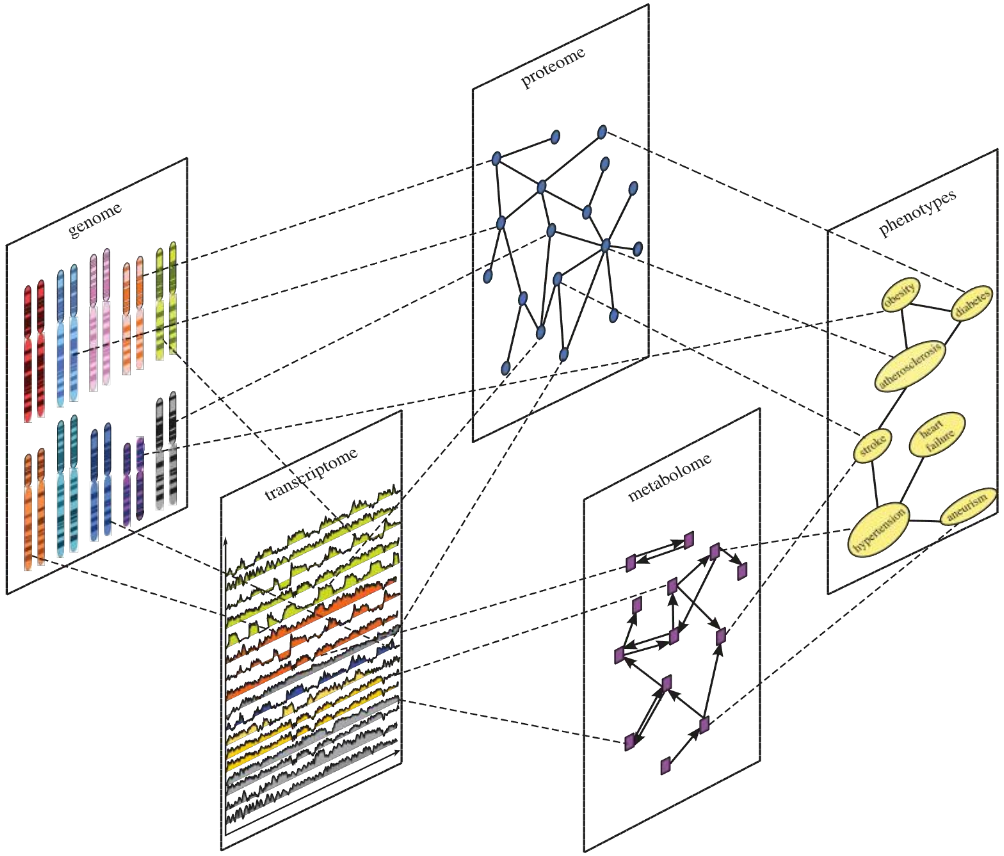
MULTIVIEW LEARNING FOR UNDERSTANDING FUNCTIONAL MULTIOMICS

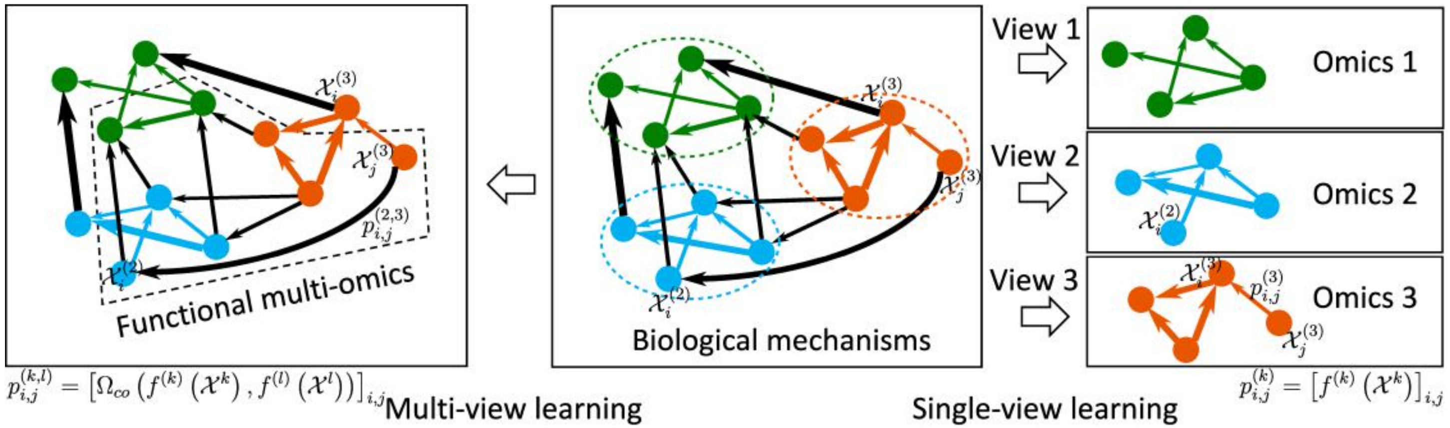
Nam D. Nguyen
Department of Computer Science, Stony Brook University, NY, USA
Daifeng Wang
Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA

MULTIOMICS DATA

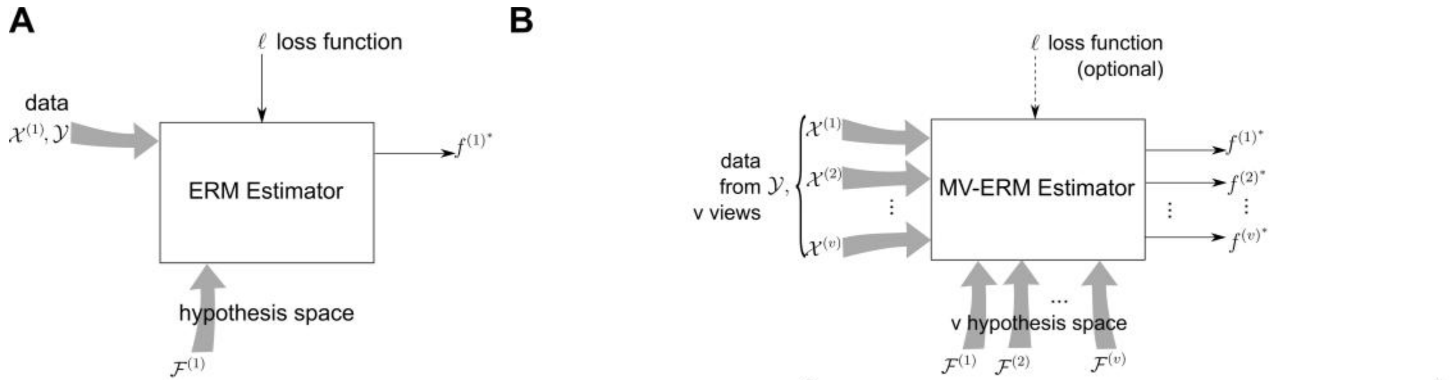


[Gligorijević, Vladimir, and Nataša Pržulj. "Methods for biological data integration: perspectives and challenges." Journal of the Royal Society Interface 12.112 (2015): 20150571.]

MULTIVIEW LEARNING



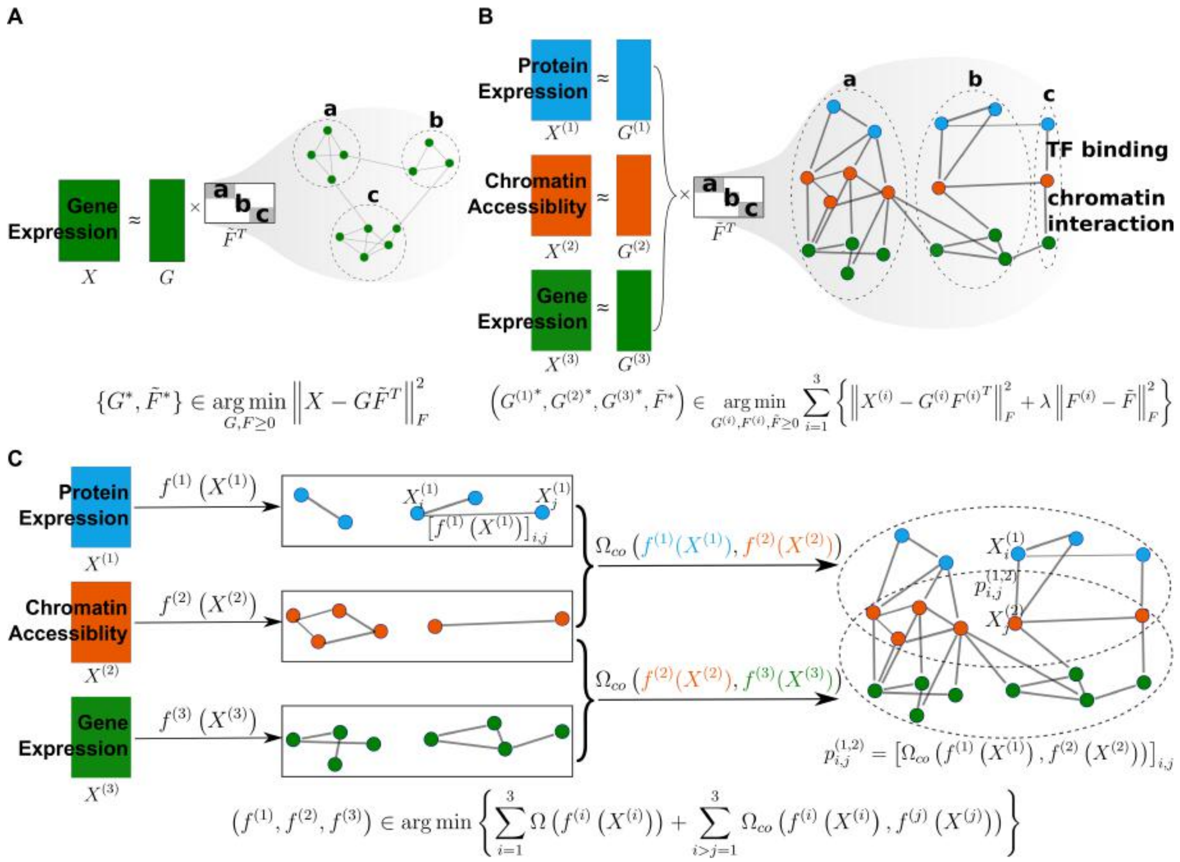
MULTIVIEW EMPIRICAL RISK MINIMIZATION (MV-ERM)



$$f^{(1)^*} \in \arg\min_{f^{(1)} \in \mathcal{F}^{(1)}} \ell\left(f\left(\mathcal{X}^{(1)}\right), \mathcal{Y}\right)$$

$$f^{(1)^*}, f^{(2)^*}, \cdots \in \arg\min_{f^{(i)} \in \mathcal{F}^{(i)}} \left\{ \underbrace{\sum_i \ell\left(f^{(i)}\left(\mathcal{X}^{(i)}\right), \mathcal{Y}\right)}_{\text{optional}} + \underbrace{\lambda \sum_i \Omega\left(f^{(i)}\right)}_{\text{complementary}} + \underbrace{\lambda_{co} \sum_{i,j} \Omega_{co}\left(f^{(i)}, f^{(j)}\right)}_{\text{consensus}} \right\}$$

FACTORIZATION-BASED VS. ALIGNMENT-BASED METHODS



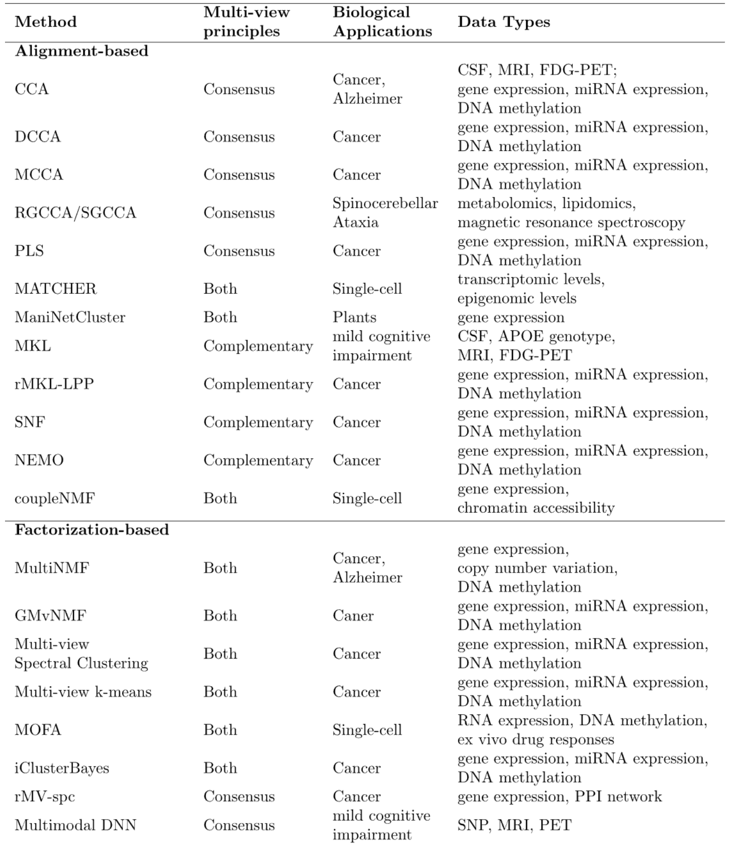
$$\{G^*, \tilde{F}^*\} \in \underset{G,F\geq 0}{\arg\min} \left\| X - G\tilde{F}^T \right\|_F^2$$

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

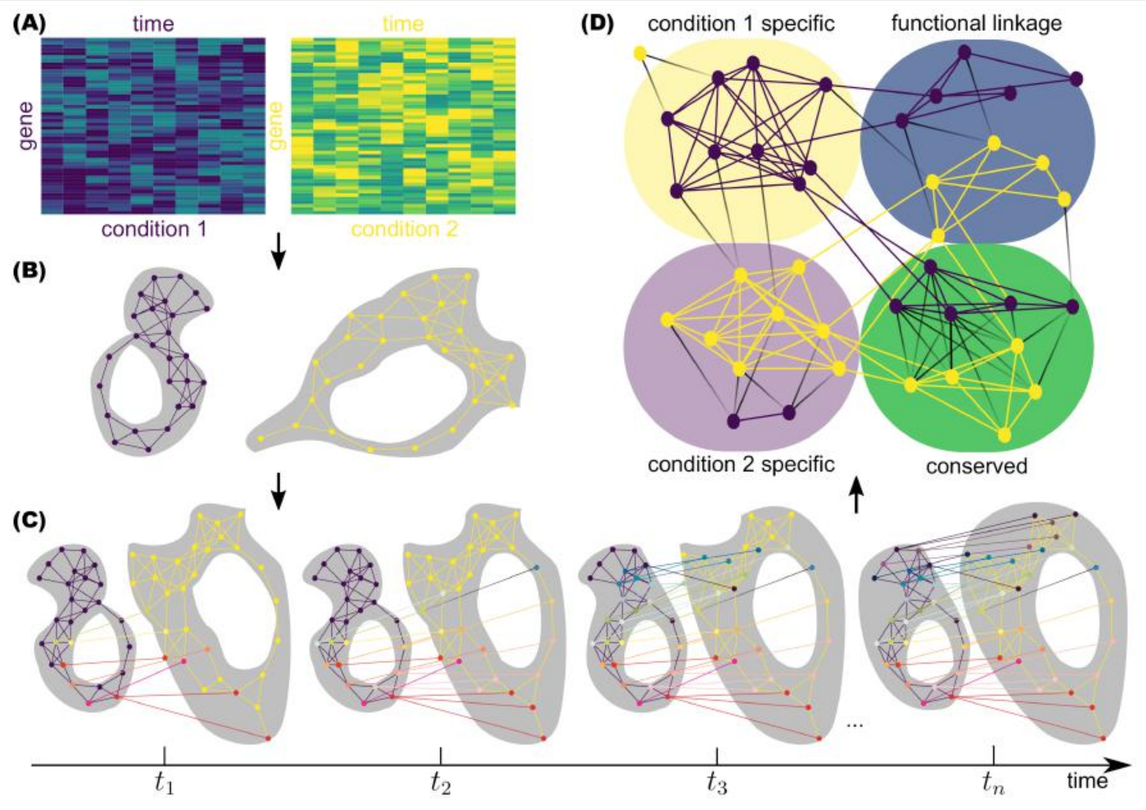
$$p_{i,j}^{(1,2)} = \left[\Omega_{co}\left(f^{(1)}\left(X^{(1)}\right), f^{(2)}\left(X^{(2)}\right)\right)\right]_{i,j}$$

$$\left(f^{(1)}, f^{(2)}, f^{(3)}\right) \in \arg\min \left\{ \sum_{i=1}^{3} \Omega\left(f^{(i)}\left(X^{(i)}\right)\right) + \sum_{i>j=1}^{3} \Omega_{co}\left(f^{(i)}\left(X^{(i)}\right), f^{(j)}\left(X^{(j)}\right)\right) \right\}$$

STATE-OF-THE-ART METHODS AND APPLICATIONS



Method	Multi-view principles	Biological Applications	Data Types
Alignment-based			
CCA	Consensus	Cancer, Alzheimer	CSF, MRI, FDG-PET; gene expression, miRNA expression, DNA methylation
DCCA	Consensus	Cancer	gene expression, miRNA expression, DNA methylation
MCCA	Consensus	Cancer	gene expression, miRNA expression, DNA methylation
RGCCA/SGCCA	Consensus	Spinocerebellar Ataxia	metabolomics, lipidomics, magnetic resonance spectroscopy
PLS	Consensus	Cancer	gene expression, miRNA expression, DNA methylation
MATCHER	Both	Single-cell	transcriptomic levels, epigenomic levels
ManiNetCluster	Both	Plants	gene expression
MKL	Complementary	mild cognitive impairment	CSF, APOE genotype, MRI, FDG-PET
rMKL-LPP	Complementary	Cancer	gene expression, miRNA expression, DNA methylation
SNF	Complementary	Cancer	gene expression, miRNA expression, DNA methylation
NEMO	Complementary	Cancer	gene expression, miRNA expression, DNA methylation
coupleNMF	Both	Single-cell	gene expression, chromatin accessibility
Factorization-based			
MultiNMF	Both	Cancer, Alzheimer	gene expression, copy number variation, DNA methylation
GMvNMF	Both	Caner	gene expression, miRNA expression, DNA methylation
Multi-view Spectral Clustering	Both	Cancer	gene expression, miRNA expression, DNA methylation
Multi-view k-means	Both	Cancer	gene expression, miRNA expression, DNA methylation
MOFA	Both	Single-cell	RNA expression, DNA methylation, ex vivo drug responses
iClusterBayes	Both	Cancer	gene expression, miRNA expression, DNA methylation
rMV-spc	Consensus	Cancer	gene expression, PPI network
Multimodal DNN	Consensus	mild cognitive impairment	SNP, MRI, PET



[Nguyen, Nam D., Ian K. Blaby, and Daifeng Wang. "ManiNetCluster: a novel manifold learning approach to reveal the functional links between gene networks." BMC genomics 20.12 (2019): 1-14.]

SUMMARY

- provided the formal framework for categorizing current multiview learning methods
- can also serve as a guideline for developing many new methods
- demonstrated that the biological applications of these methods are thriving and promising because of the growing use of multiomics data
- Biological problems always involve of many diverse facets, and multiview learning is an efficient strategy for tackling those problems.

REFERENCES

- **Nguyen, Nam D., and Daifeng Wang. "Multiview learning for understanding functional multiomics." PLOS Computational Biology 16.4 (2020): e1007677.**
- Nguyen, Nam D., Ian K. Blaby, and Daifeng Wang. "ManiNetCluster: a novel manifold learning approach to reveal the functional links between gene networks." BMC genomics 20.12 (2019): 1-14.
- Gligorijević, Vladimir, and Nataša Pržulj. "Methods for biological data integration: perspectives and challenges." Journal of the Royal Society Interface 12.112 (2015): 20150571.

THANK YOU!