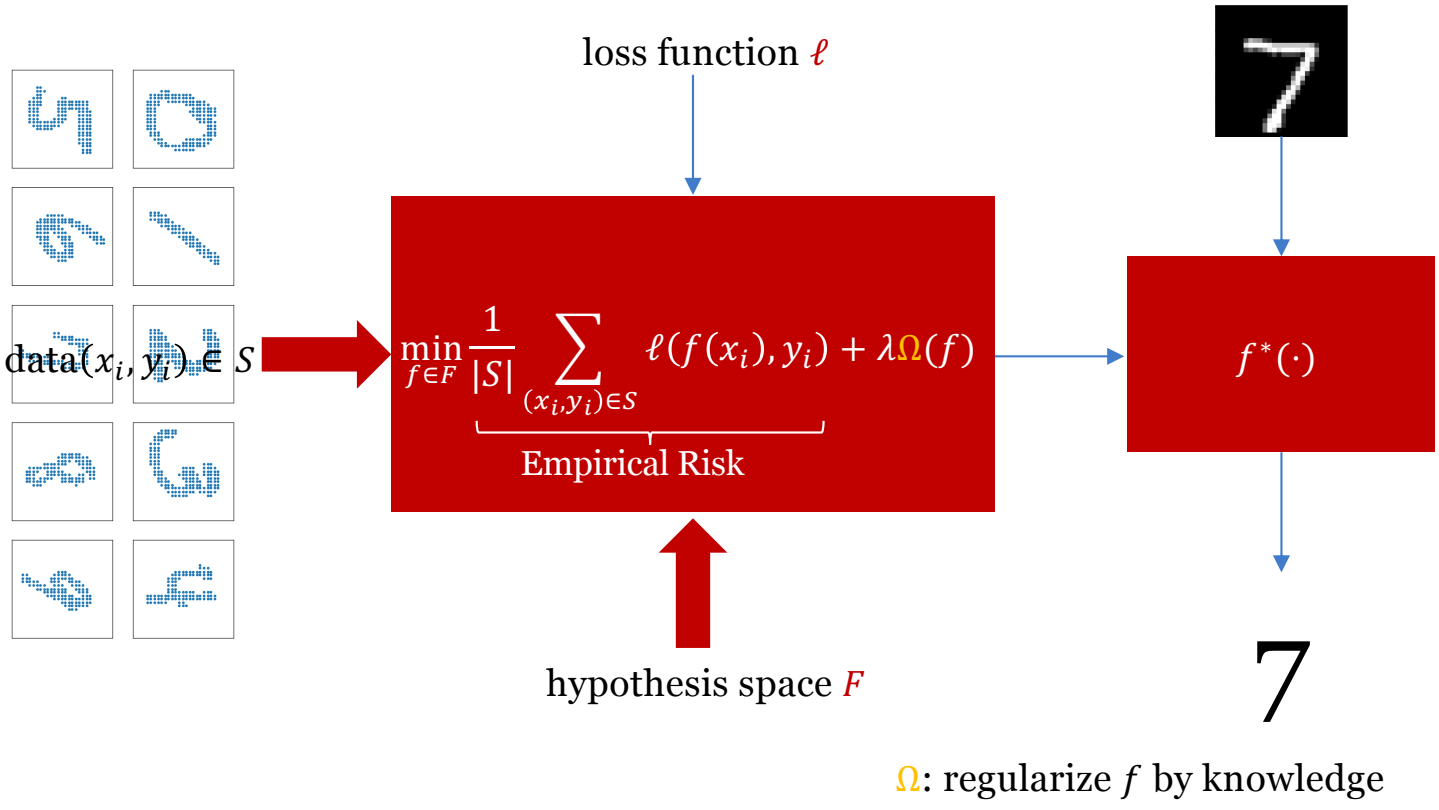
Outline

- Background
  - Functional genomics & multi-omics
  - Gene regulation

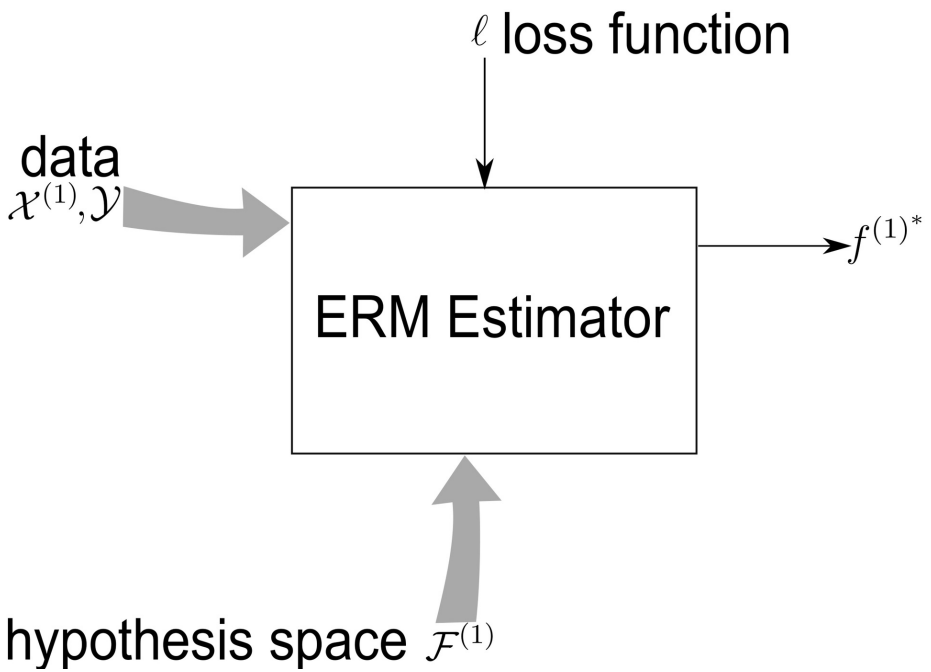
- Multi-view learning for multi-modal data
  - Multi-view Empirical Risk Minimization (MV-ERM)
  - Manifold alignment

- App 1: *single-cell multi-modal data integration*
  - Predicting neuronal electrophysiology from gene expression
- App 2: *deep manifold-regularized classification*
  - Predicting cortical layers of neurons from multi-modalities
- App 3: *biologically interpretable neural network modeling*
  - Disease prediction and functional prioritization

Empirical Risk Minimization (ERM)



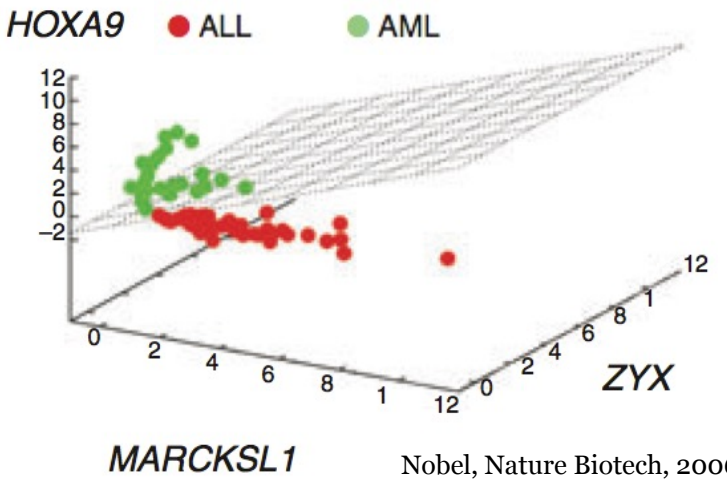
Empirical Risk Minimization (ERM) for single-omics (single-view learning)



$$\mathrm{R}(f) = \frac{1}{|S|} \sum_{(x_i, y_i) \in S} \ell(f(x_i), y_i)$$

$$f^* \in \underset{f}{\operatorname{argmin}} \{\mathrm{R}(f) + \lambda \Omega(f)\}$$

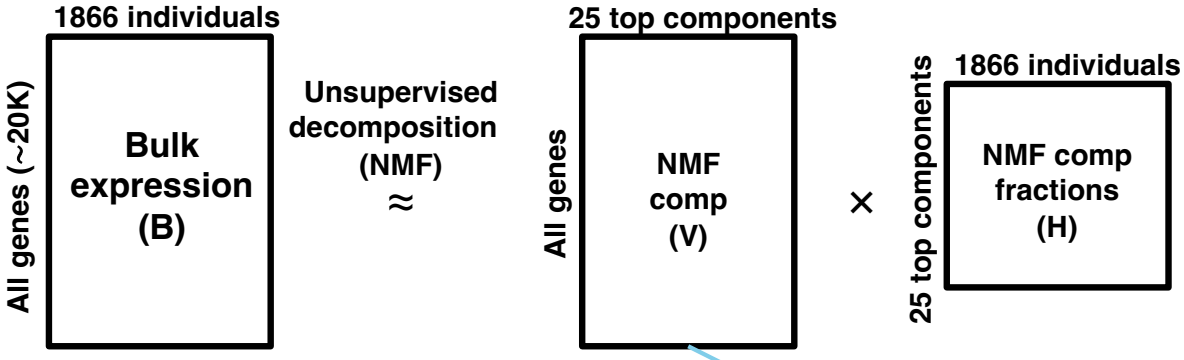
- e.g., Leukemia patient classification
  - y_i: Acute lymphoblastic leukemia (ALL) vs. Acute myeloid leukemia (AML)
  - x_i: gene expression
  - f: SVM

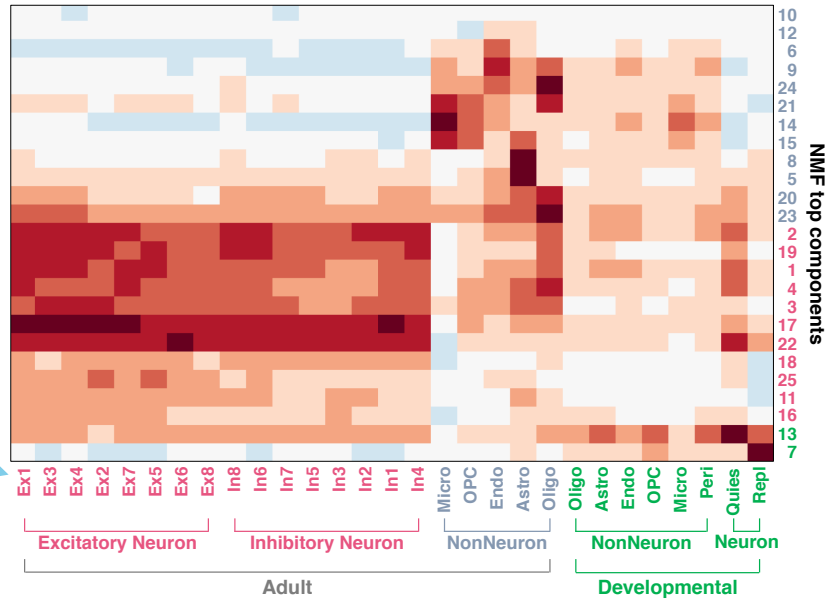


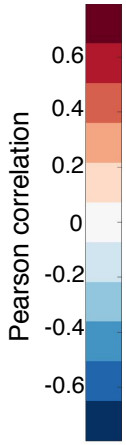
Nobel, Nature Biotech, 2006

Example: single-view unsupervised learning

Non-negative matrix factorization (NMF)

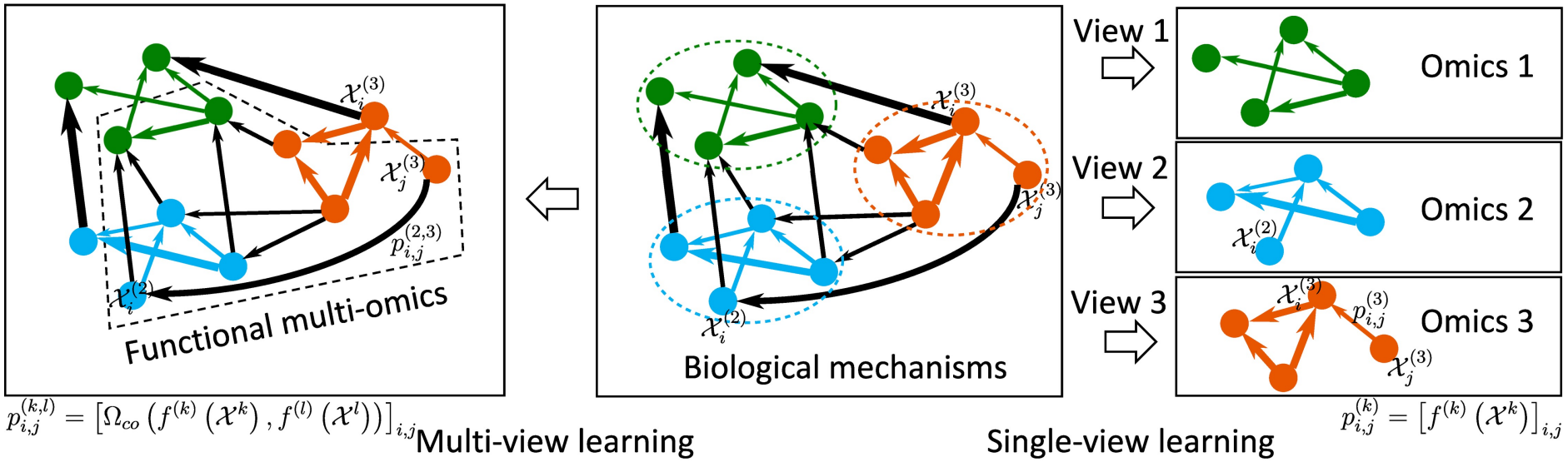






Single cell signatures
- ~14,000 cells (Lake et al., Science, 2016&2018)
- ~400 cells (Darmanis et al., PNAS, 2015)
- ~18,000 cells (PsychENCODE)

Multi-view learning for multi-omics integration



- For example, gene regulation involves
  1. Genomics; e.g., SNPs
  2. Transcriptomics; e.g., genes
  3. Proteomics; e.g., transcription factors (TFs)

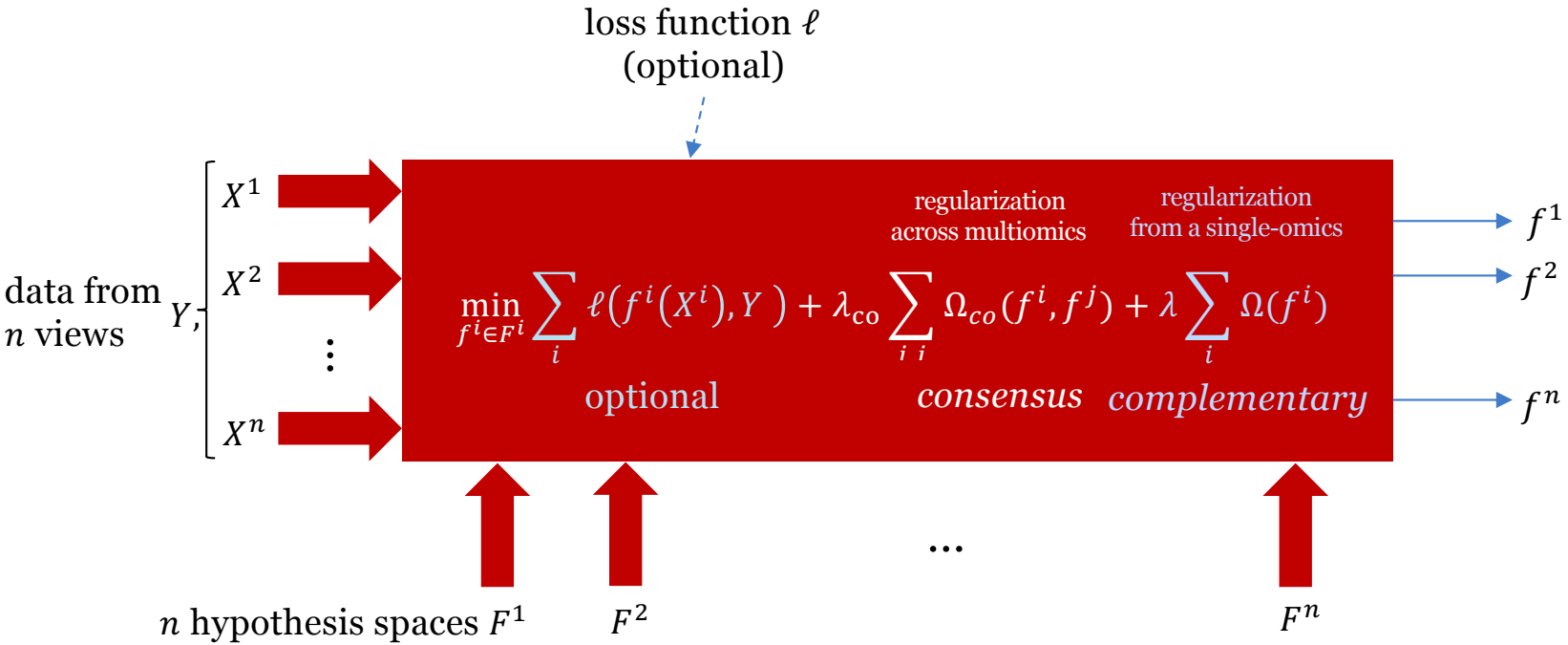
Cross-omics interactions

$\Omega_{co}(f^{(1)},f^{(3)})$: SNPs break TF binding sites

$\Omega_{co}(f^{(2)},f^{(3)})$: TFs control gene expression

$\Omega_{co}(f^{(1)},f^{(2)})$: SNPs associate with gene expression (e.g, eQTLs)

Multiview Empirical Risk Minimization (MV-ERM)



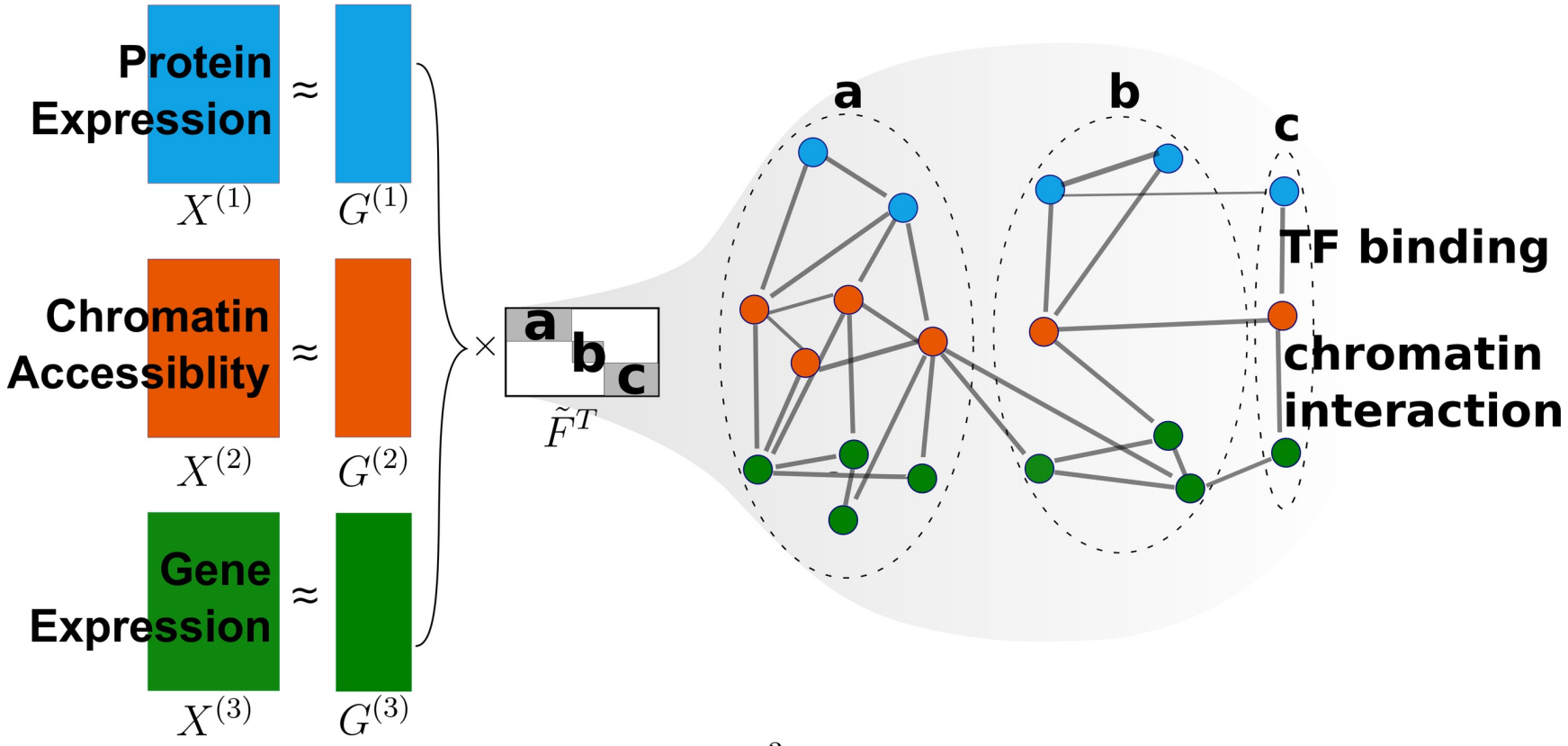
Complementary regularization

- Unique information from each view

Consensus regularization

- Relationship information across views

Factorization-based MV-ERM

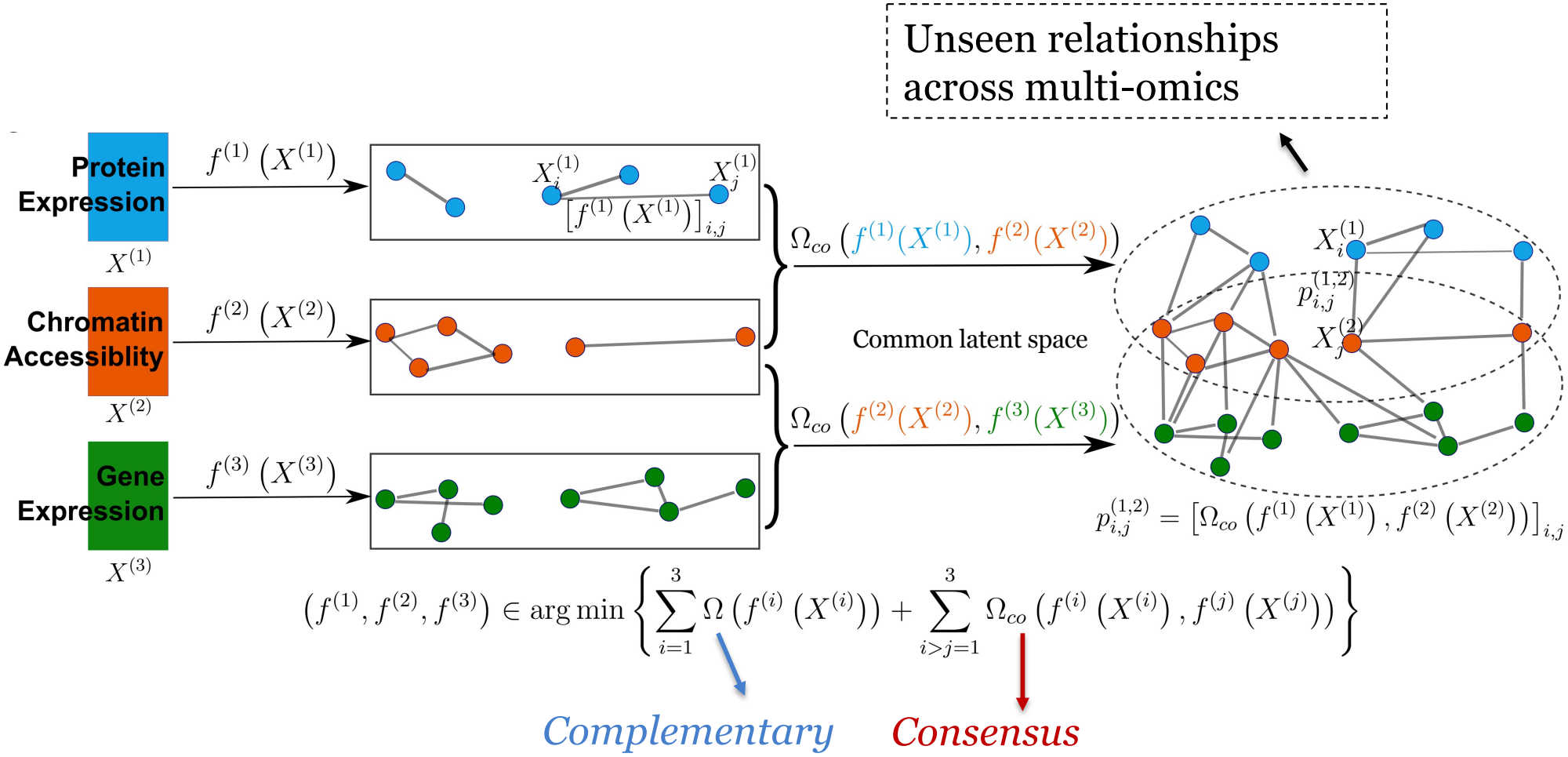


$$\left(G^{(1)*}, G^{(2)*}, G^{(3)*}, \tilde{F}^*\right) \in \underset{G^{(i)}, F^{(i)}, \tilde{F} \geq 0}{\arg\min} \sum_{i=1}^{3} \left\{ \left\| X^{(i)} - G^{(i)} F^{(i)^T} \right\|_F^2 + \lambda \left\| F^{(i)} - \tilde{F} \right\|_F^2 \right\}$$

Complementary $G^{(i)}$ *Consensus* $\tilde{F}$

- e.g., solved by Multi-view NMF (Liu et al., SIAM ICDM, 2013)

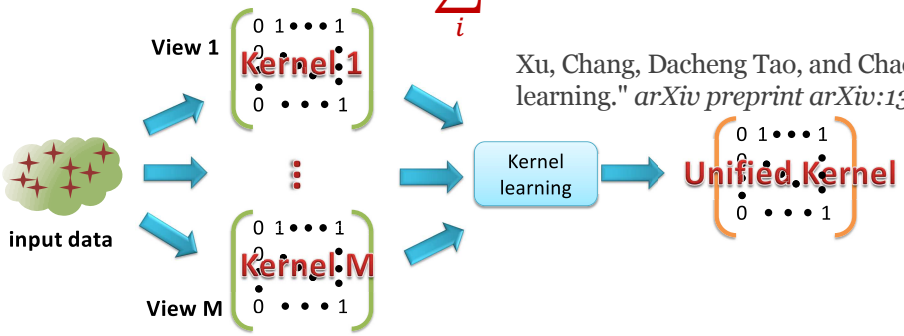
Alignment-based MV-ERM



Alignment-based multi-view learning methods

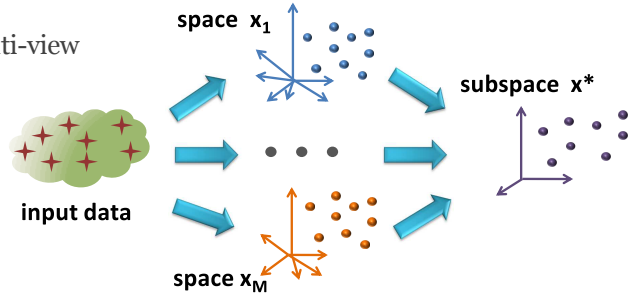
Multiple kernel learning

- learn a combination of predefined kernels
- not data dependent
- expensive
  - Only support complementary principle
  - **non-linear** $\sum_i \beta_i K^{(i)}$



Subspace learning

- obtains a common latent subspace
- **data dependent** → more general
- **non-expensive**
- CCA (Canonical Correlation Analysis)
  - Only support consensus principle
  - linear $\max_{f,g} corr\big(f(X), g(Y)\big)$



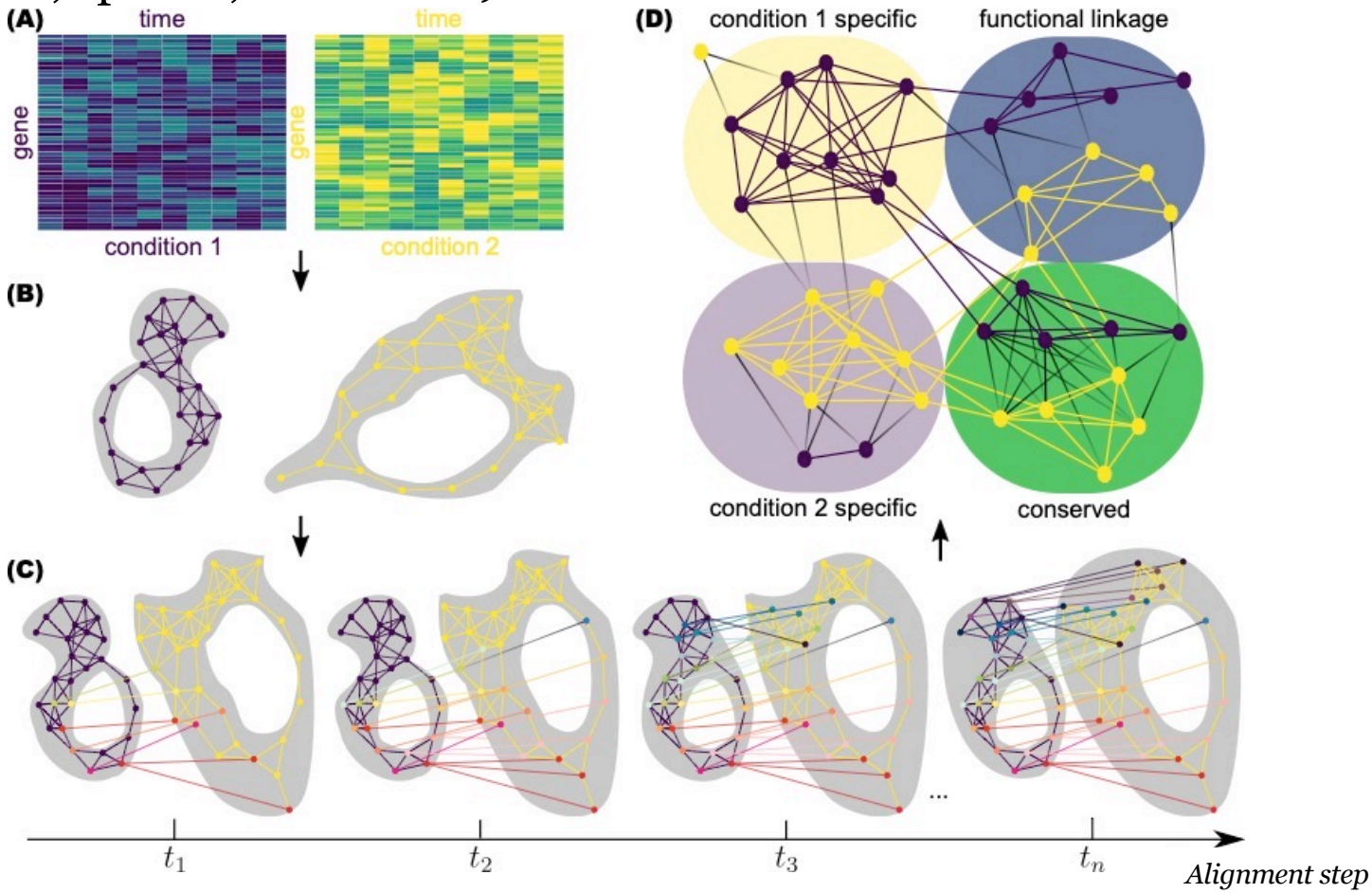
Xu, Chang, Dacheng Tao, and Chao Xu. "A survey on multi-view learning." *arXiv preprint arXiv:1304.5634* (2013)

Can we support both principles?

ManiNetCluster: manifold alignment to reveal the functional links between gene networks

Multi-view datasets (e.g., diseases, species, conditions)

Functional linkages across views

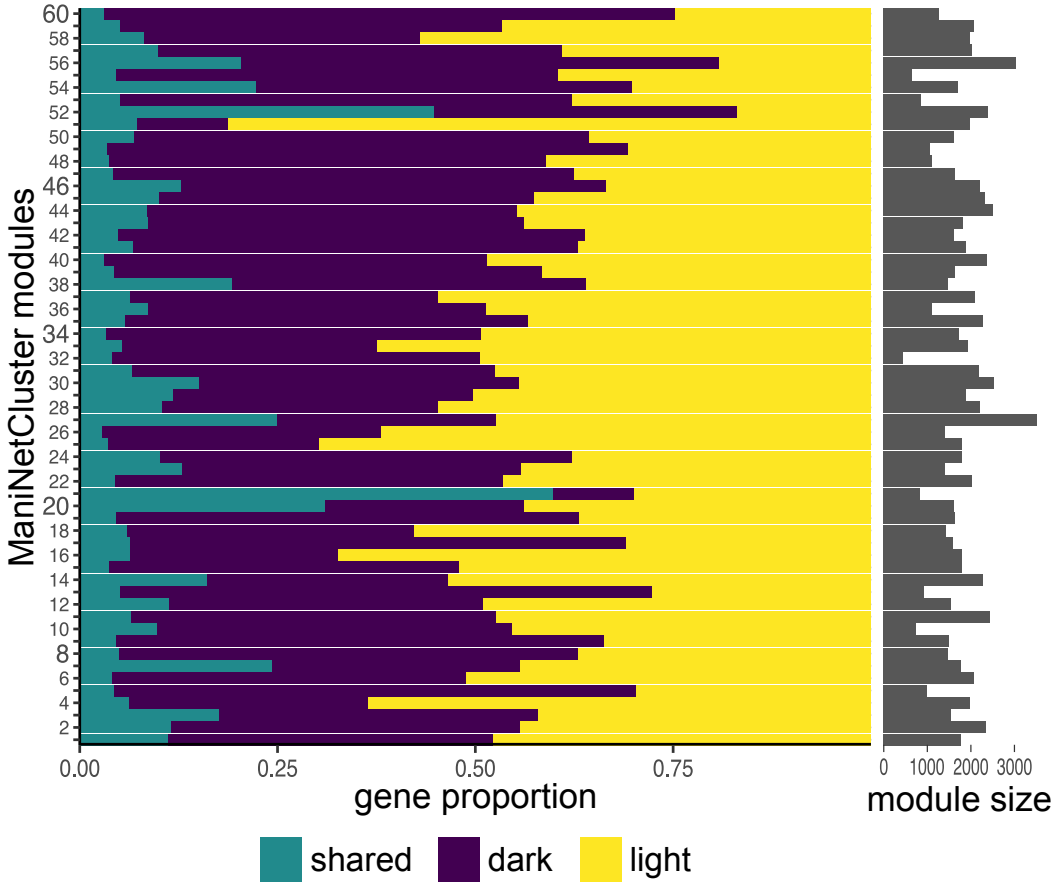


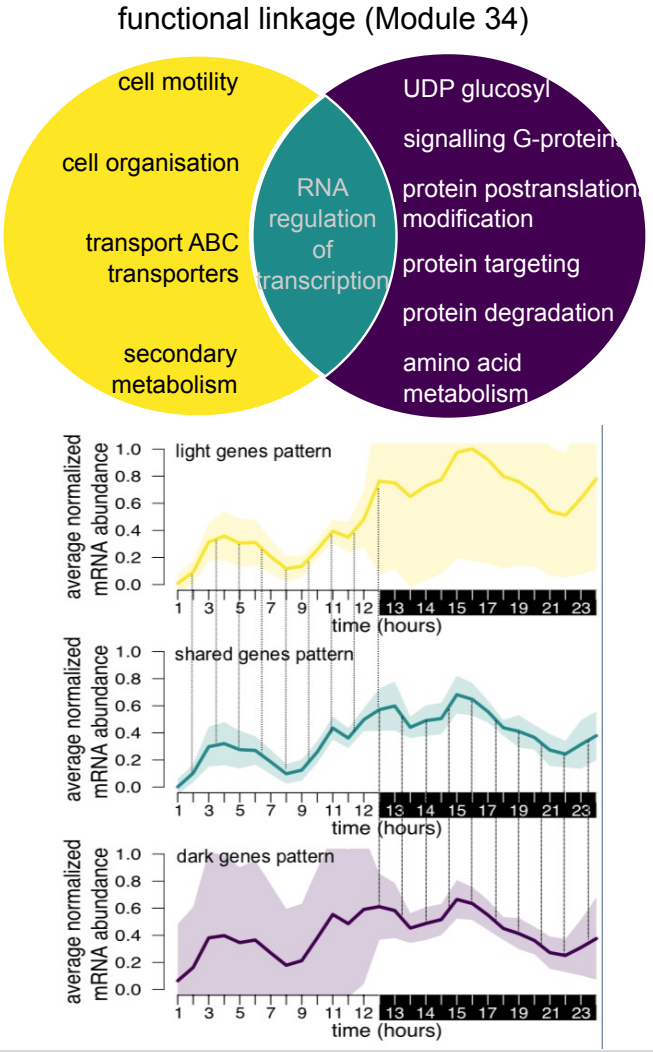
ManiNetCluster is a *subspace learning*, **non-linear** and non-expensive (by employing **manifold learning**) & support both principles

Nguyen, Blaby, Wang, BMC Genomics, 2019 (*Best Poster Award*, ACM BCB 2018)

ManiNetCluster: manifold alignment to reveal the functional links between gene networks

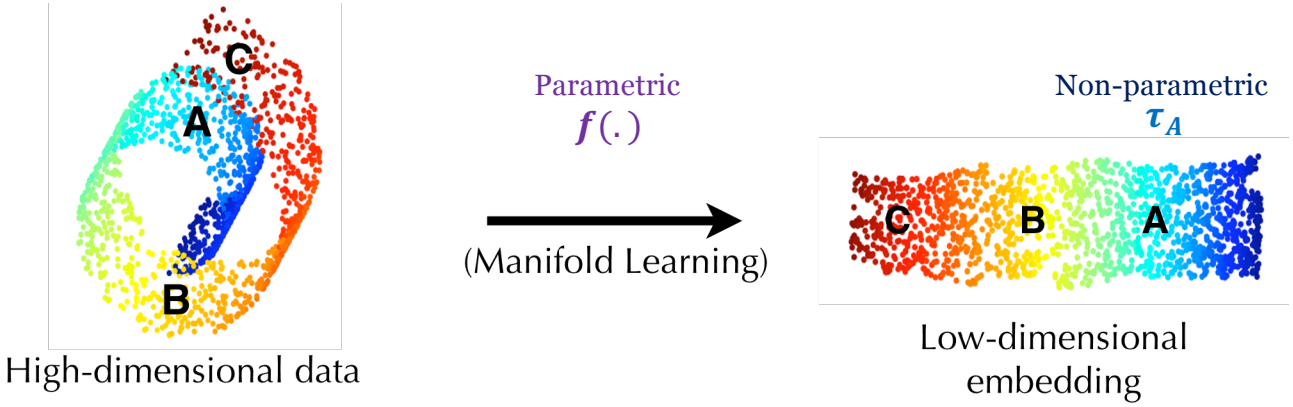
Application: genomic functional linkages between light and dark periods of green alga





Nguyen, Blaby, Wang, BMC Genomics, 2019 (*Best Poster Award*, ACM BCB 2018)

Manifold Learning



A d dimensional manifold M is embedded in a m dimensional space, and there is an explicit mapping f: $\mathbb{R}^m \rightarrow \mathbb{R}^d$ where $d \leq m$. Given samples $x_i \in \mathbb{R}^m$ with noise

$$\boldsymbol{\tau_i = f(x_i)}$$

→ **find** $f(.)$ **or** τ_i from given x_i

Parametric & generalizable (e.g., linear manifold learning)

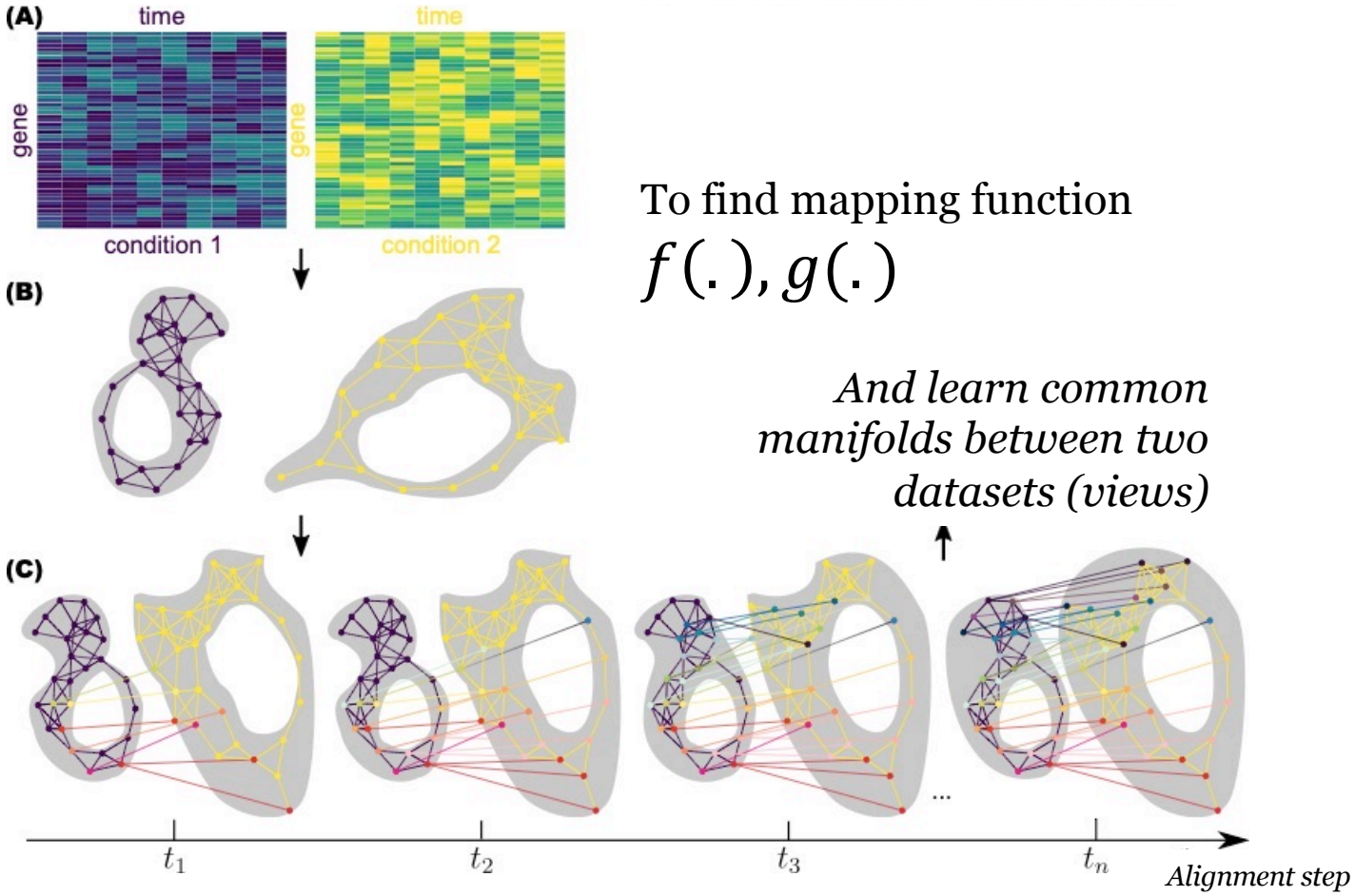
Nonparametric & nongeneralizable (e.g., nonlinear manifold learning)

Manifold Alignment

two datasets $X = [x_1, \dots, x_m], x_i \in \mathbb{R}^p$

$Y = [y_1, \dots, y_n], y_j \in \mathbb{R}^q$

$x_i \leftrightarrow y_i \; for \; i \in [1, l]$



To find mapping function

$f(.), g(.)$

And learn common manifolds between two datasets (views)

Manifold Alignment framed by MV-ERM

Learn common manifolds between two datasets (views)

$$X = [x_1, \dots, x_m], x_i \in \mathbb{R}^p \qquad Y = [y_1, \dots, y_n], y_j \in \mathbb{R}^q \qquad x_i \leftrightarrow y_i \; for \; i \in [1, l]$$

To find mapping function $f(.), g(.)$ to minimize the cost function

Inter-view correspondence

Intra-view correspondence

$$\sum_{i,j} \|f(x_i) - g(y_j)\|^2 W^{i,j} + \sum_{i,j} \|f(x_i) - f(x_j)\|^2 W_X^{i,j} + \sum_{i,j} \|g(y_i) - g(y_j)\|^2 W_Y^{i,j}$$

global consistency
consensus

local smoothness
complementary

- Extract and optimally align local geometry to minimize overall differences
- Generalization of CCA $\sum_{i,j} \|f(x_i) - g(y_j)\|^2$
- Interpreted as a manifold regularization

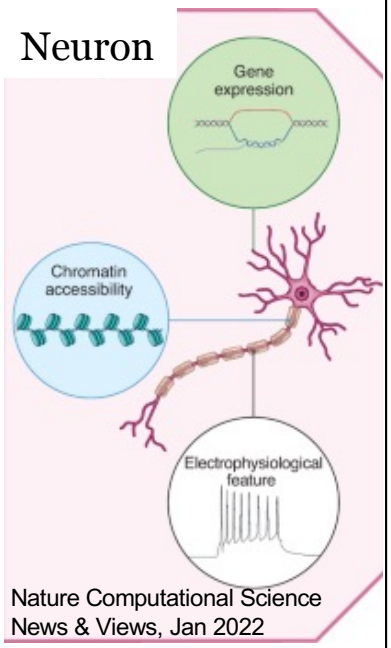
$$\sum_{i,j} \|f(x_i) - g(y_j)\|^2 W^{i,j} + tr(\mathbb{f}^T L_X \mathbb{f}) + tr(\mathbb{g}^T L_Y \mathbb{g})$$

- Eigen-decomposition to solve nonlinear manifold alignment

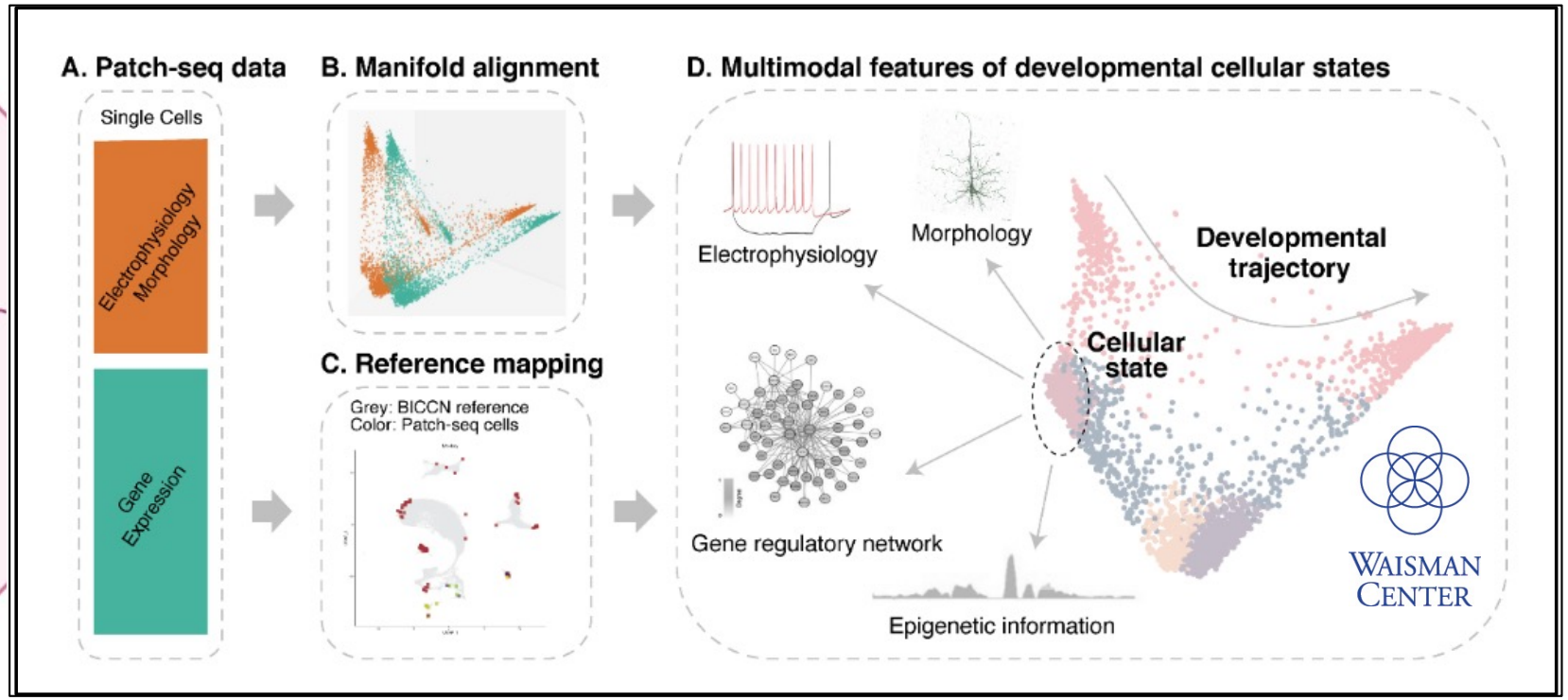
Outline

- Background
  - Functional genomics & multi-omics
  - Gene regulation
- Multi-view learning for multi-modal data
  - Multi-view Empirical Risk Minimization (MV-ERM)
  - Manifold alignment
- App 1: *single-cell multi-modal data integration*
  - Predicting neuronal electrophysiology from gene expression
- App 2: *deep manifold-regularized classification*
  - Predicting cortical layers of neurons from multi-modalities
- App 3: *biologically interpretable neural network modeling*
  - Disease prediction and functional prioritization

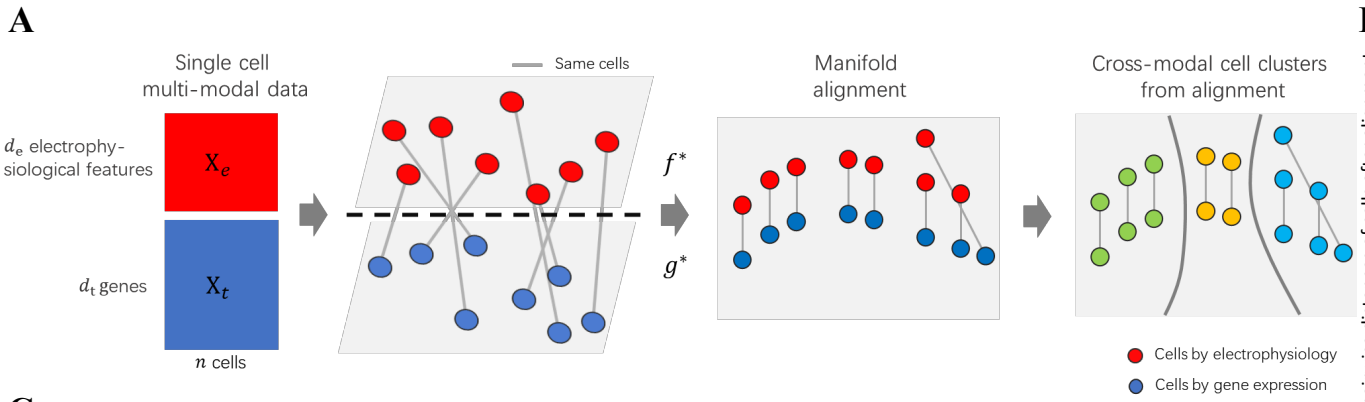
Single-cell multi-modal data by Patch-seq (beyond omics)

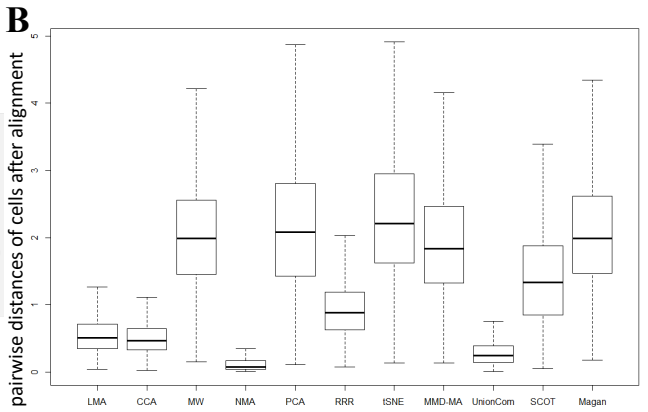


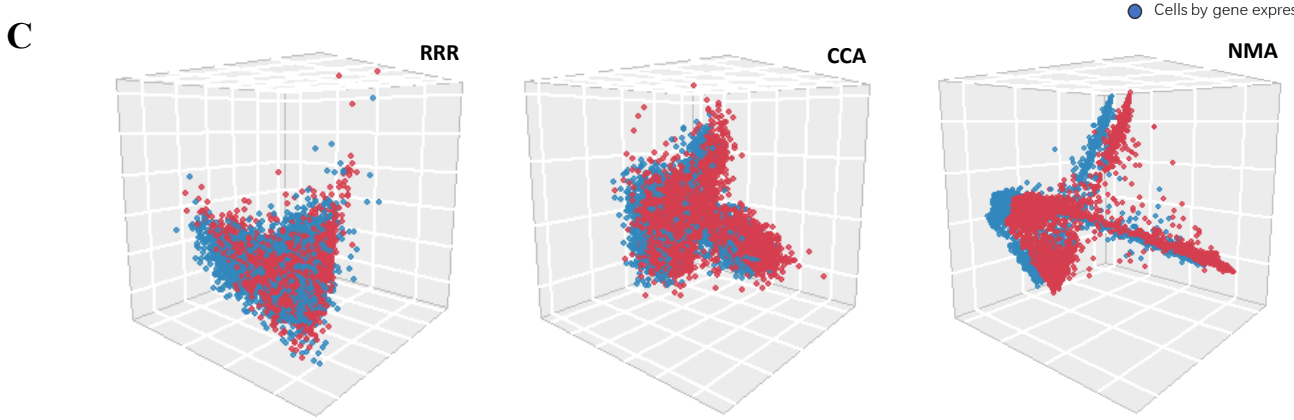
Nature Computational Science
News & Views, Jan 2022



Manifold alignment of single neurons by electrophysiology and gene expression







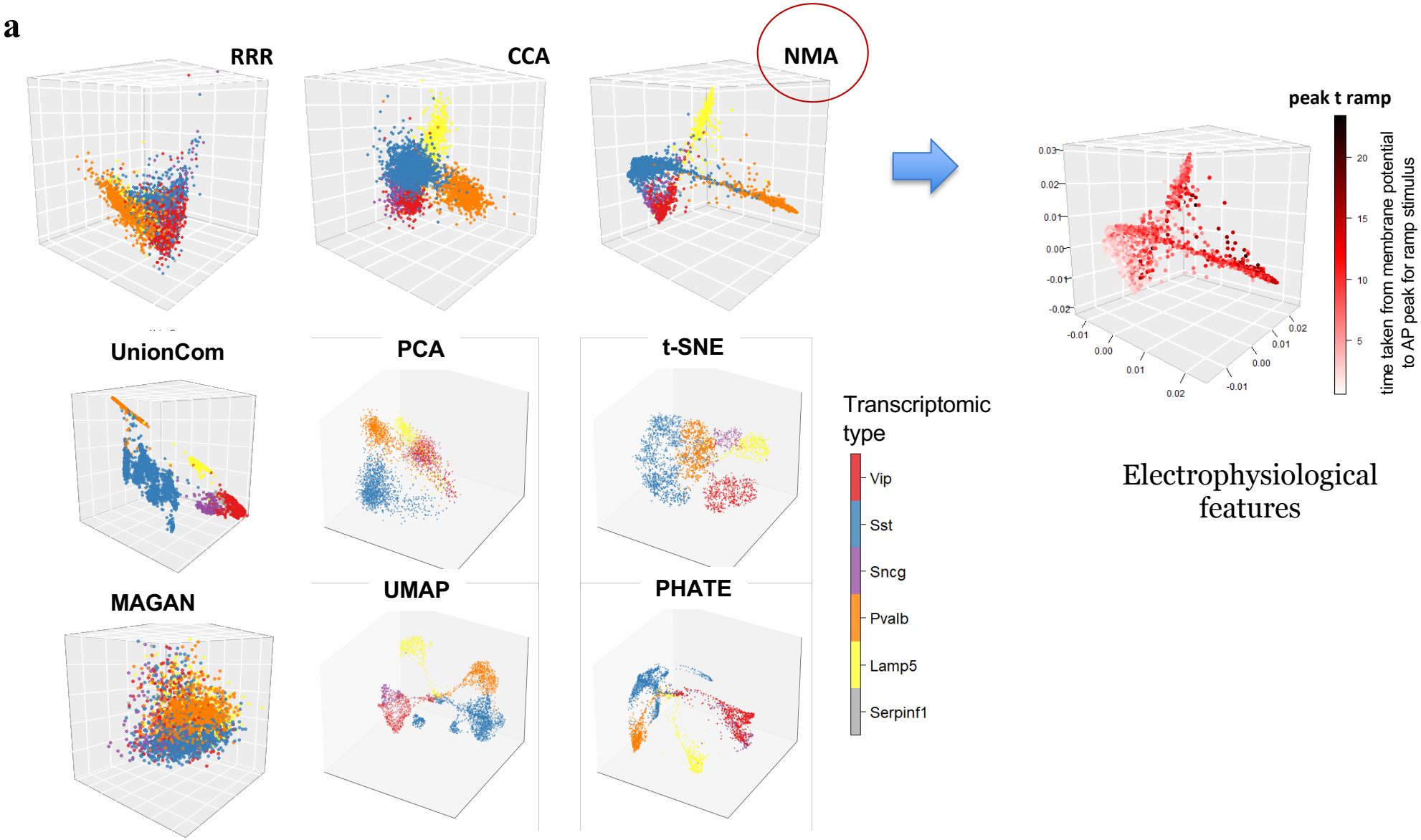
LMA = Linear Manifold Alignment
CCA = Canonical Correlation Analysis
MW = Manifold Warping
NMA = Nonlinear Manifold Alignment
PCA = Principal Component Analysis
RRR = Reduced Rank Regression
tSNE = Stochastic Neighbor Embedding
MMD-MA = Manifold Alignment with maximum mean discrepancy
UnionCom = Unsupervised Topological Alignment of Single-cell Multi-omics Integration
SCOT = Single-Cell alignment using Optimal Transport
MAGAN = Manifold Aligning GAN

Jiawei Huang (M.S., Statistics, 2021)

Patch-seq data of ~3k neuronal cells in the mouse visual cortex from BRAIN Initiative

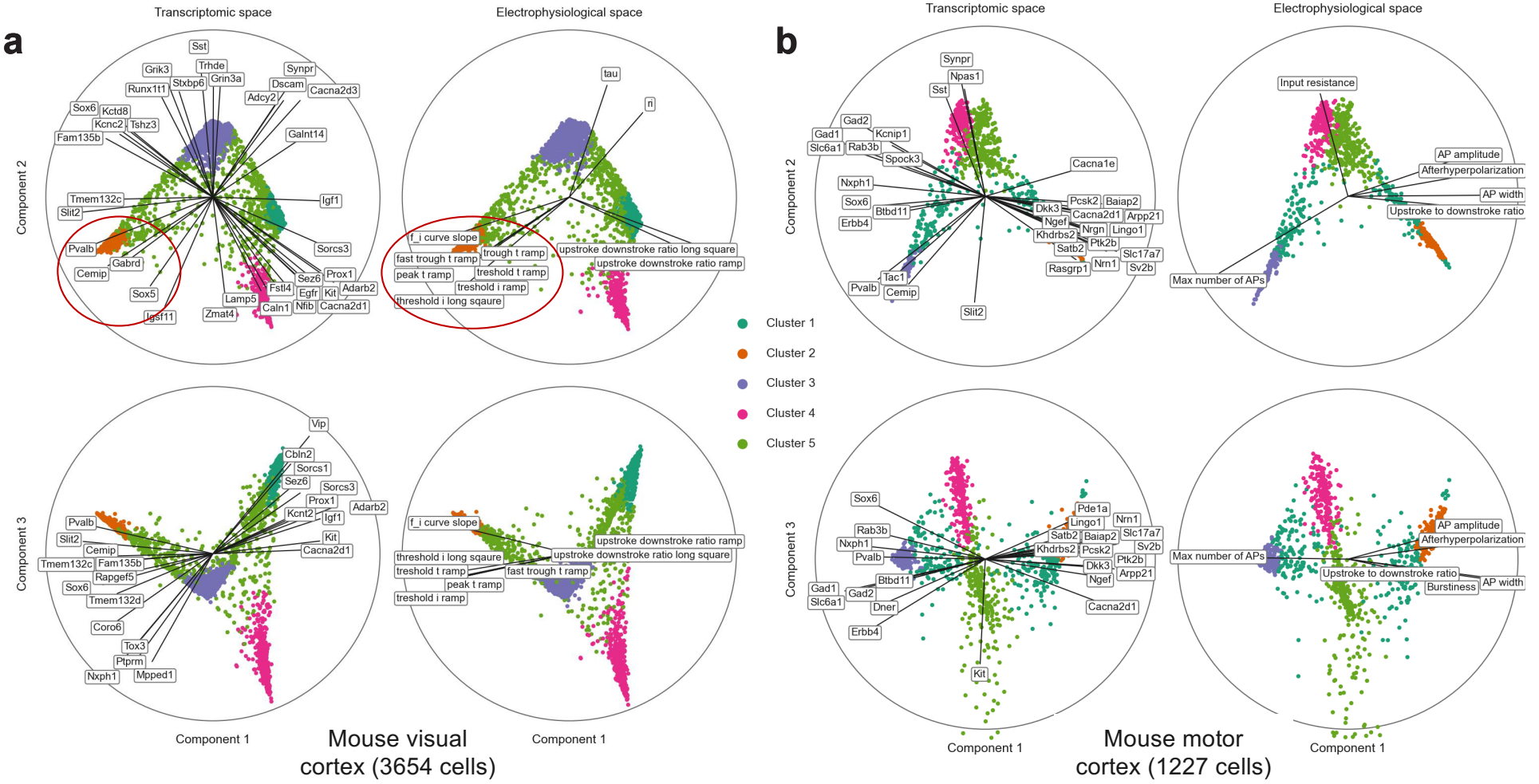
Huang, Sheng, Wang, Communications Biology, 2021

Nonlinear Manifold Alignment (NMA) uncovers trajectory with multi-modal changes



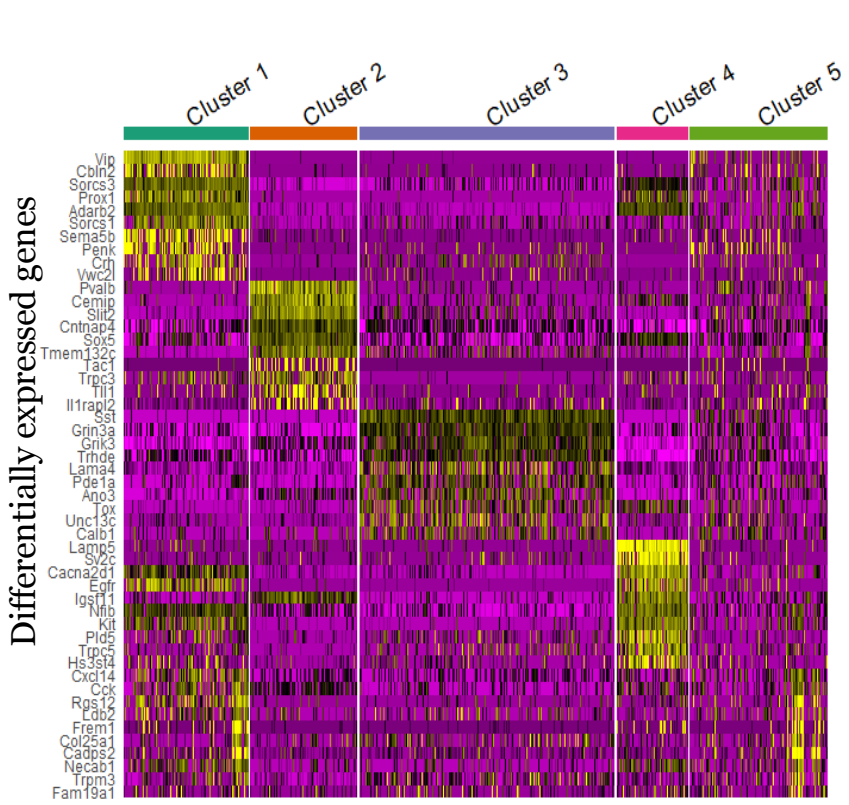
Huang, Sheng, Wang, Communications Biology, 2021

NMA aligned cells form cross-modal clusters, suggesting related genes and electrophysiological features

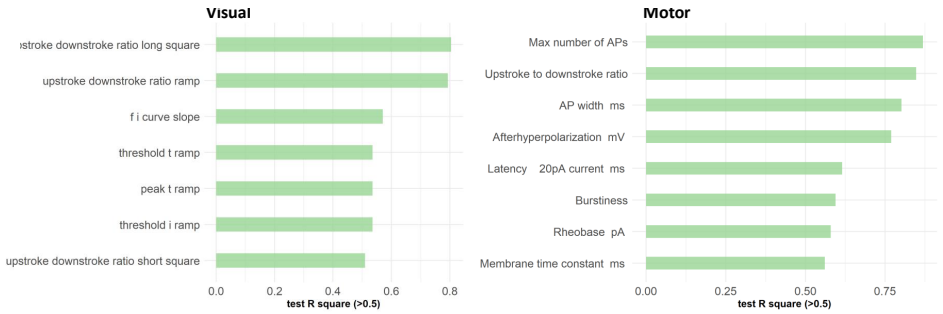


Huang, Sheng, Wang, Communications Biology, 2021

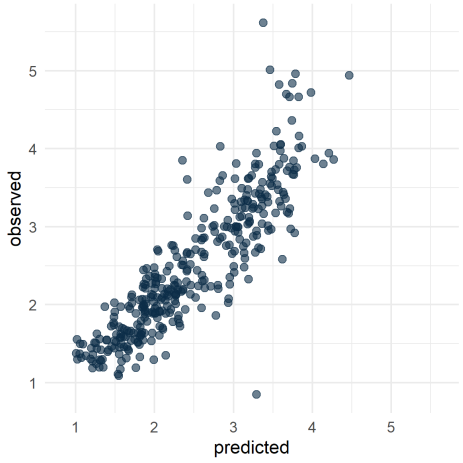
Predicting electrophysiological features by differentially expressed genes in cross-modal clusters



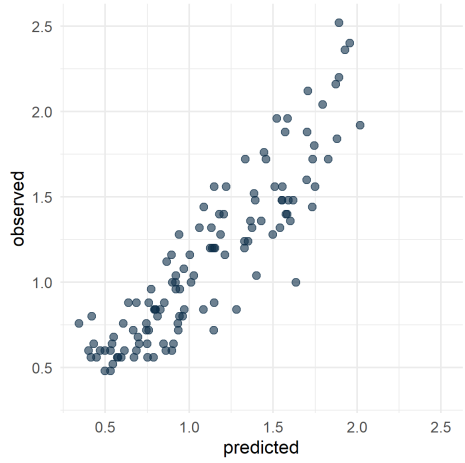
Multivariate regression



Highly predictive e-phys features



Upstroke Downstroke ratio (Long Square, Visual)



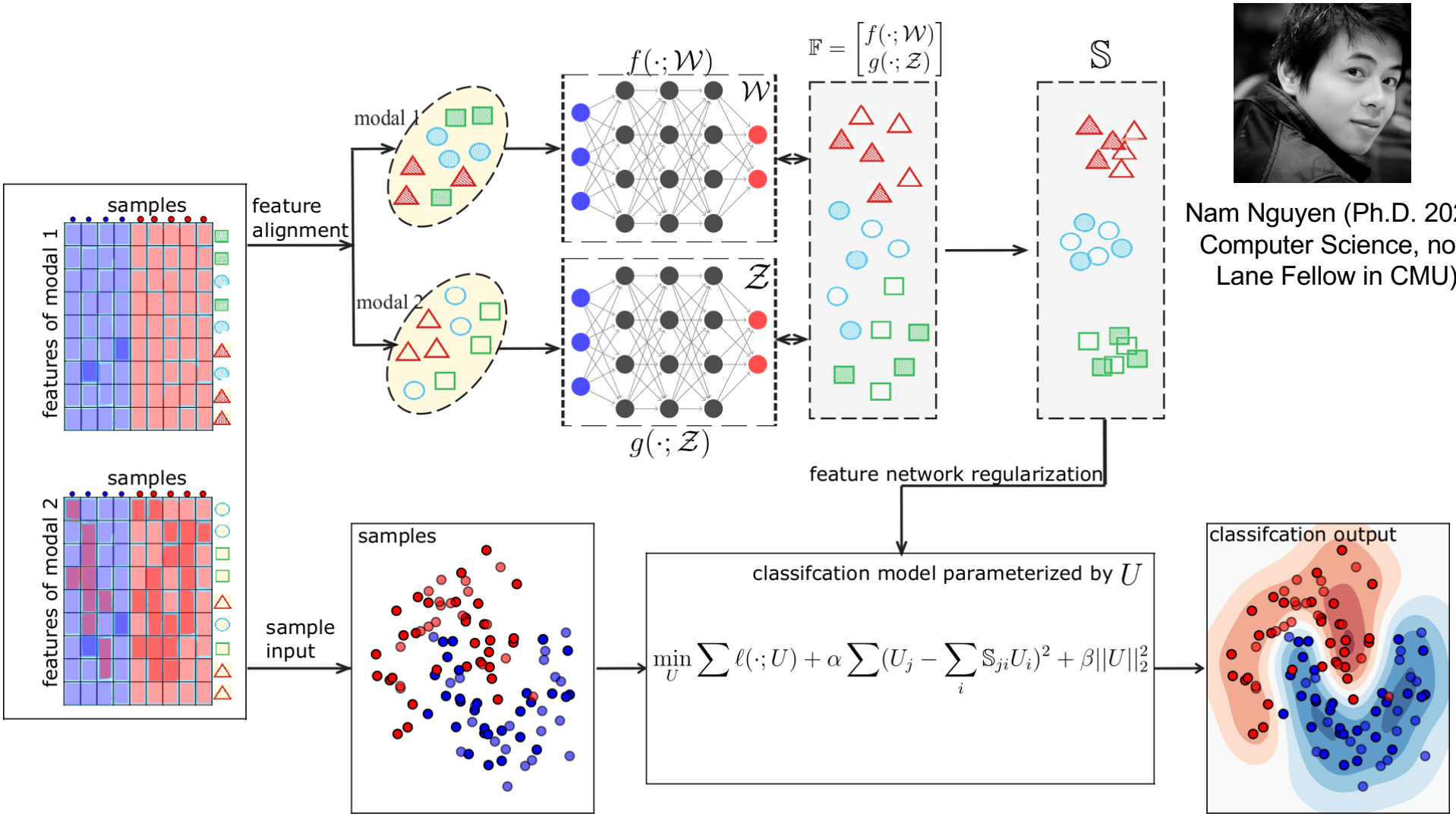
Ap width (Motor)

Can we align multi-modal features to improve phenotype prediction?

Outline

- Background
  - Functional genomics & multi-omics
  - Gene regulation
- Multi-view learning for multi-modal data
  - Multi-view Empirical Risk Minimization (MV-ERM)
  - Manifold alignment
- App 1: *single-cell multi-modal data integration*
  - Predicting neuronal electrophysiology from gene expression
- App 2: *deep manifold-regularized classification*
  - Predicting cortical layers of neurons from multi-modalities
- App 3: *biologically interpretable neural network modeling*
  - Disease prediction and functional prioritization

deepManReg: a deep manifold-regularized learning model for phenotype prediction from multi-modal data

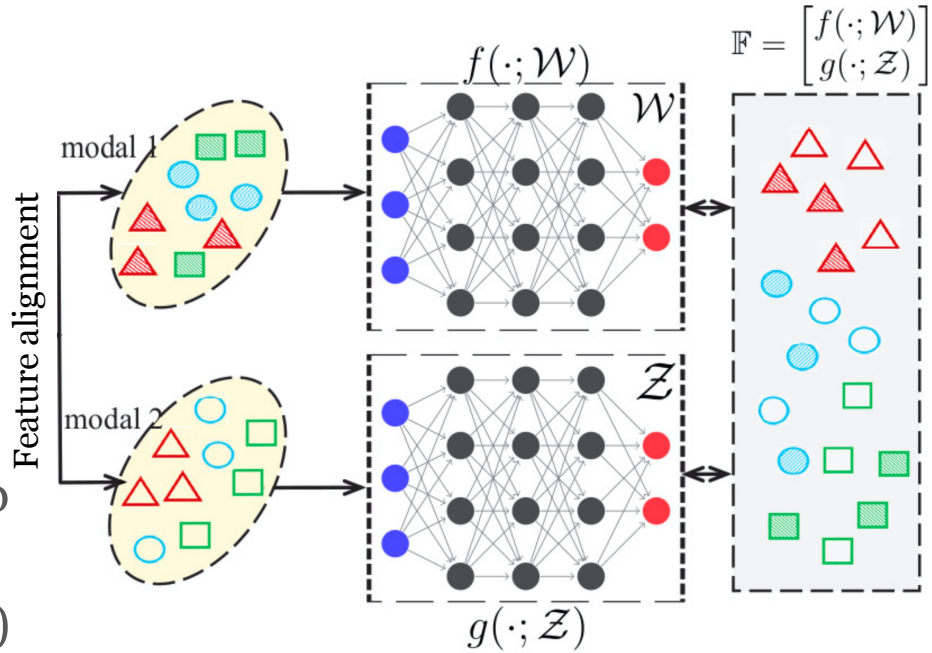


Nam Nguyen (Ph.D. 2021, Computer Science, now Lane Fellow in CMU)

Nguyen, Huang, Wang, Nature Computational Science, 2022

Phase 1: deep manifold alignment of multi-modal features

- Goal: reveal nonlinear relationships across multi-modal features
- Trade-off:
  - Features → Parametric
  - Nonlinear → Nonparametric
- deepManReg solution:
  - Parameterize $f(\cdot,\mathcal{W})$ & $g(\cdot,\mathcal{Z})$ by deep neural networks
  - Nonlinear manifold alignment (NMA)
- Eigen-decomposition to solve NMA:
  - Computationally intensive
  - Use gradient descent on a non-Euclidean space!



Nonlinear manifold alignment

$$\min_{f,g} \operatorname{tr}(\mathbb{F}^{\mathrm{T}} L \mathbb{F})$$

$$\text{s.t. } \mathbb{F}^{\mathrm{T}} D \mathbb{F} = \mathbb{I}$$

Nguyen, Huang, Wang, Nature Computational Science, 2022

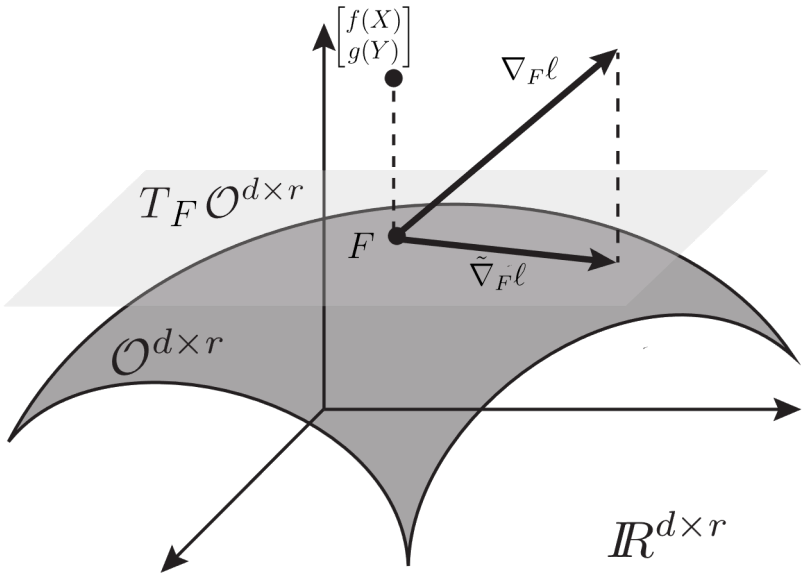
Optimization for deepManReg alignment

$\mathbb{F} = [f(X;\mathcal{W})^T, g(Y;\mathcal{Z})^T]^T$ can be solved by

$$\min_{f,g} tr(\mathbb{F}^T L \mathbb{F}) \text{ s.t. } \mathbb{F}^T D \mathbb{F} = \mathbb{I},$$

where L and D are joint Laplacian and diagonal matrices of prior correspondence across modalities

→ an optimization problem on Stiefel manifold $\mathcal{O}^{d\times r}$



- Forward pass — project the output $\begin{bmatrix} f(X) \\ g(Y) \end{bmatrix}$ onto Stiefel manifold

$$F = \pi \circ \begin{bmatrix} f(X) \\ g(Y) \end{bmatrix} = UIV^T$$

, where $\begin{bmatrix} f(X) \\ g(Y) \end{bmatrix} = U\Sigma V^T$ is the SVD of output

- Backward pass
  - Compute Euclidean gradient

$$\nabla_F \ell = \frac{\partial tr(F^T L F)}{\partial F} = LF + L^T F$$

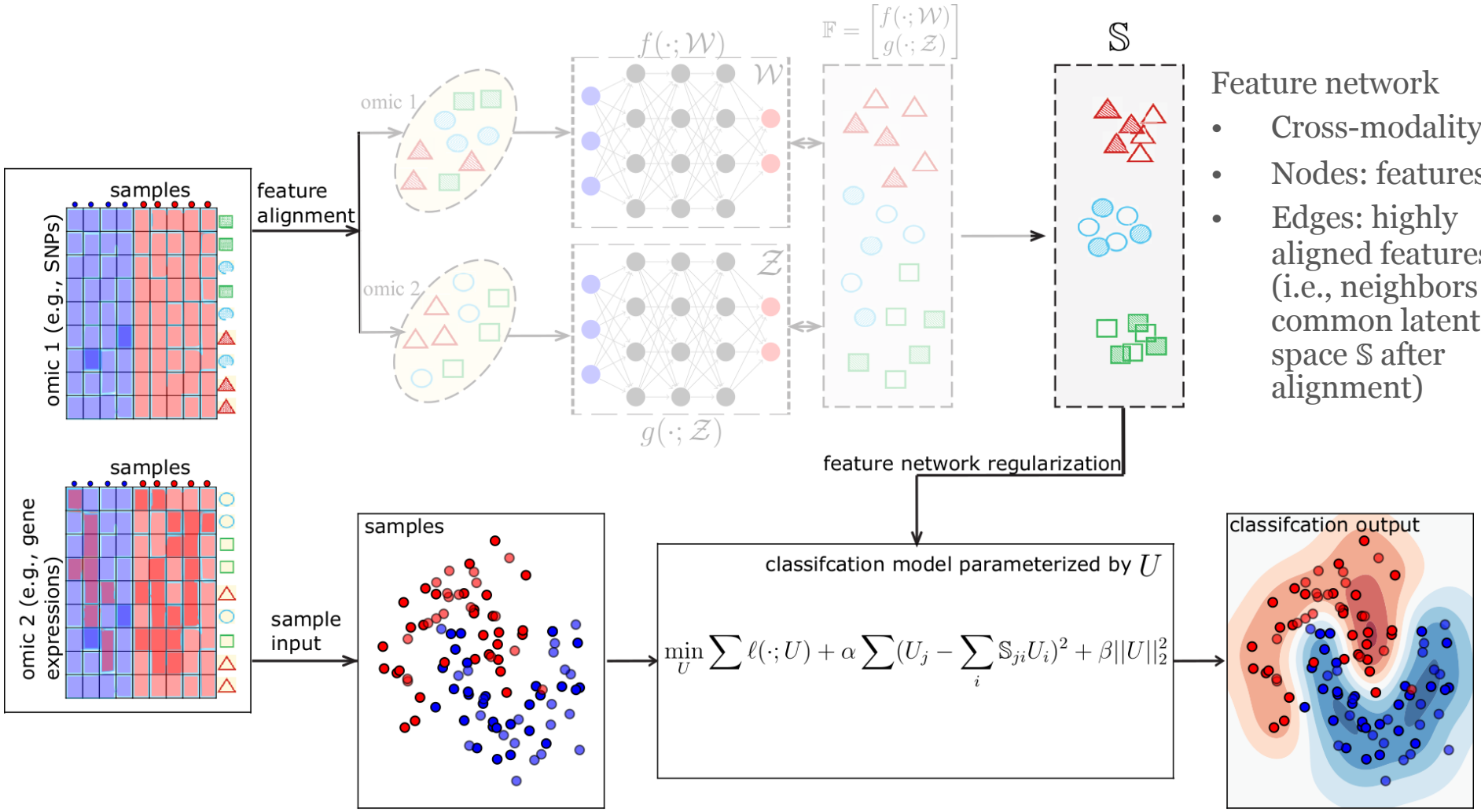
  - Project Euclidean gradient onto the tangent space of Stiefel manifold

$$\tilde{\nabla}_F \ell = \pi(\nabla_F \ell) = Fskew(F^T \nabla_F \ell) + (I - FF^T)\nabla_F \ell$$

  - Backpropagate the Riemannian gradient $\tilde{\nabla}_F \ell$ to update $\mathcal{W}, \mathcal{Z}$

Cunningham et al. *The Journal of Machine Learning Research* 16.1 (2015): 2859-2900.

Phase 2: classification regularized by cross-modal feature network (aligned features)



Feature network

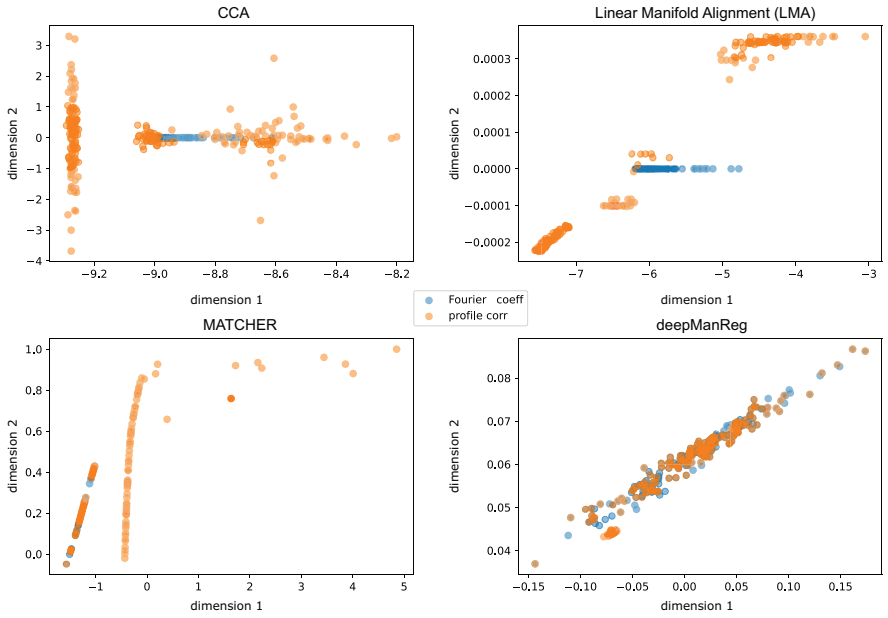
- Cross-modality
- Nodes: features
- Edges: highly aligned features (i.e., neighbors on common latent space $\mathbb{S}$ after alignment)

- Sample classification regularized by cross-modal feature network

Nguyen, Huang, Wang, Nature Computational Science, 2022

deepManReg alignment & classification of handwritten digits

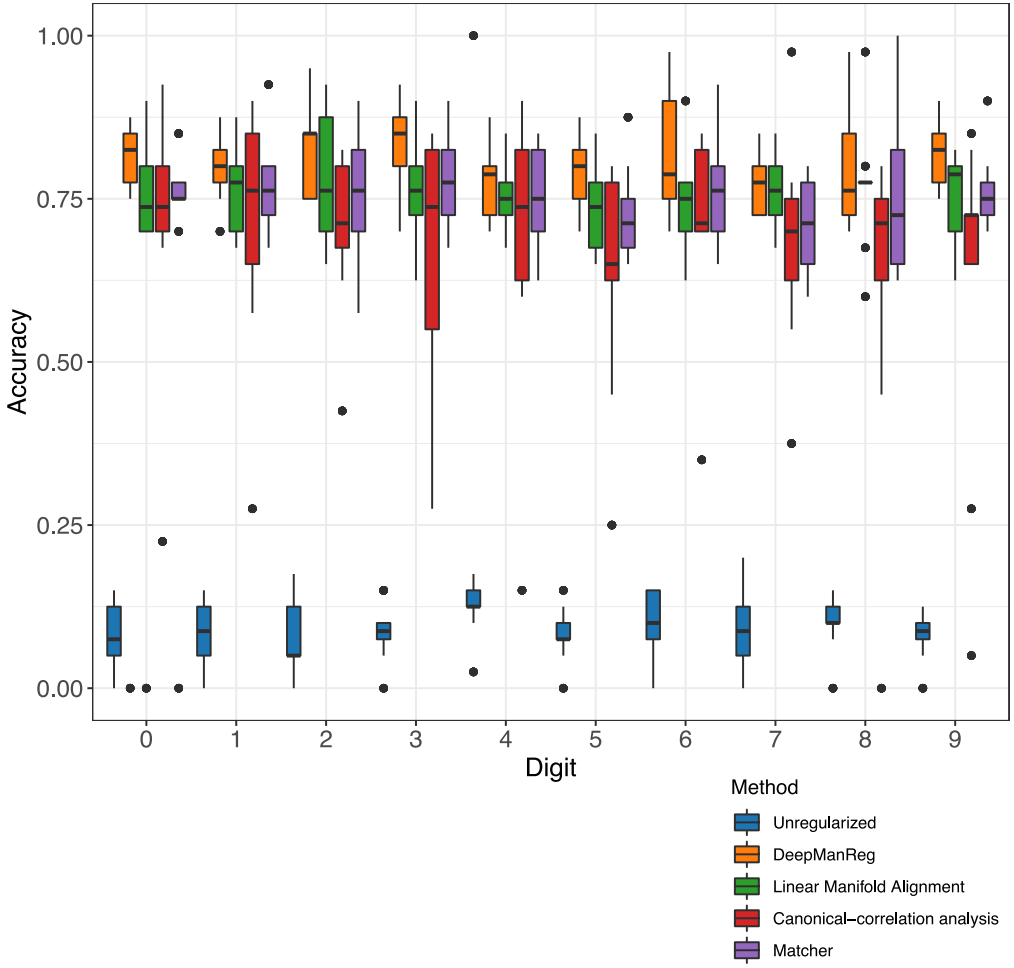
Multi-modal feature alignment



- 2000 images of handwritten digits 0-9 (mfeat data) Breukelen et al. Kybernetika, 34(4):381–386, 1998.
- Two types of features:

216 profile correlations, 76 Fourier coefficients

Digit classification

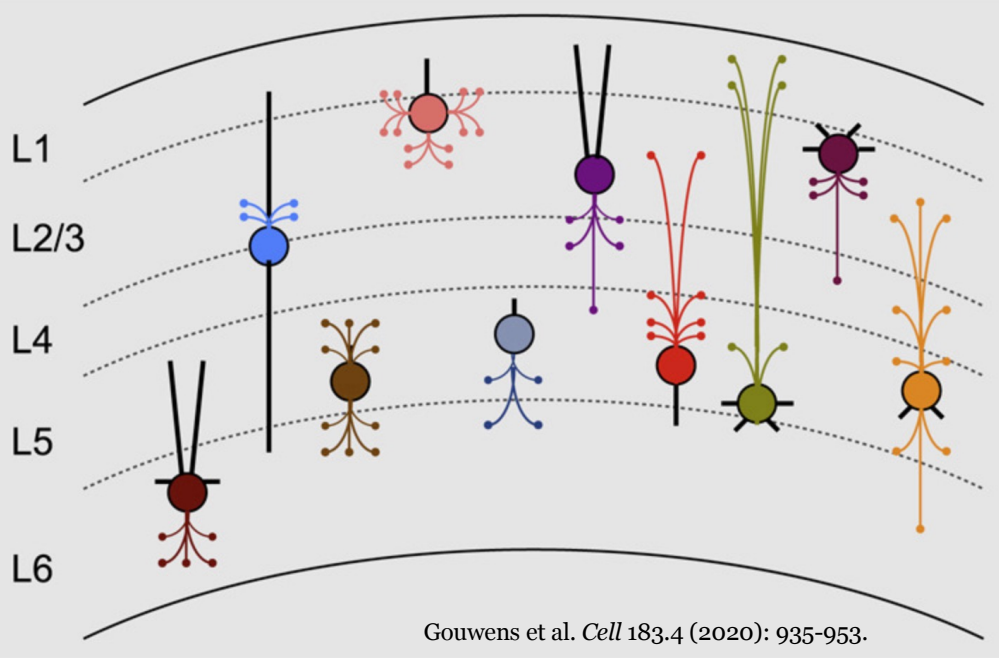


Nguyen, Huang, Wang, Nature Computational Science, 2022

deepManReg alignment & classification of neuronal cells

Single-cell multi-modal data by Patch-seq

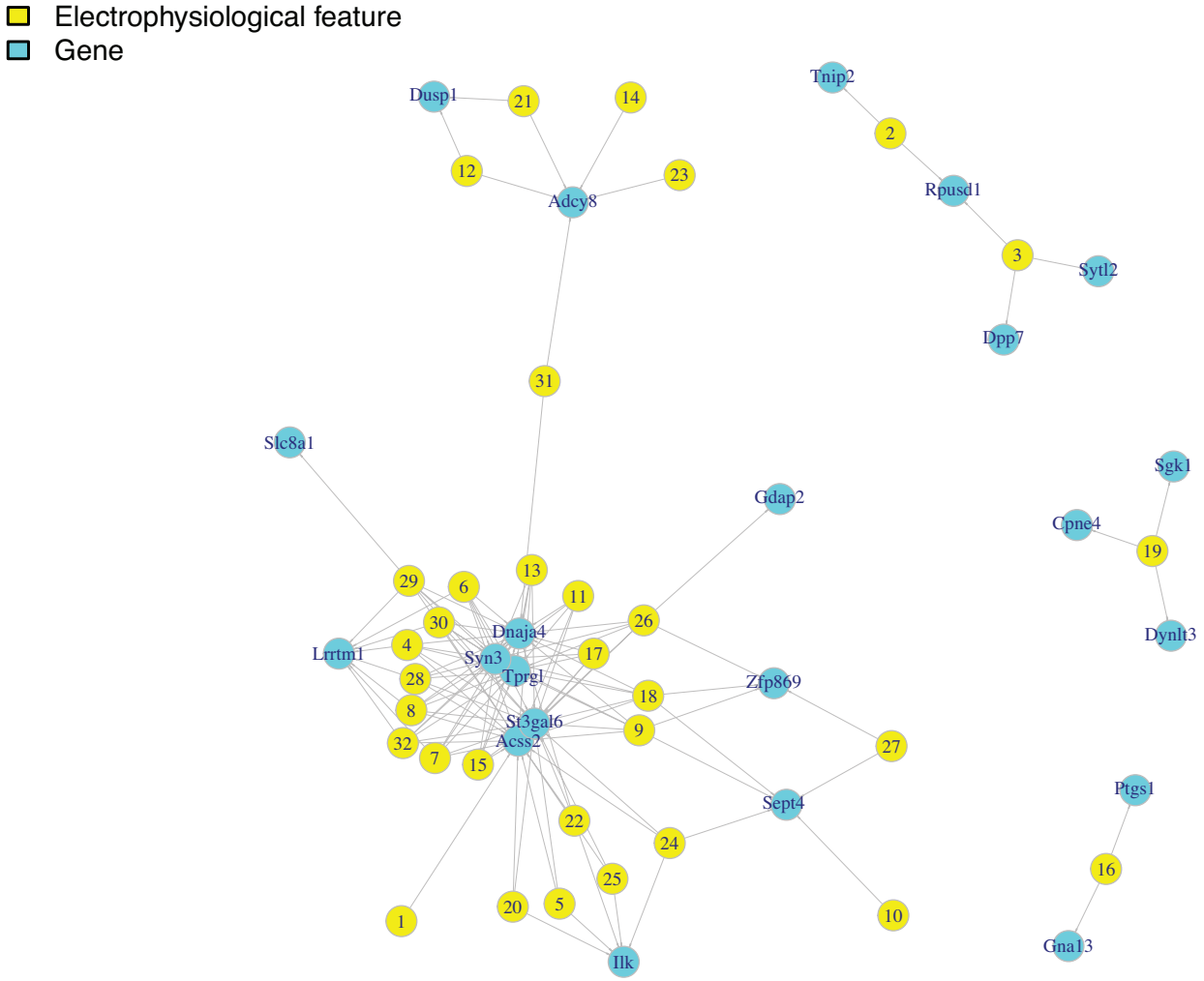
- 3654 GABAergic cortical neurons in mouse visual cortex
- Modality 1: Electrophysiological features
  - By hyperpolarizing and depolarizing current injection stimuli and responses of short (3 ms) current pulses, long (1 s) current steps, and slow (25 pA/s) current ramps.
- Modality 2: Genes
  - Expression levels
- Phenotype: 5 cortical layers, L1, L2/3, L4, L5, and L6
- Running time (alignment only)
  - CCA (725.96 seconds)
  - Manifold Alignment (663.43 seconds) MATCHER (150.94 seconds)
  - deepManReg (57.90 seconds by GPUs GTX 1060Ti and 90.10 seconds by CPU i5-8250U)



Gouwens et al. *Cell* 183.4 (2020): 935-953.

Nguyen, Huang, Wang, Nature Computational Science, 2022

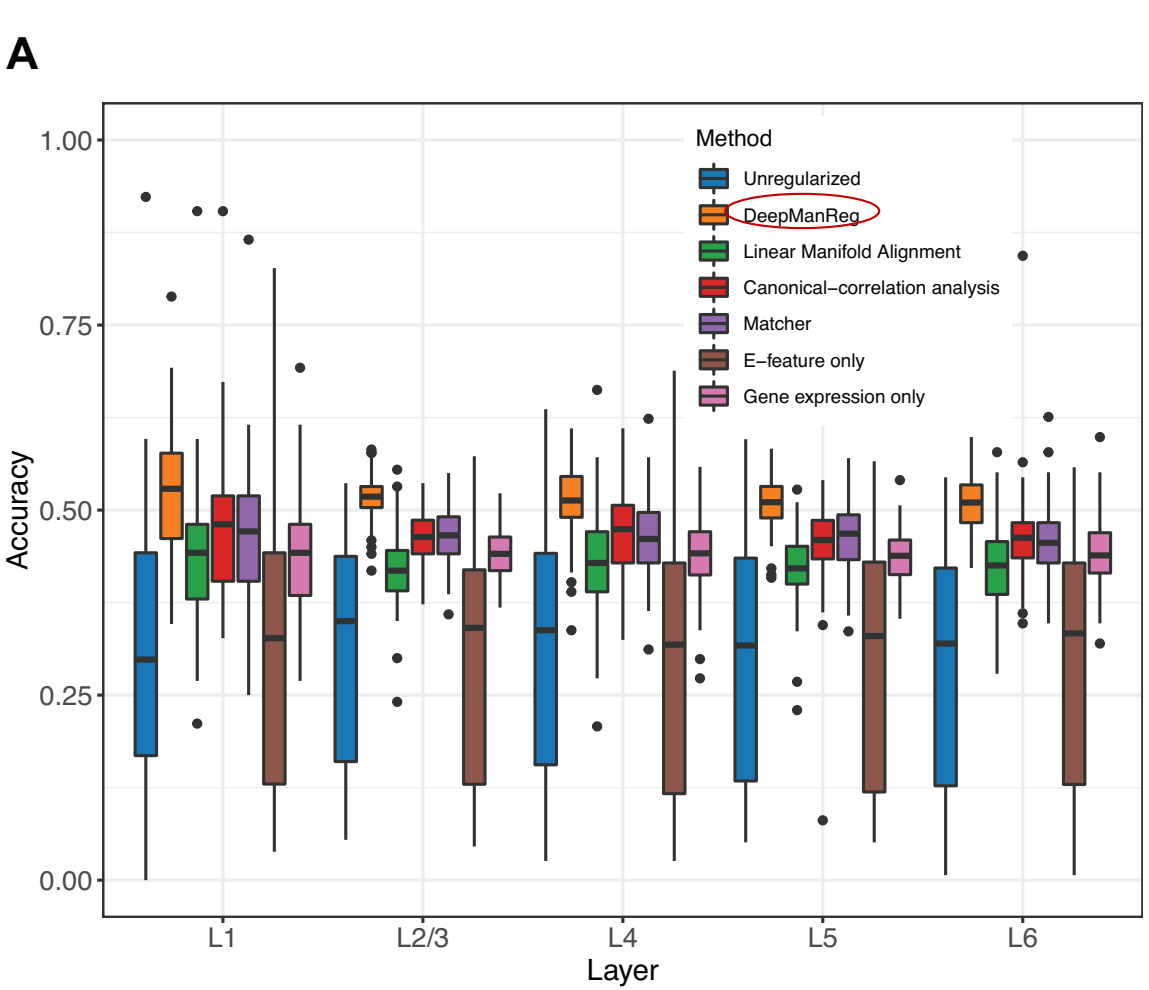
Cross-modal feature network linking aligned genes and electrophysiological features

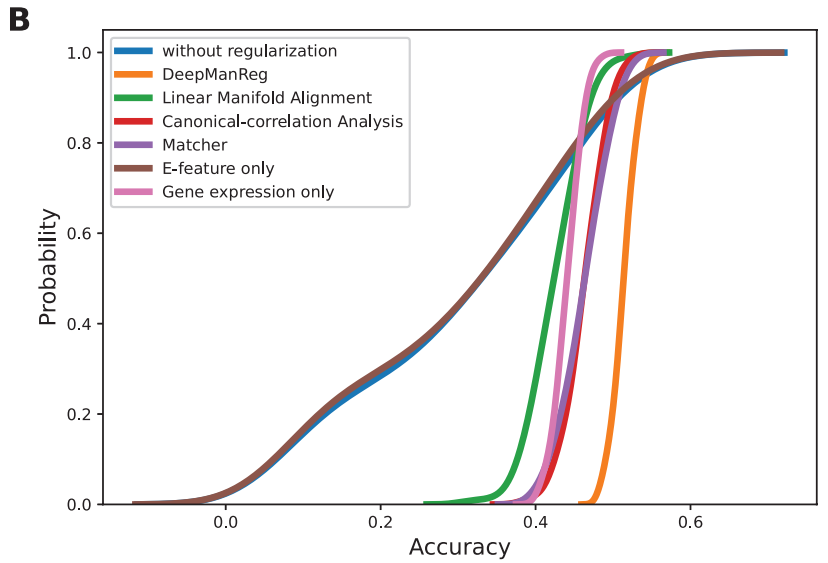


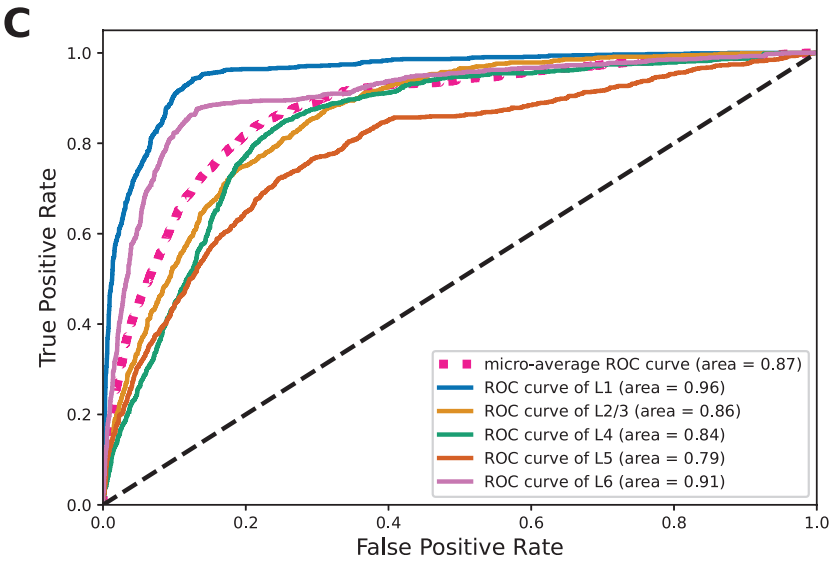
e-Features	Index
rheobase_sweep_num	1
thumbnail_sweep_num	2
vrest	3
sag	4
tau	5
f_i_curve_slope	6
adaptation	7
latency	8
upstroke_downstroke_ratio_long_square	9
peak_v_long_square	10
peak_t_long_square	11
trough_v_long_square	12
trough_t_long_square	13
fast_trough_v_long_square	14
fast_trough_t_long_square	15
threshold_i_long_square	16
threshold_t_long_square	17
upstroke_downstroke_ratio_ramp	18
peak_v_ramp	19
peak_t_ramp	20
trough_v_ramp	21
trough_t_ramp	22
fast_trough_v_ramp	23
fast_trough_t_ramp	24
threshold_t_ramp	25
upstroke_downstroke_ratio_short_square	26
peak_v_short_square	27
peak_t_short_square	28
trough_t_short_square	29
fast_trough_t_short_square	30
threshold_v_short_square	31
threshold_t_short_square	32

Nguyen, Huang, Wang, Nature Computational Science, 2022

Feature-network-regularized classification of cortical layers of neuronal cells





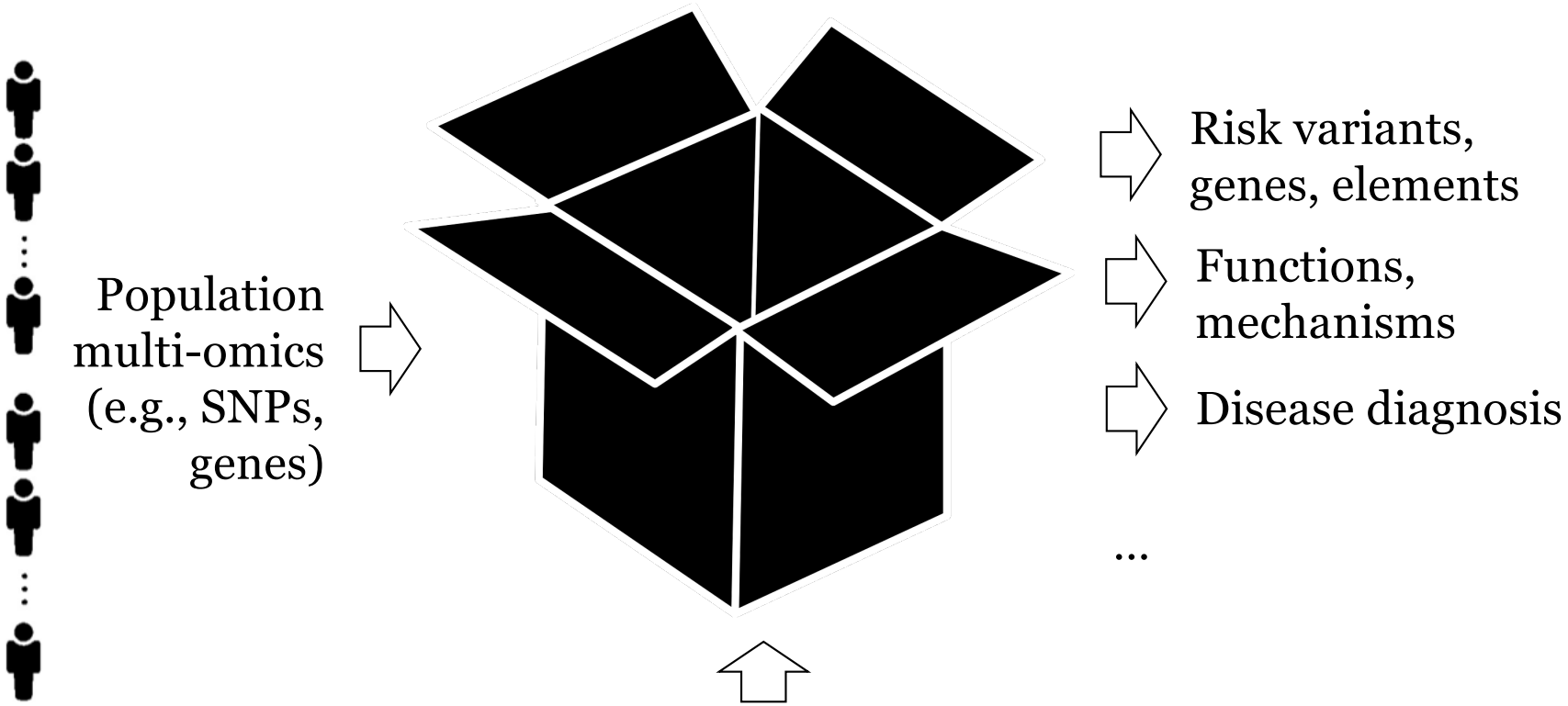


Nguyen, Huang, Wang, Nature Computational Science, 2022

Outline

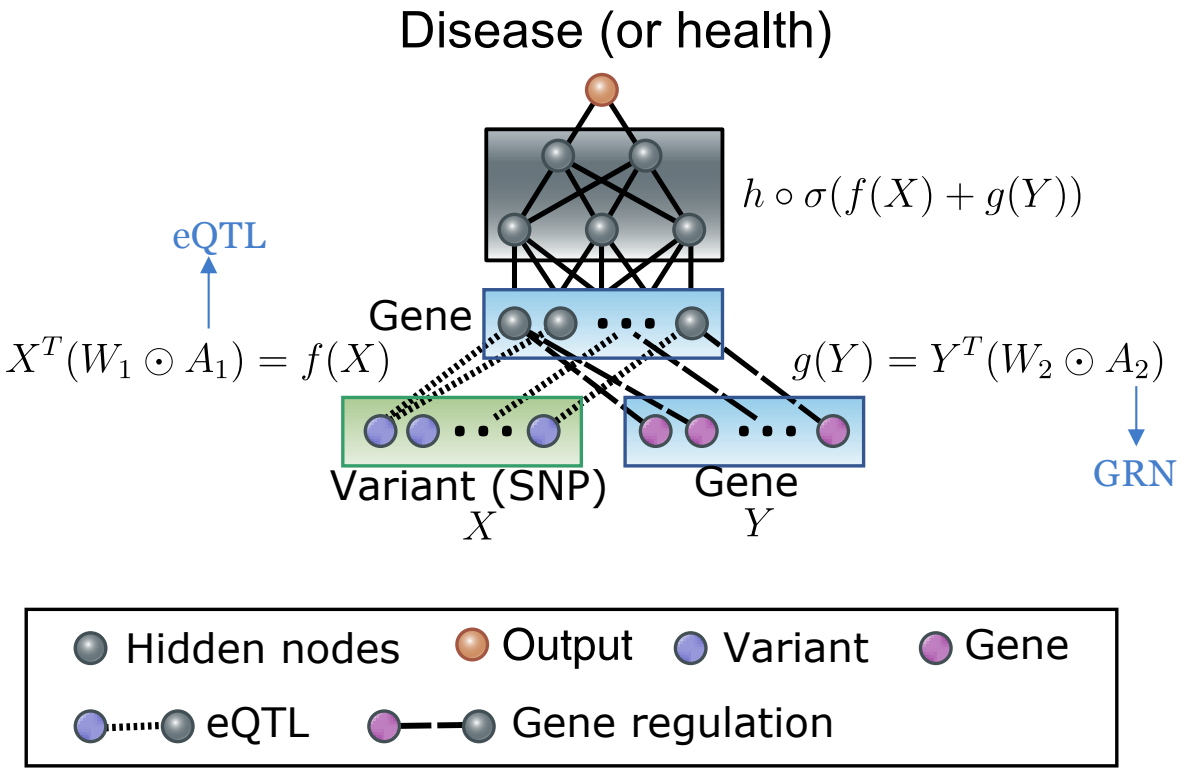
- Background
  - Functional genomics & multi-omics
  - Gene regulation
- Multi-view learning for multi-modal data
  - Multi-view Empirical Risk Minimization (MV-ERM)
  - Manifold alignment
- App 1: *single-cell multi-modal data integration*
  - Predicting neuronal electrophysiology from gene expression
- App 2: *deep manifold-regularized classification*
  - Predicting cortical layers of neurons from multi-modalities
- App 3: *biologically interpretable neural network modeling*
  - Disease prediction and functional prioritization

Machine learning opens “black box” for disease prediction and functional interpretation



Machine Learning

Varmole: Interpretable deep neural network model prioritizes disease variants and genes via DropConnect



- Input form 2 omics, X, Y (SNPs & genes)
- First layer embed A_1 and A_2 – eQTL and gene regulatory network (GRN)

→ From variants (& gene regulations) to gene expression

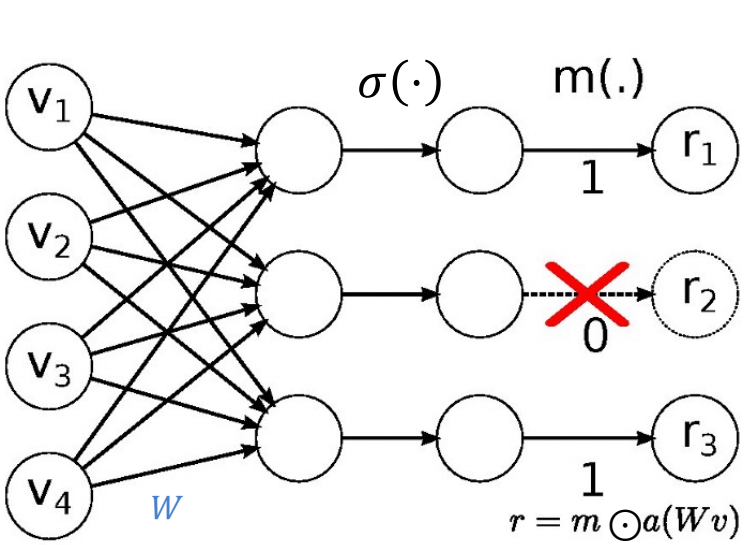
- Other fully connected hidden layers: h;

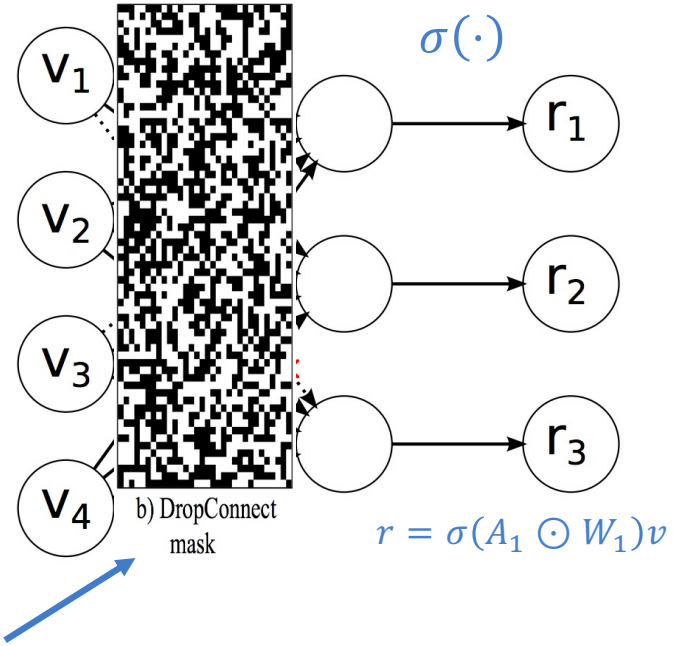
→ From gene expression to phenotypes

- Softmax classification layer: $o = \delta\left(h \circ \sigma\left(f(X) + g(Y)\right)\right)$;
- The Cross-Entropy: $L(o, \hat{o}) = -\frac{1}{n\sum_{i=1}^{n} y_i log(\hat{y}_i)}$
- Varmole: $\min\ L(o, \hat{o}) + \|W\|_1$

Drop-connect

- Drop-out and drop-connect are 2 simple but effective regularization techniques

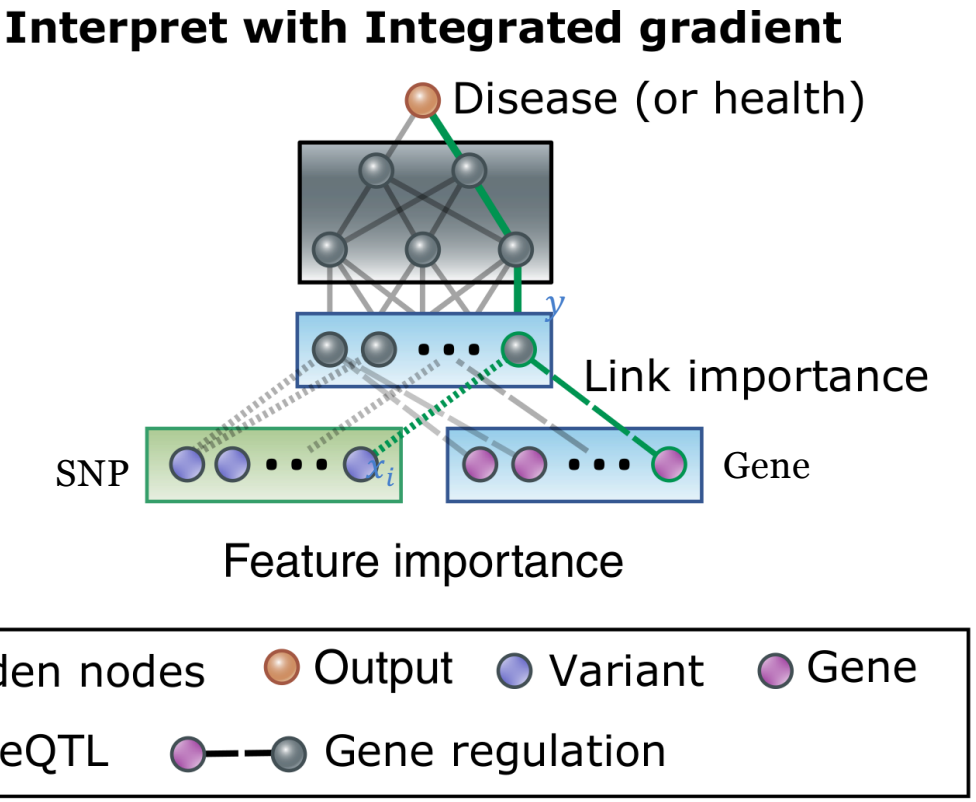




The drop-connect mask is eQTL or GRN (A_1 or A_2)

Interpretation: prioritization via Integrated Gradients

- Given a model F, an input x, and the output $F(x)$ of the model for input in question, an attribution methods returns the 'relevance' of each input feature i to the output
- How importance a gene (SNP, eQTL link, TF-gene link) to the disease outcome



$$\boxed{\text{Cond}_i^y(x)} ::= (x_i - x_i') \cdot \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha(x - x'))}{\partial y} \cdot \frac{\partial y}{\partial x_i} \, d\alpha$$

Importance score of link $x_i \to y$

$$\boxed{\text{IG}_i(x)} ::= (x_i - x_i') \cdot \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} \, d\alpha$$

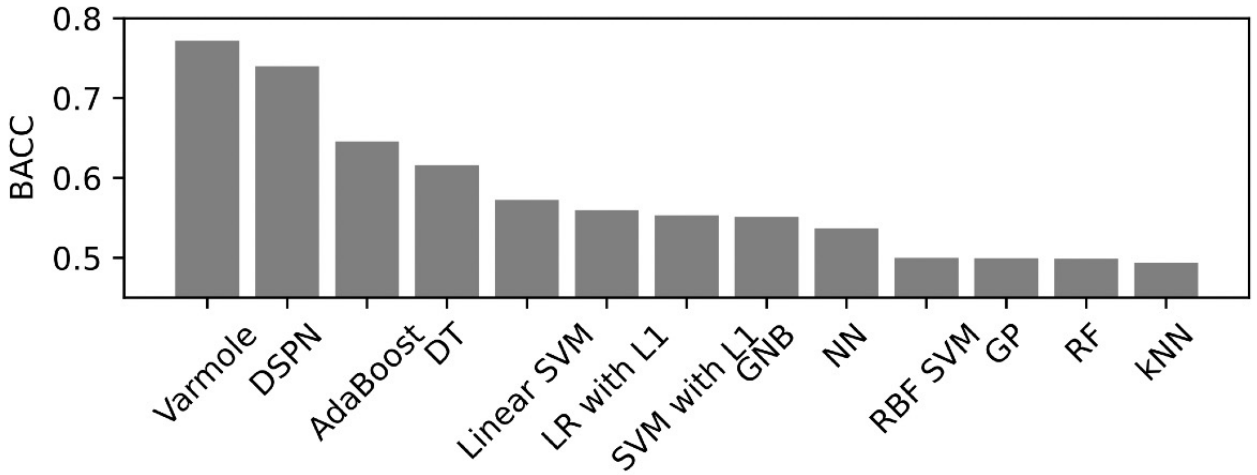
Importance score of feature i of input x

Sundararajan, M., Taly, A., & Yan, Q. (2017). *arXiv preprint arXiv:1703.01365*.

$\frac{\partial F(x)}{\partial x_i}$ is the gradient of F along the i^{th} dimension at x

Application for schizophrenia

- Dataset:
  - RNA-seq gene expression & genotype data (dosage) for 487 schizophrenia (scz) vs. 891 non-scz human brain samples (front cortex)
  - Embedding GTEx eQTLs & PsychENCODE GRN for human brain front cortex
  - → 127304 SNPs, 2598 genes



Nguyen, Jin, Wang, Bioinformatics, 2021

Prioritized gene functions & regulatory links for schizophrenia

- A list of enriched functions (FDR<0.05) from prioritized genes:
  - neuron development
  - axon guidance
  - cell adhesion
  - calcium signaling
  - response to external stimulus
  - NMDA receptor
  - insulin secretion
- Prioritized SNP-gene pairs
  - SNP-gene pairs on the interacting enhancers and promoters (Hi-C) have significantly higher importance scores (p<5e-5)
  - Potential regulatory roles of prioritized SNPs to genes via enhancers

Nguyen, Jin, Wang, Bioinformatics, 2021

Outline

- Background
  - Functional genomics & multi-omics
  - Gene regulation

- Multi-view learning for multi-modal data
  - Multi-view Empirical Risk Minimization (MV-ERM)
  - Manifold alignment

- App 1: *single-cell multi-modal data integration*
  - Predicting neuronal electrophysiology from gene expression
- App 2: *deep manifold-regularized classification*
  - Predicting cortical layers of neurons from multi-modalities
- App 3: *biologically interpretable neural network modeling*
  - Disease prediction and functional prioritization

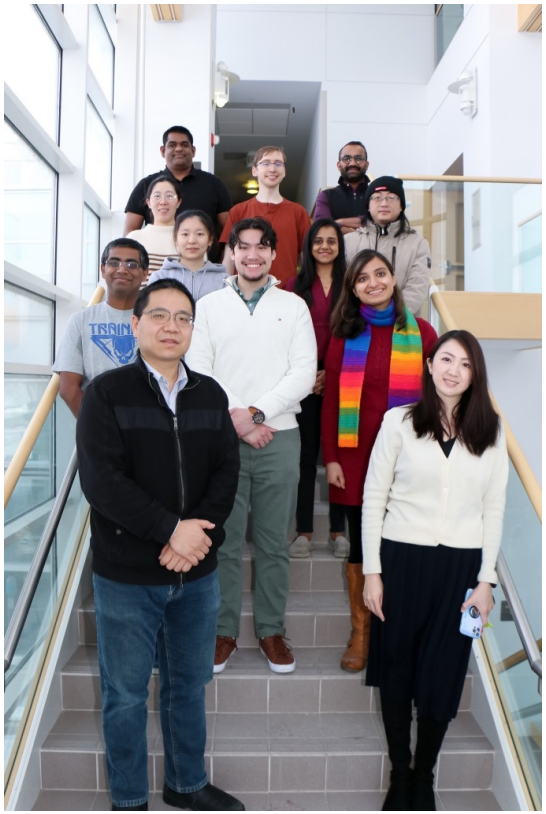
Thank you!

Current lab members

- Dr. Shuang Liu (Data Scientist III)
- Dr. Pramod Chandrashekar (Postdoc)
- Dr. Chirag Gupta (Postdoc)
- Dr. Chenfeng He (Postdoc)
- Dr. Kalpana H. Arachchilage (Postdoc)
- Saniya Khullar (CIBM fellow & Ph.D. candidate, Biomedical Data Science)
- Ting Jin (Ph.D. student, Biomedical Data Science)
- Sayali Alatkar (Ph.D. student, Computer Sciences)
- Noah Cohen Kalafut (Ph.D. student, Computer Sciences)
- Jie Sheng (Biostatistician, Waisman Center)
- Jonathan Bryan (Undergraduate, Neurobiology)

Recent graduates

- Nam Nguyen (Ph.D. 2021, Computer Science, Lane Fellow in Carnegie Mellon University)
- Jiawei Huang (M.S. 2021, Statistics, Ph.D. student in University of Cincinnati)
- Yudi Mu (M.S. 2021, Statistics, Ph.D. student in University of California Riverside)
- Mufang Yin (M.S. 2020, Statistics, Ph.D. student in Rutgers University)
- Peter Rehani (B.S. 2020, Neurobiology, Morgridge Institute for Research)



Funding acknowledgment

- NSF Career 2144475
- NIH R01AG067025
- NIH R21CA237955
- NIH R21NS127432
- NIH R03NS123969

- NIH U01MH116492
- NIH P50HD105353
- NIH R01HD106197
- DOE QPSI

daifeng.wang@wisc.edu
https://daifengwanglab.org/

Various positions available!