Applied Machine Learning Part II

BMI/CS 776

www.biostat.wisc.edu/bmi776/

Spring 2021

Daifeng Wang

daifeng.wang@wisc.edu

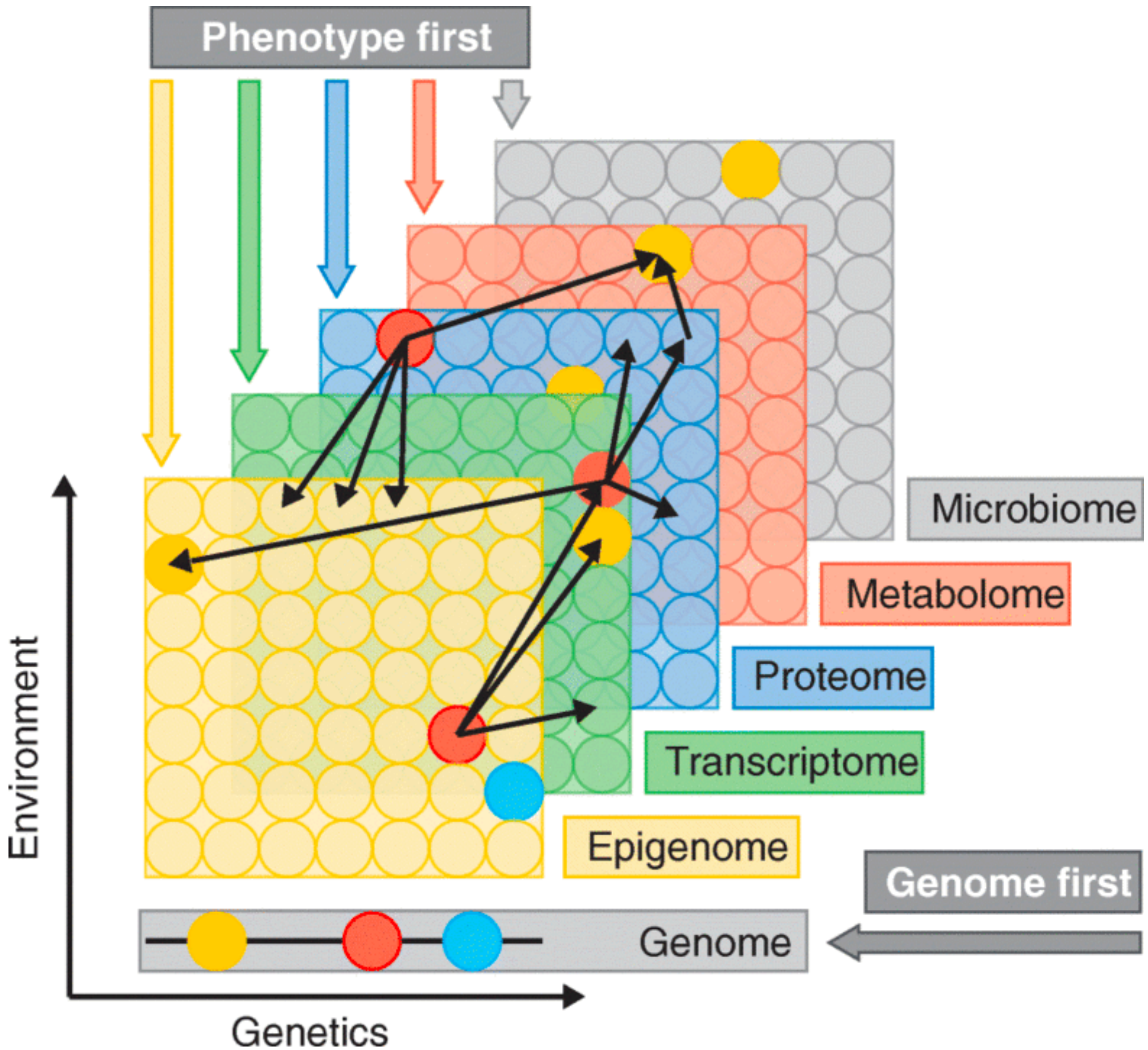
Goals for lecture

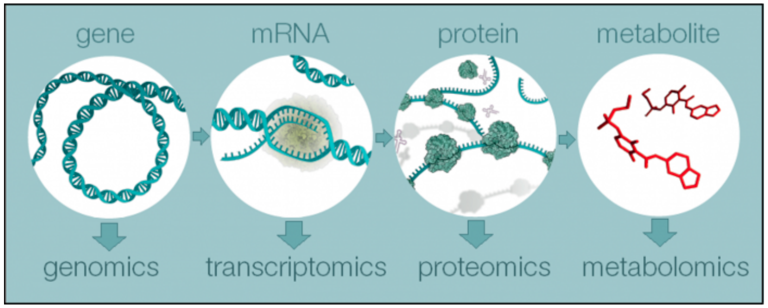
- Multi-omics data
- Machine learning modeling
  - Empirical risk minimization (ERM)
- Multi-layer network clustering
- Dimensionality reduction & Spectral methods
- Decision tree
- Neural network

Goals for lecture

- Multi-omics data
- Machine learning modeling
  - Empirical risk minimization (ERM)
- Multi-layer network clustering
- Dimensionality reduction & Spectral methods
- Decision tree
- Neural network

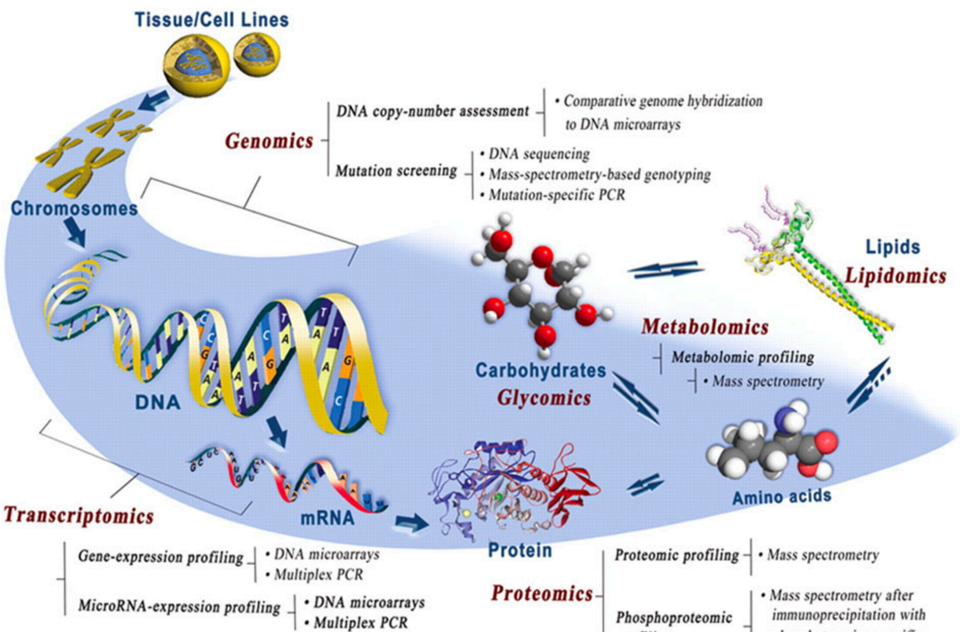
Multi-omics





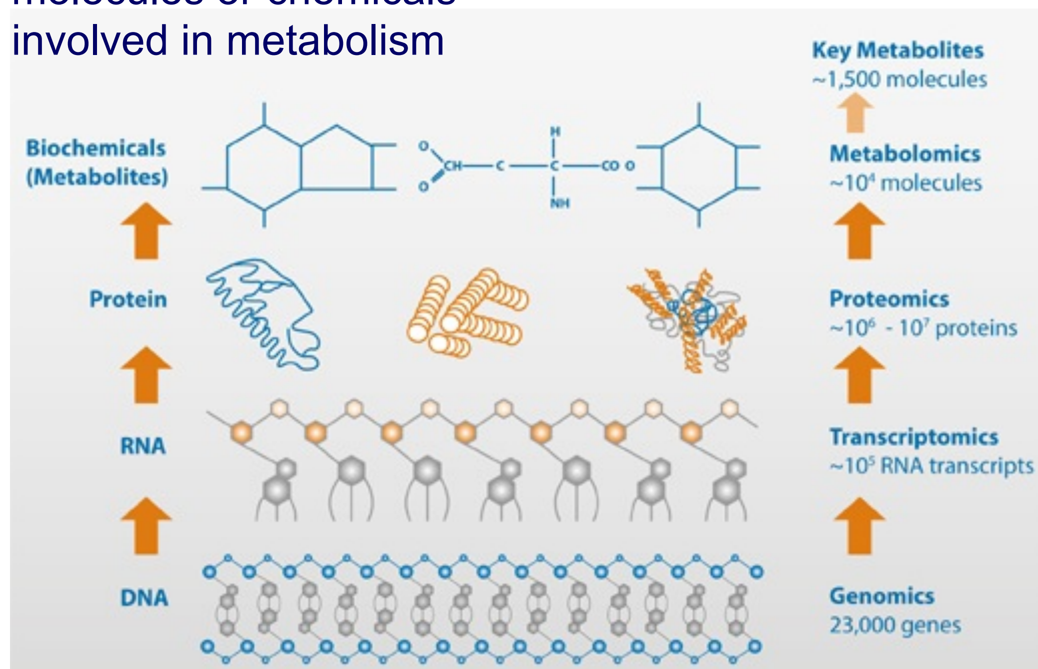
https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1215-1

Metabolites and Metabolomics

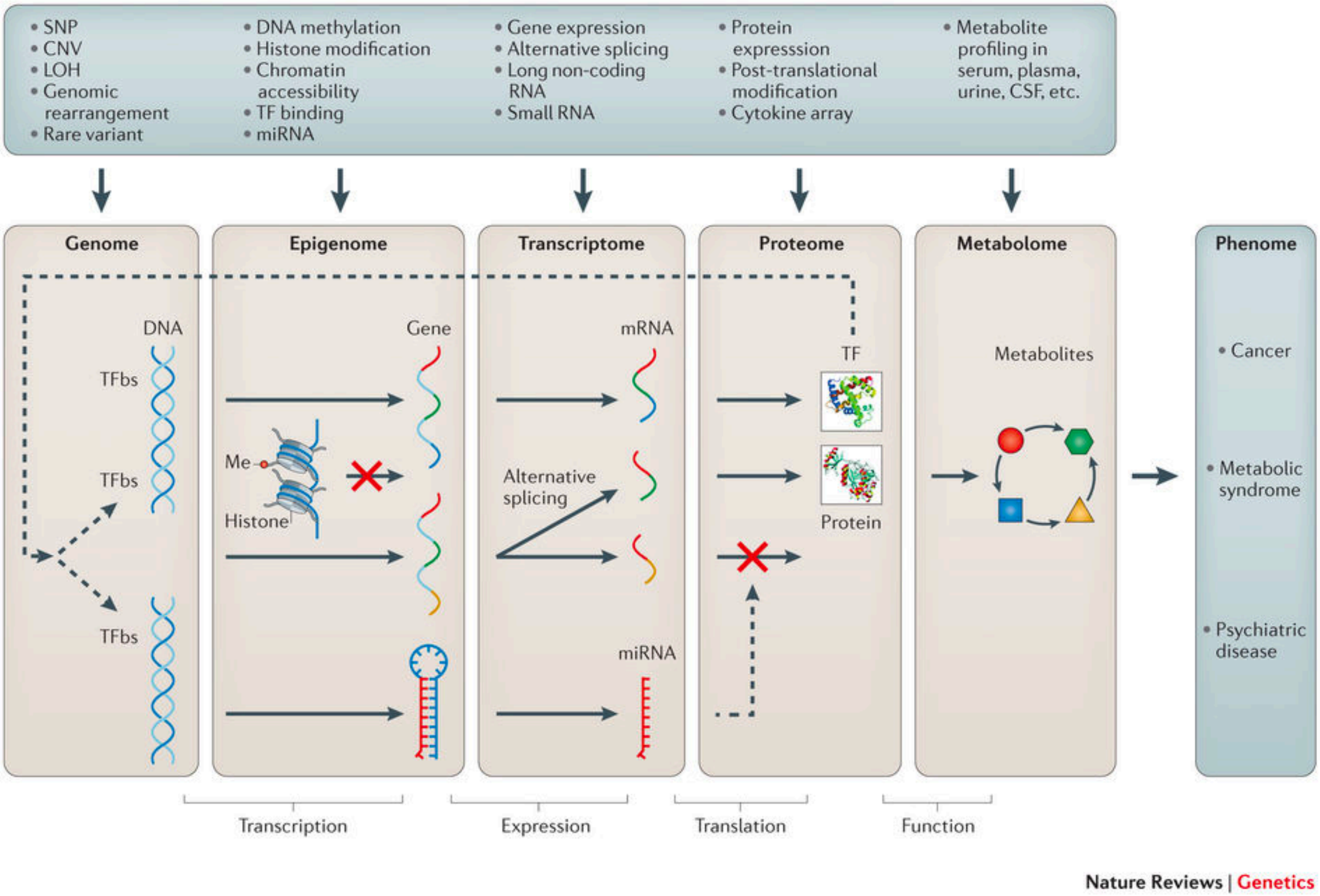


Wu RQ, J Dent Res. 2011

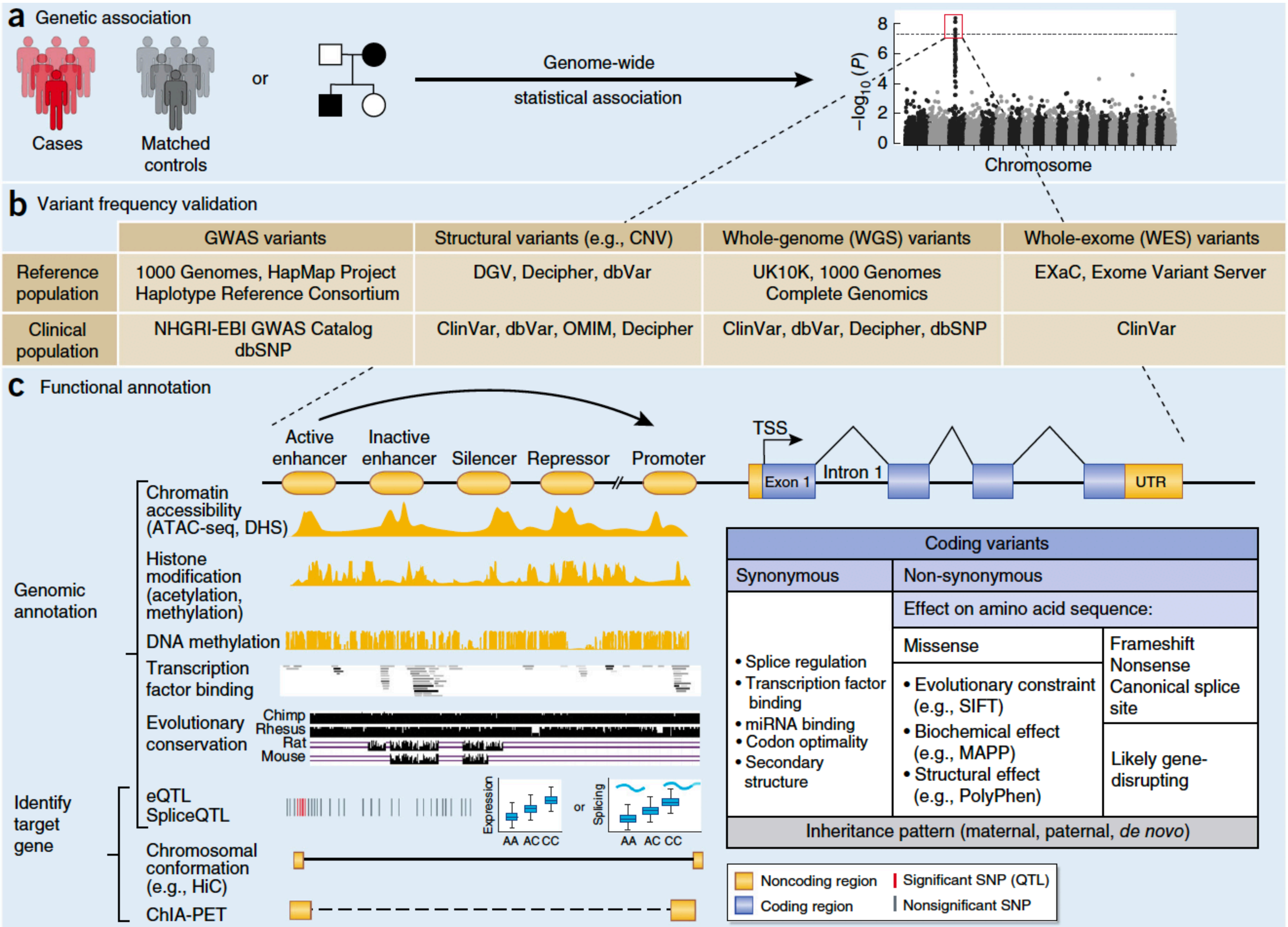
Metabolites are small molecules or chemicals involved in metabolism



Multi-scale mechanisms



Functional genomics to understand mechanisms

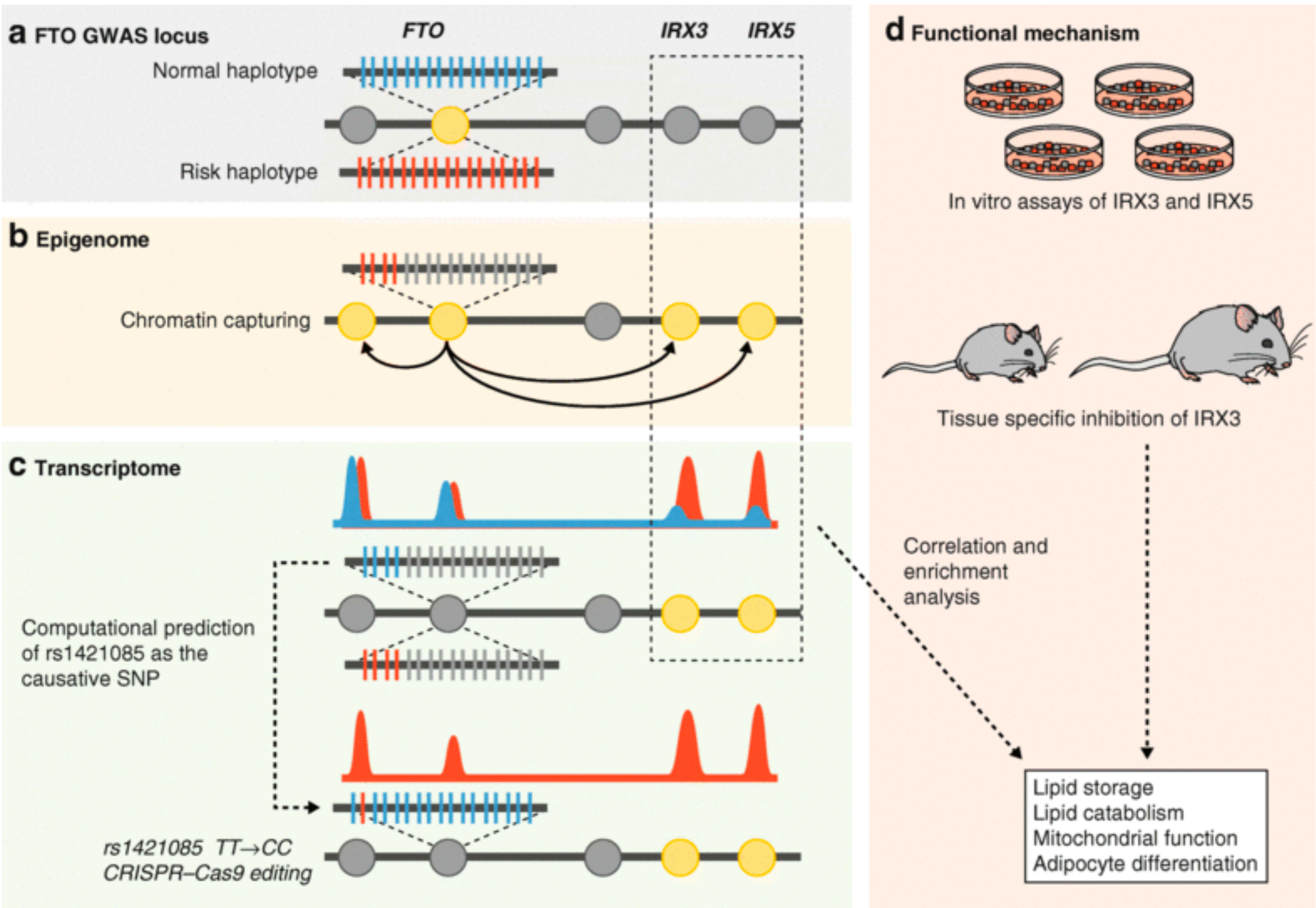


Gandal et al., Nature Neuroscience, 2016

Disease-associated genomic variants

How do variants function?

Example



N Engl J Med 2015; 373:895-907
DOI: 10.1056/NEJMoa1502214

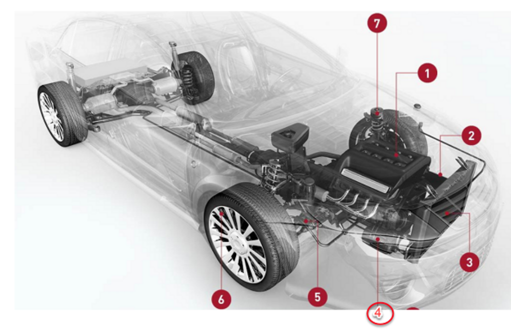
Hierarchical understanding from genotype to phenotype

Elements
- variants
- genes
- regulatory regions

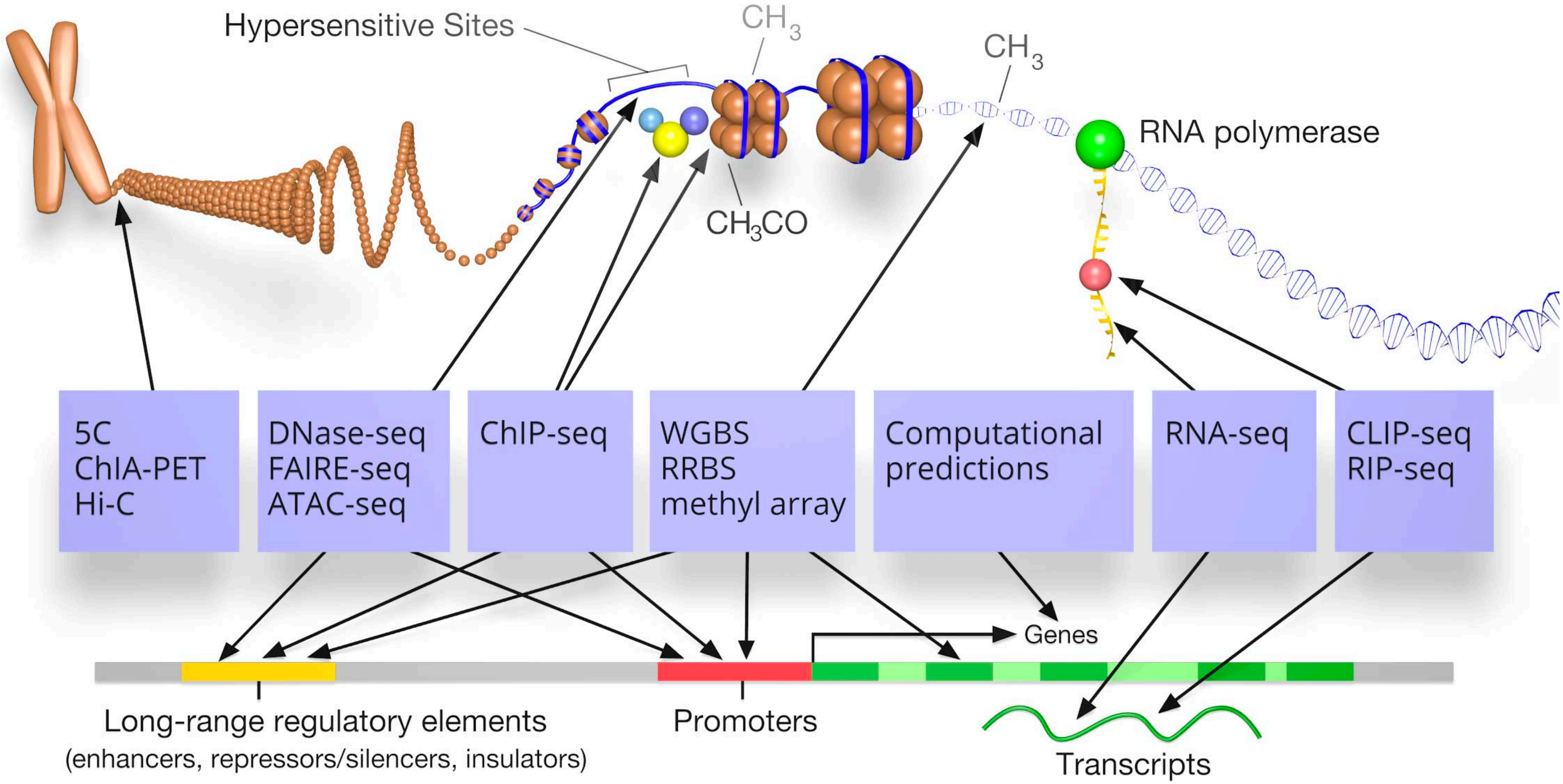
Interactions
- gene regulation
- chromatin interaction
- TF binding

Mechanisms
- pathways
- circuits
- functions

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



Multi-omics for understanding functional genomics and gene regulation



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

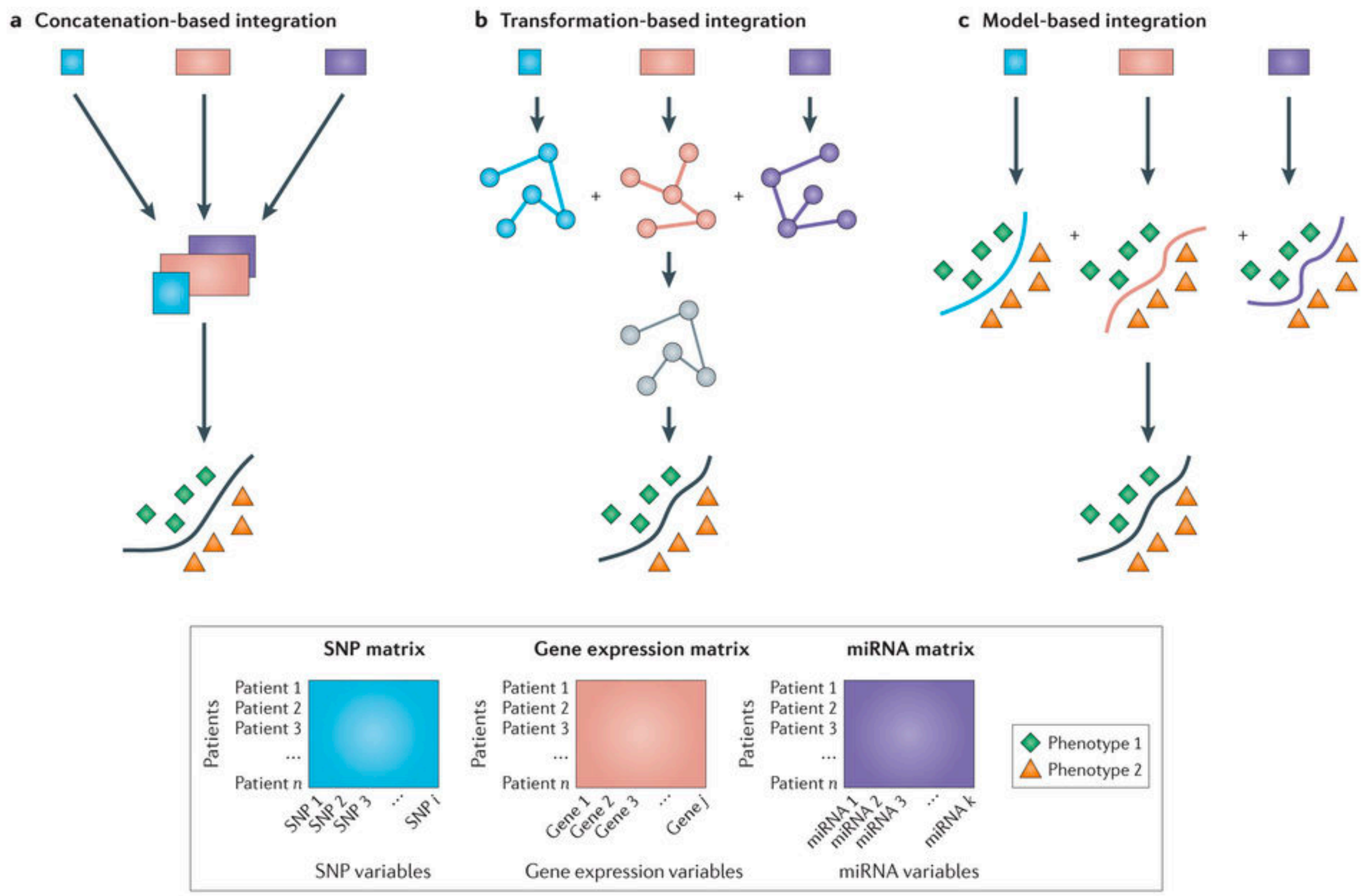
Some multi-omics datasets

Human	20,000 genes (2% genome)	Other genomic elements: non-coding RNAs, gene regulatory regions, repeats, and so on… (98% genome)
Cell lines	ENCODE (Encyclopedia of DNA Elements) Consortium (> 300 cell types)	National Human Genome Research Institute ENCODE
Tissues	GTEx PROJECT	Genotype-Tissue Expression (GTEx) (> 40 tissues)
Cancers	THE CANCER GENOME ATLAS National Cancer Institute National Human Genome Research Institute	The Cancer Genome Atlas (TCGA) (> 40 cancer types)
Development	BRAINSPAN ATLAS OF THE DEVELOPING HUMAN BRAIN	(13 developmental stages, 16 brain regions)
Psychiatric disorders	PsychENCODE	PsychENCODE Consortium (~2,000 tissues incl. health, Schizophrenia, Autism, Bipolar)
Neurodegenerative diseases	ADGC Alzheimer's Disease Genetics Consortium	Religious Orders Study and Memory and Aging Project (ROSMAP); International Parkinson's Disease Genomics Consortium (IPDGC)

Goals for lecture

- Multi-omics data
- Machine learning modeling
  - Empirical risk minimization (ERM)
- Multi-layer network clustering
- Dimensionality reduction & Spectral methods
- Decision tree
- Neural network

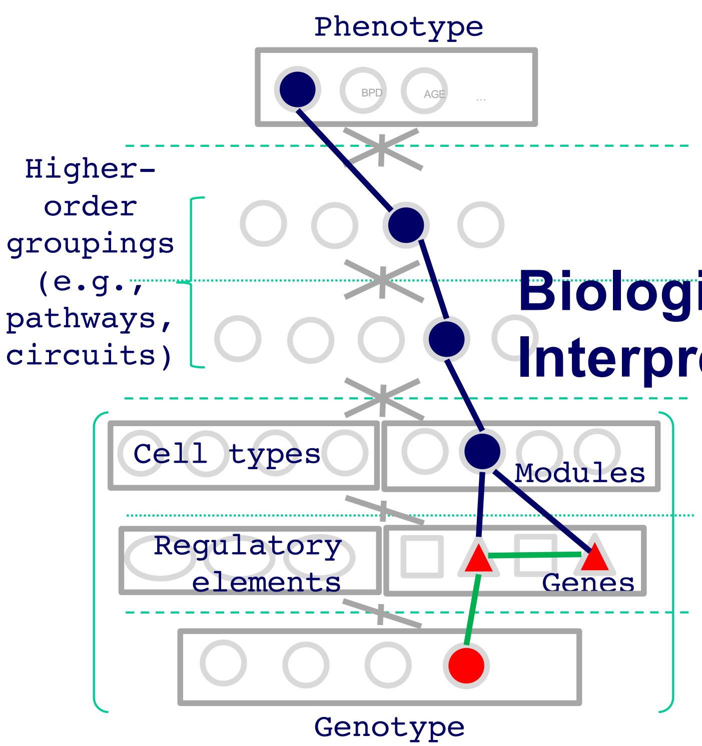
Multi-omics data integration



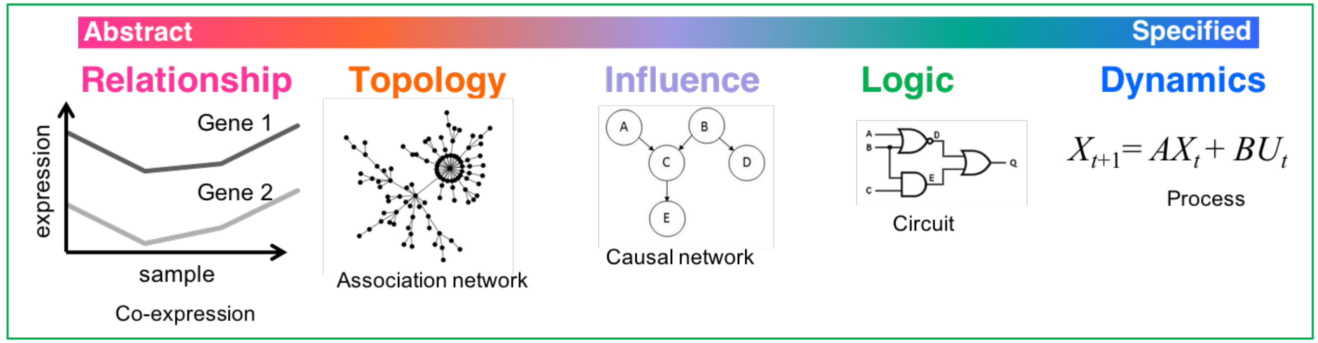
Nature Reviews | Genetics

http://www.nature.com/nrg/journal/v16/n2/full/nrg3868.html

Multi-omics data modeling



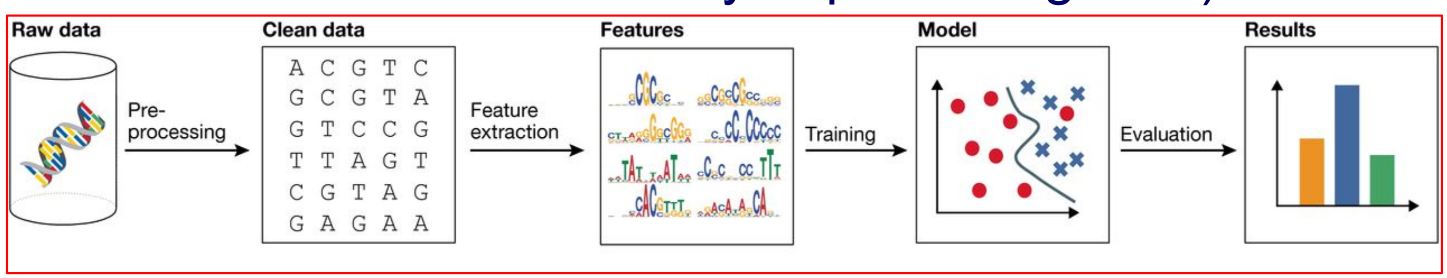
Biological Interpretation



Interactions between elements (e.g., co-expression, regulation)

Integrative & Predictive model (e.g., deep neural network)

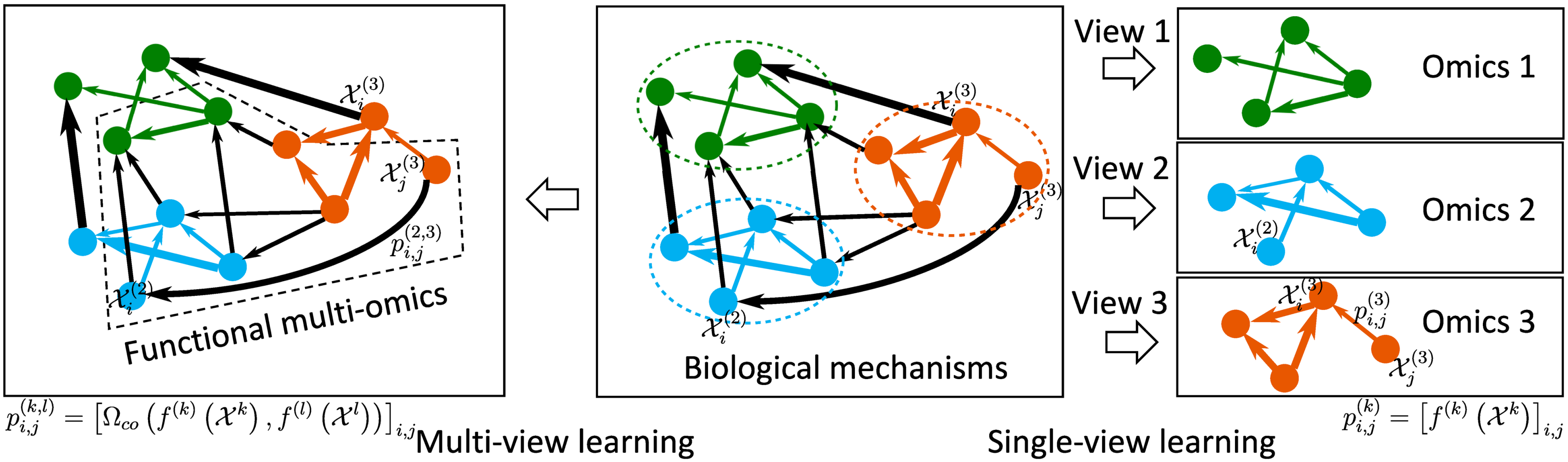
Genomic elements (e.g., GWAS, differentially expressed genes)



* Spectrum inspired from Ideker & Lauffenburger, Trends in Biotechnology, 2003

* Christof Angermueller et al. Mol Syst Biol 2016;12:878

Multiview learning for understanding functional multi-omics



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

Cross-omics interactions

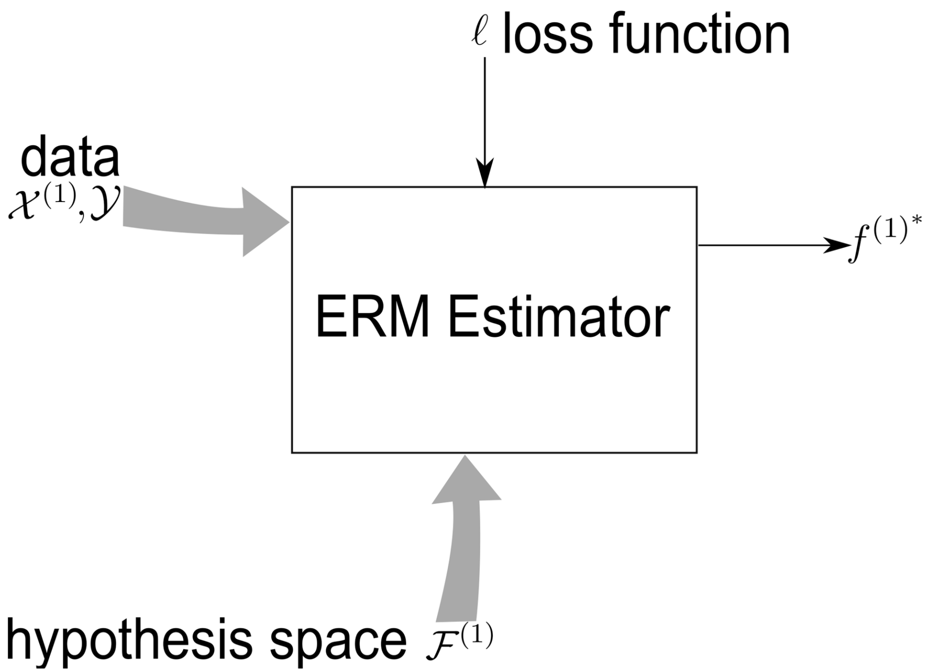
$\Omega_{co}(f^{(1)}, f^{(3)})$: SNPs break TF binding sites

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

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

Nguyen, Wang, PLoS Computational Biology, 2020

Empirical risk minimization (ERM) for machine learning modeling

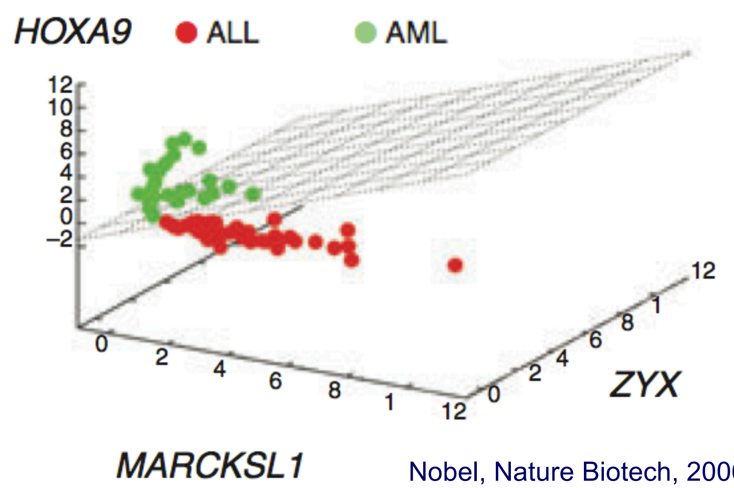


$$\mathrm{R}(f) = \frac{1}{|S|} \sum_{(x_i, y_i) \in S} \ell(f(x_i), y_i)$$

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

Regularize f by biological knowledge Ω

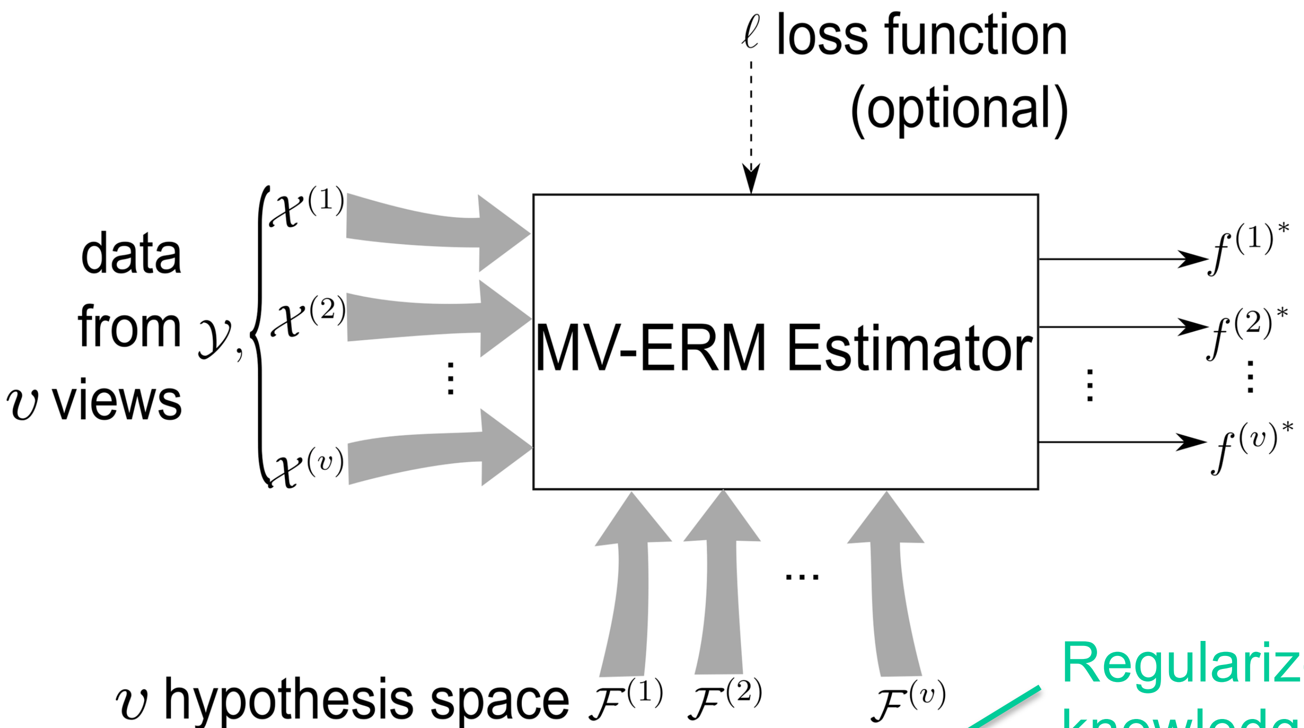
- e.g., Leukemia patient classification
  - y_i: Acute lymphoblastic leukemia (ALL) vs. Acute myeloid leukemia (AML)
  - x_i: gene expression
  - f: SVM



Nobel, Nature Biotech, 2006

Nguyen, Wang, PLoS Computational Biology, 2020

Empirical risk minimization for multi-view learning (MV-ERM)



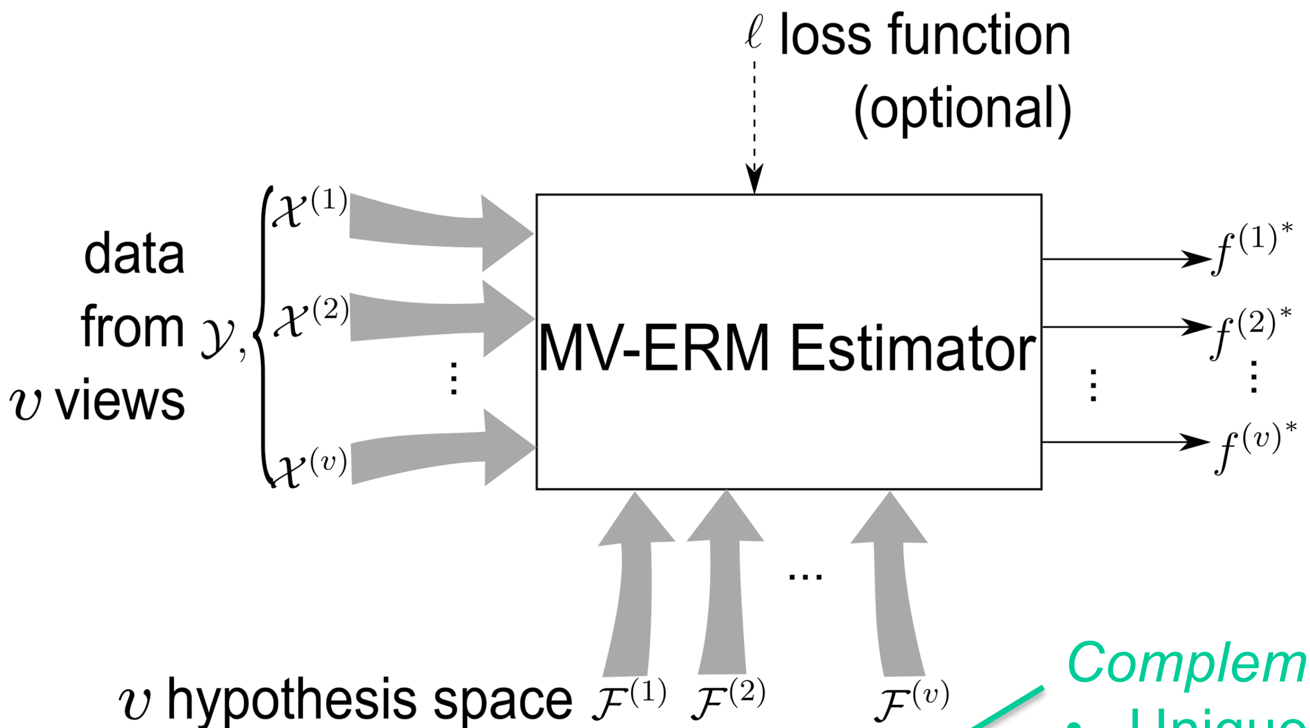
$$f^{(1)^*}, f^{(2)^*}, \cdots \in \underset{f^{(i)} \in \mathcal{F}^{(i)}}{\operatorname{argmin}} \Bigg\{ \underbrace{\sum_i \ell\left(f^{(i)}\left(\mathcal{X}^{(i)}\right), \mathcal{Y}\right)}_{\text{optional}} + \underbrace{\lambda \sum_i \Omega\left(f^{(i)}\right)}_{\text{complementary}} + \underbrace{\lambda_{co}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

Regularize f by biological knowledge Ω from single omics

Regularize f by biological knowledge Ω_{co} across multi-omics

Nguyen, Wang, PLoS Computational Biology, 2020

Consensus and complementary principles



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

Complementary principle

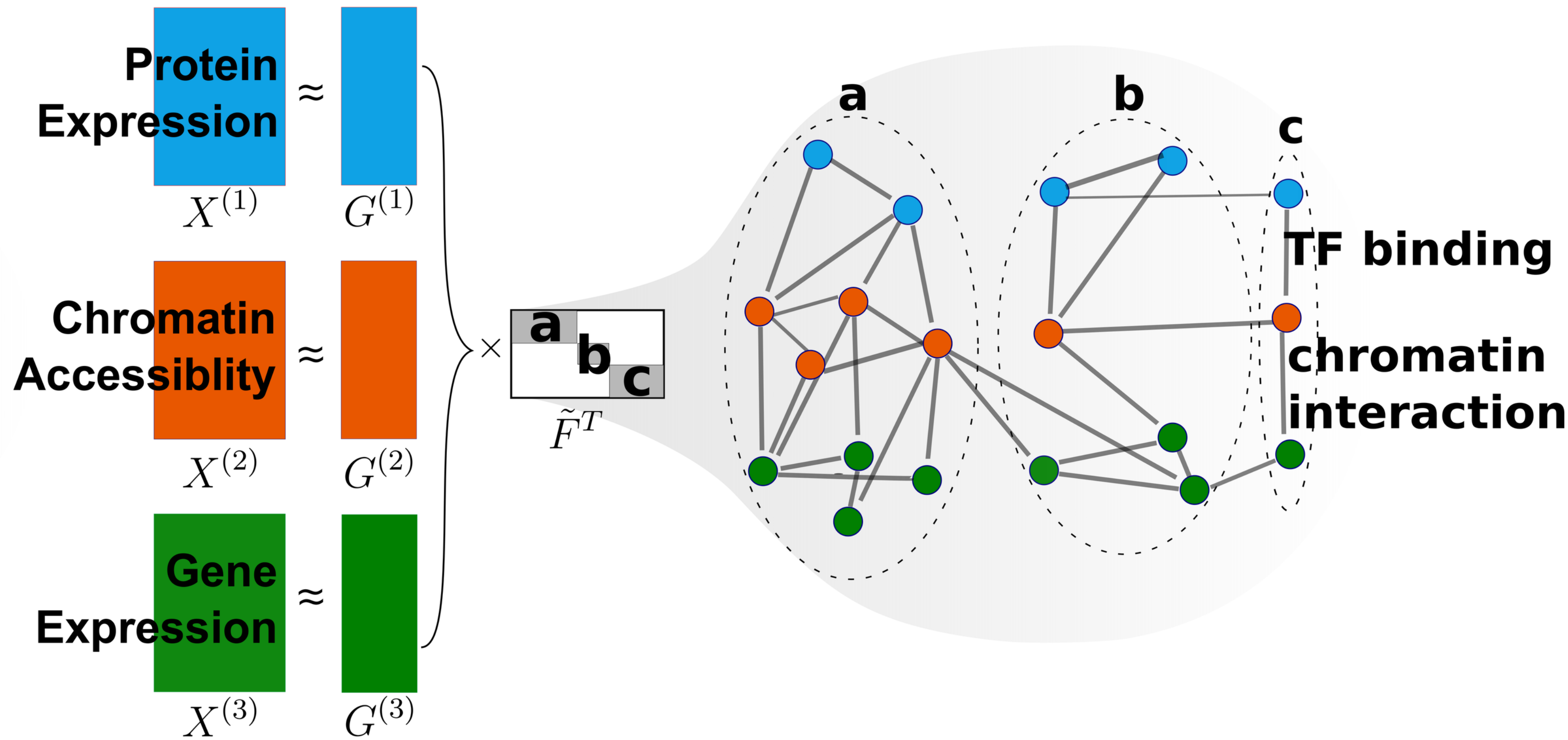
- Unique information from each view

Consensus principle

- Relationship information across views

Nguyen, Wang, PLoS Computational Biology, 2020

Factorization-based MV-ERM framework

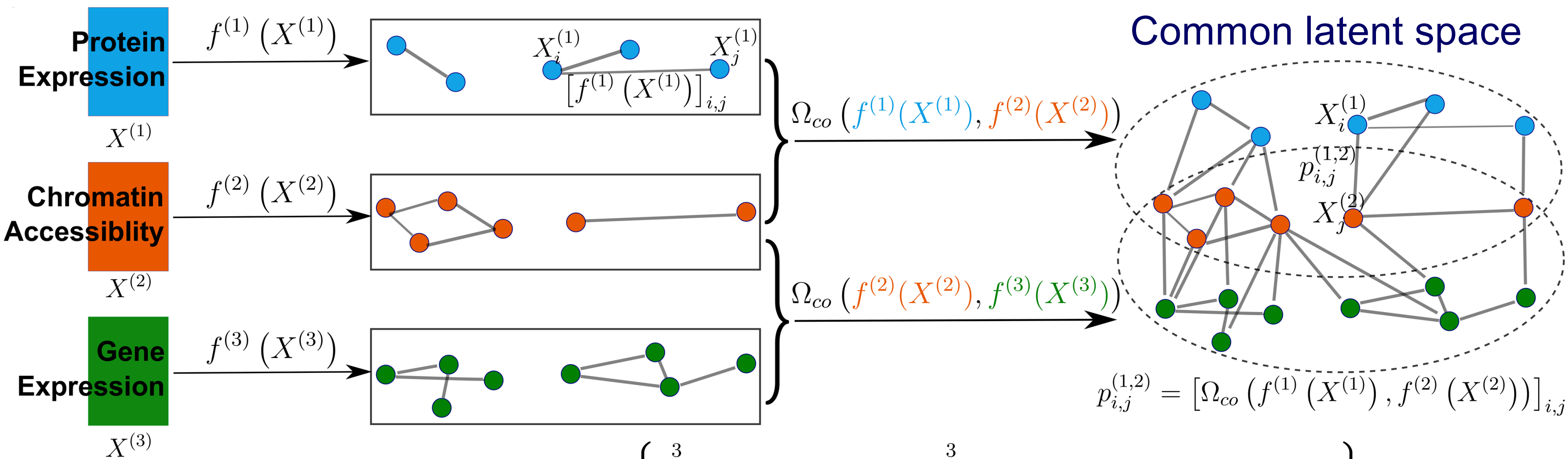


$$\left(G^{(1)*}, G^{(2)*}, G^{(3)*}, \tilde{F}^{*}\right) \in \underset{G^{(i)}, F^{(i)}, \tilde{F} \geq 0}{\arg\min} \sum_{i=1}^{3} \left\{ \left\| X^{(i)} - G^{(i)} F^{(i)^T} \right\|_F^2 + \lambda \left\| F^{(i)} - \tilde{F} \right\|_F^2 \right\}$$

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

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

Alignment-based MV-ERM



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

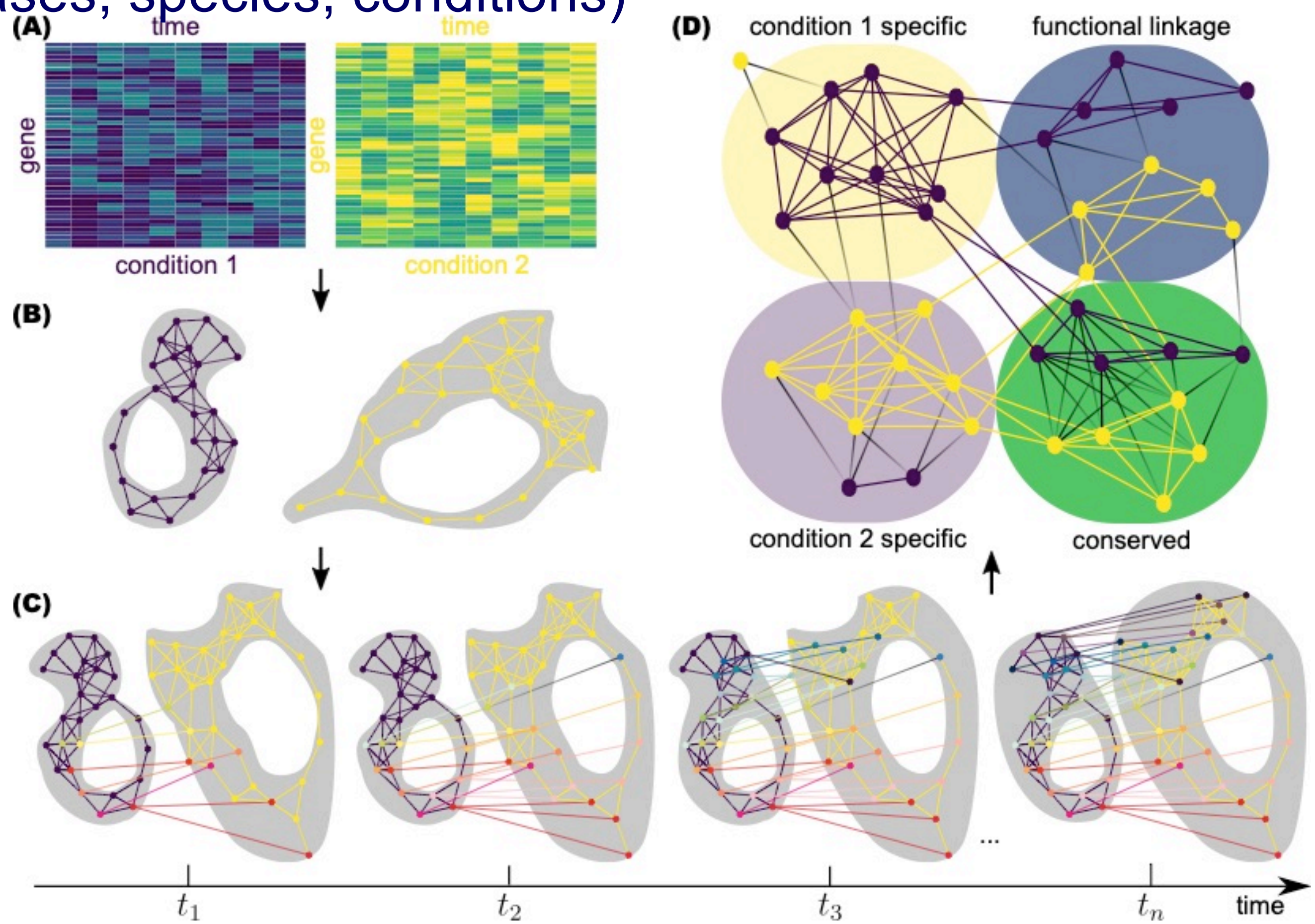
- For instance, Canonical correlation analysis (CCA)
  - Consensus only
  - $\Omega_{co}(.) = -tr(F_1^T X_1 X_2^T F_2)$ for two views X_1 and X_2 with linear projections F_1 and F_2

Nguyen, Wang, PLoS Computational Biology, 2020

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

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

Functional linkages across dimensions

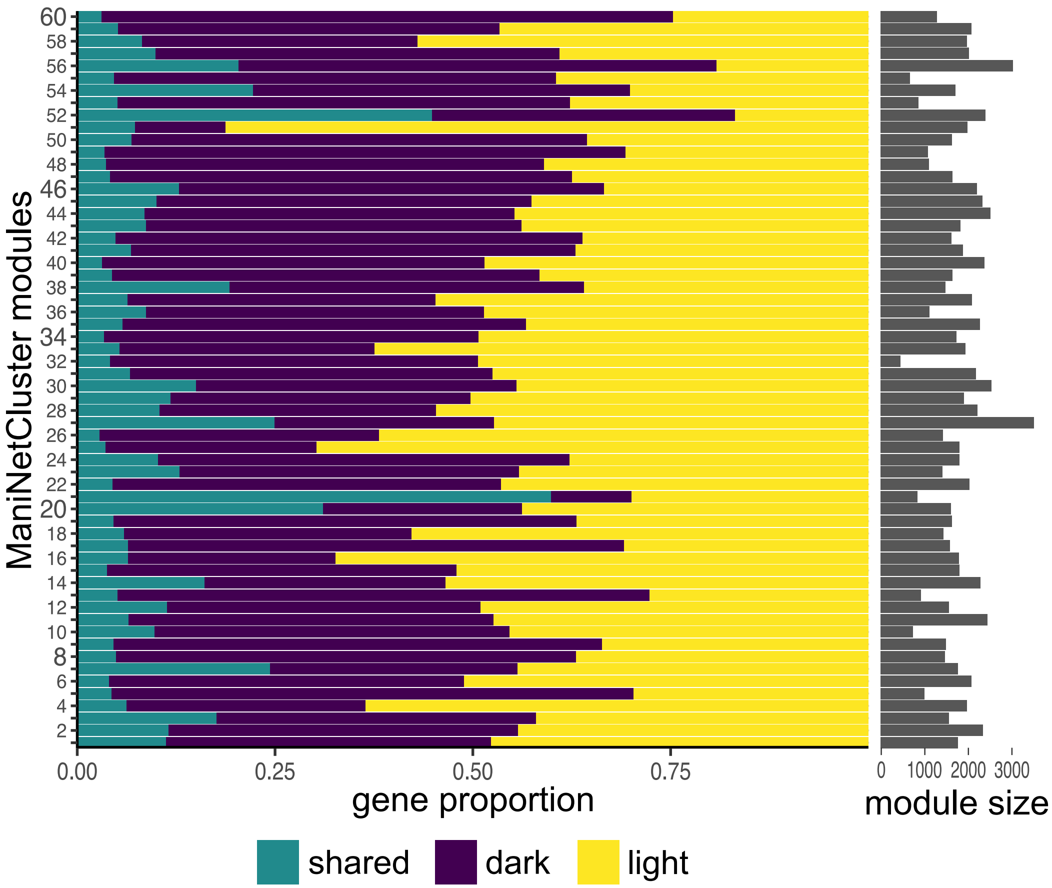


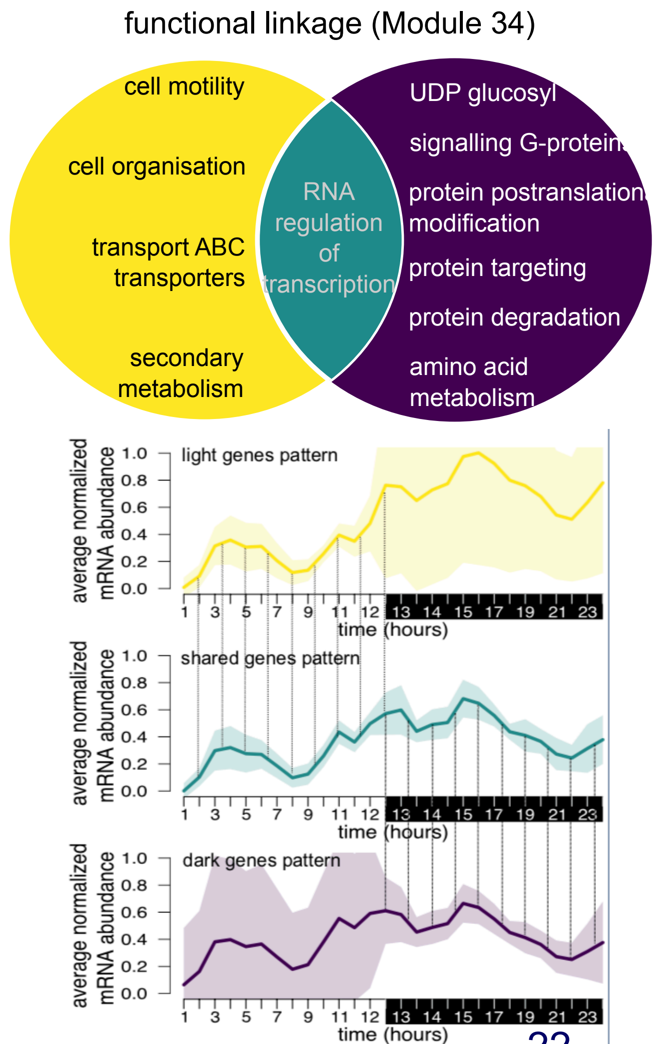
$$argmin_{f_x,f_y}\ \lambda \sum_{i,j} \|f_x(X_i) - f_x(X_j)\|^2 S_x(i,j) + \lambda \sum_{i,j} \|f_y(Y_i) - f_y(Y_j)\|^2 S_y(i,j) + (1-\lambda) \sum_{i,j} \|f_x(X_i) - f_y(Y_j)\|^2 W(i,j)$$

Nguyen, Blaby, Wang, BMC Genomics, 2019

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

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





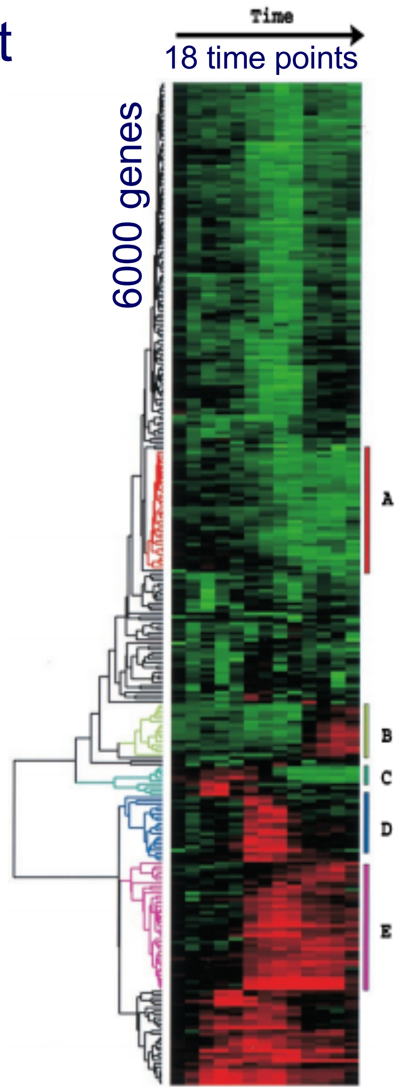
Nguyen, Blaby, Wang, BMC Genomics, 2019

Goals for lecture

- Multi-omics data
- Machine learning modeling
  - Empirical risk minimization (ERM)
- Multi-layer network clustering
- Dimensionality reduction & Spectral methods
- Decision tree
- Neural network

Co-expressed genes have similar functions in single species

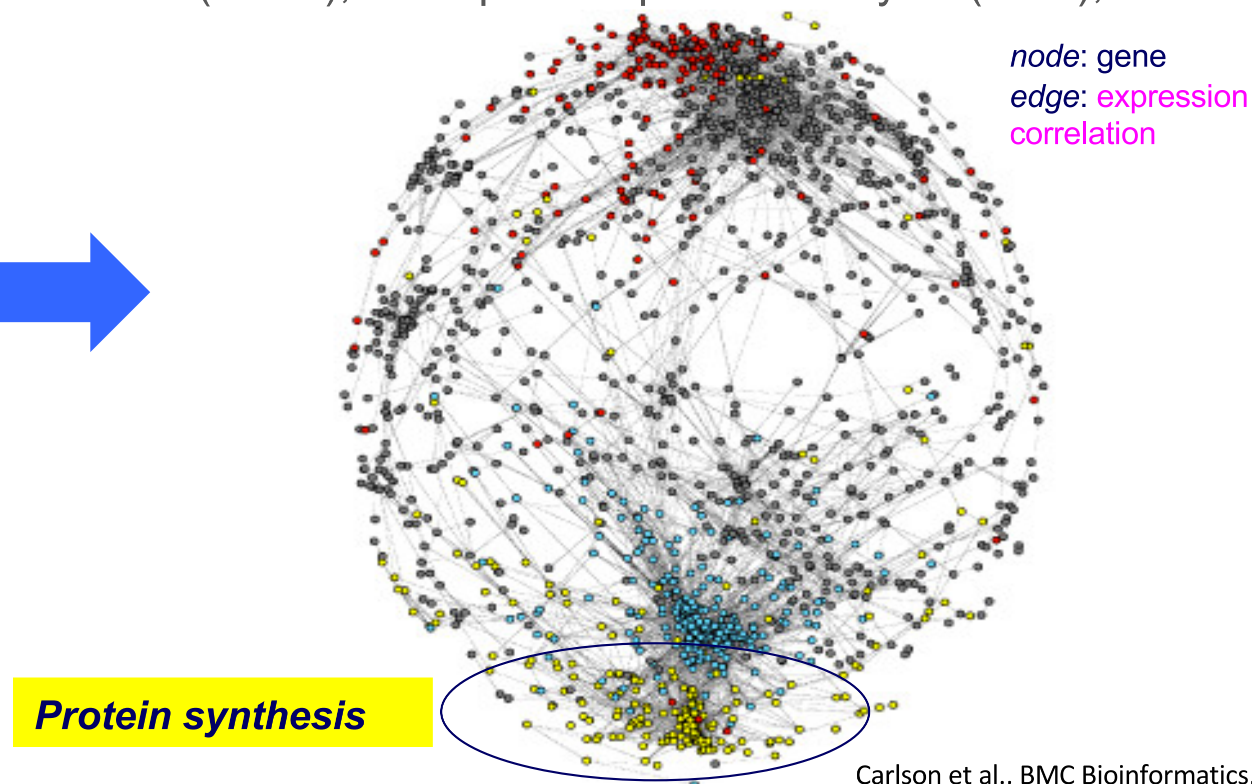
Yeast cell cycle



Eisen et al., PNAS, 1998.

A gene co-expression network (relationship) can reveal ***functional groupings***

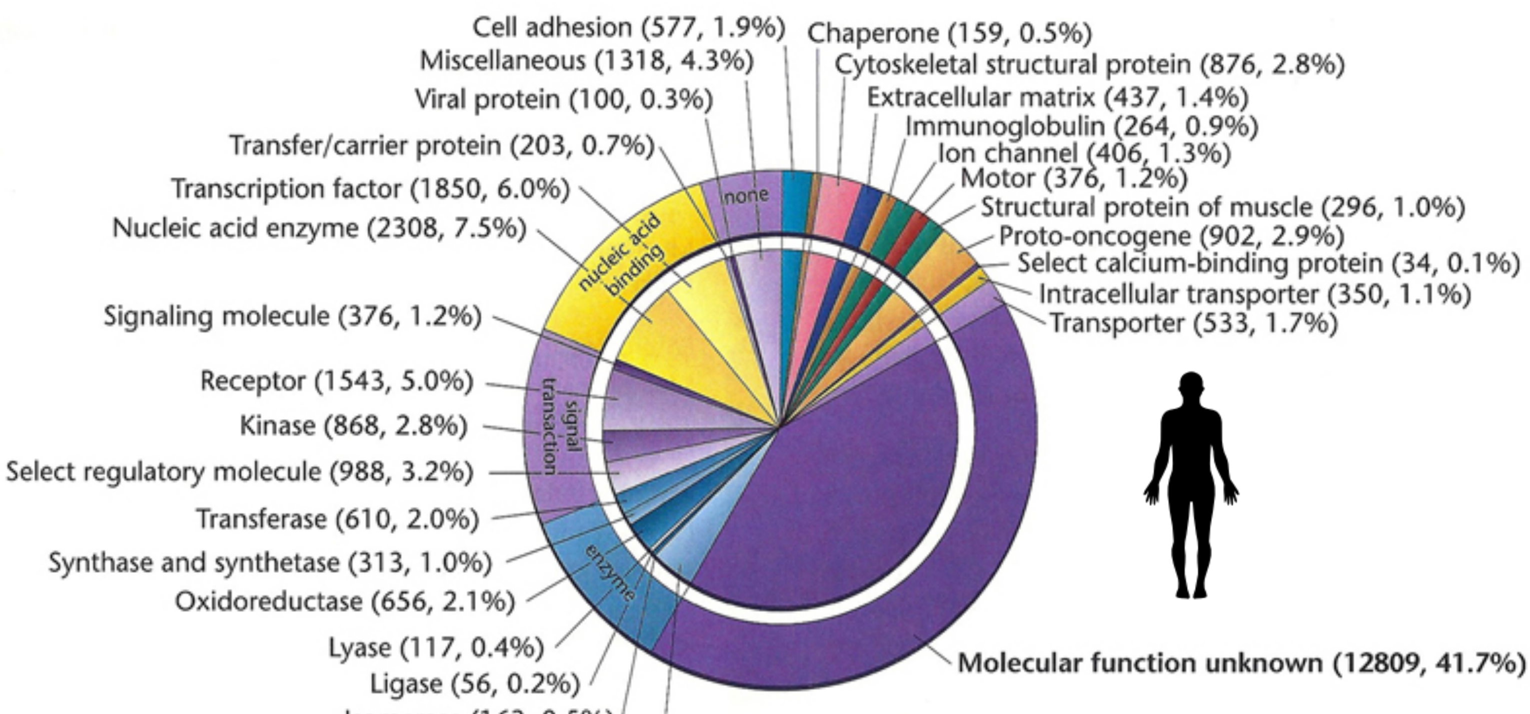
- Hierarchal clustering, K-means, Gaussian mixture model (GMM), Principal component analysis (PCA), …



Carlson et al., BMC Bioinformatics, 2006.

Limited knowledge in single species

~ half human genes, 1% human genome plus other 98% genomic elements (non-coding regions) with unknown functions

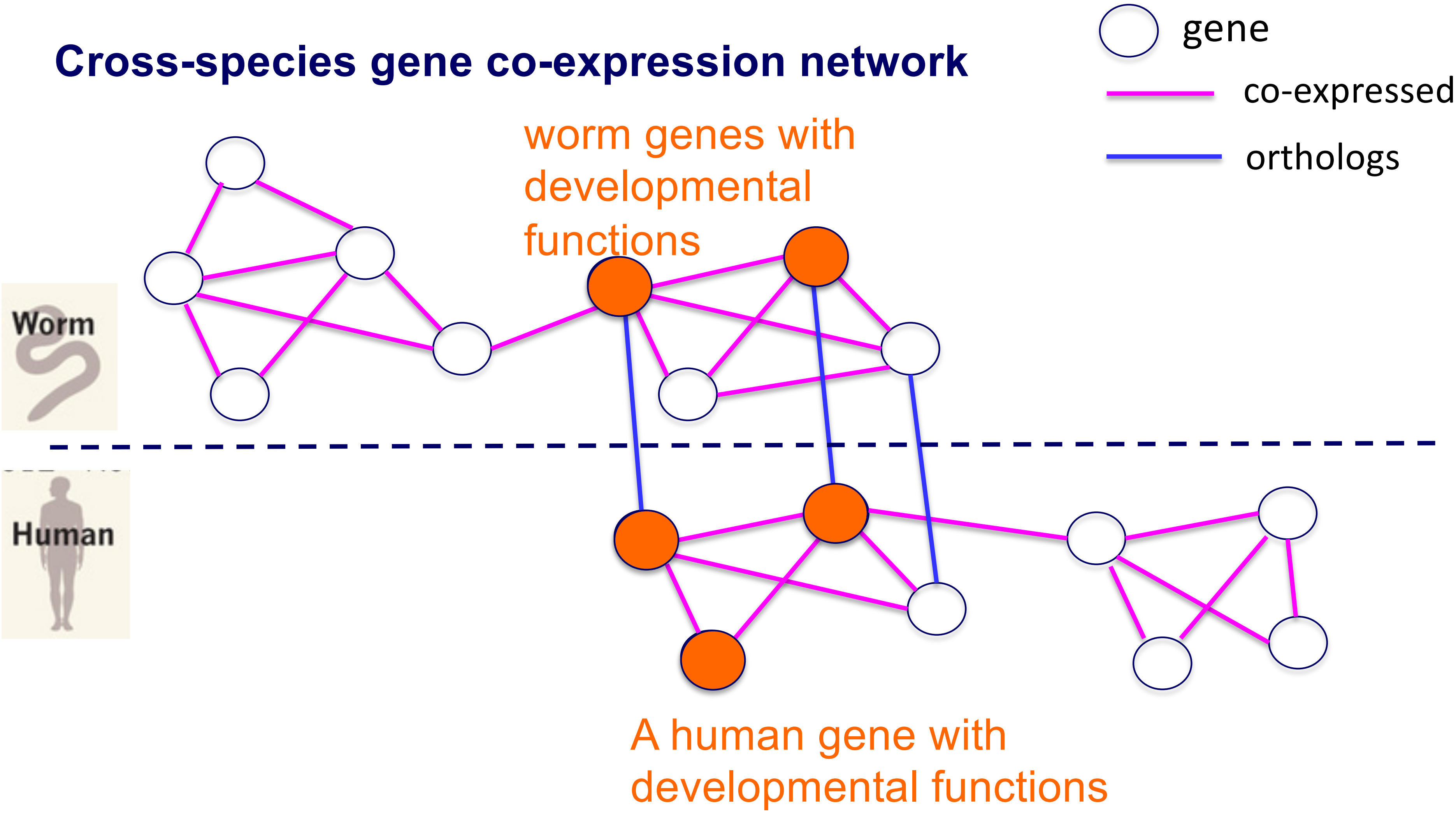


How do we know human gene functions during embryonic or brain development?

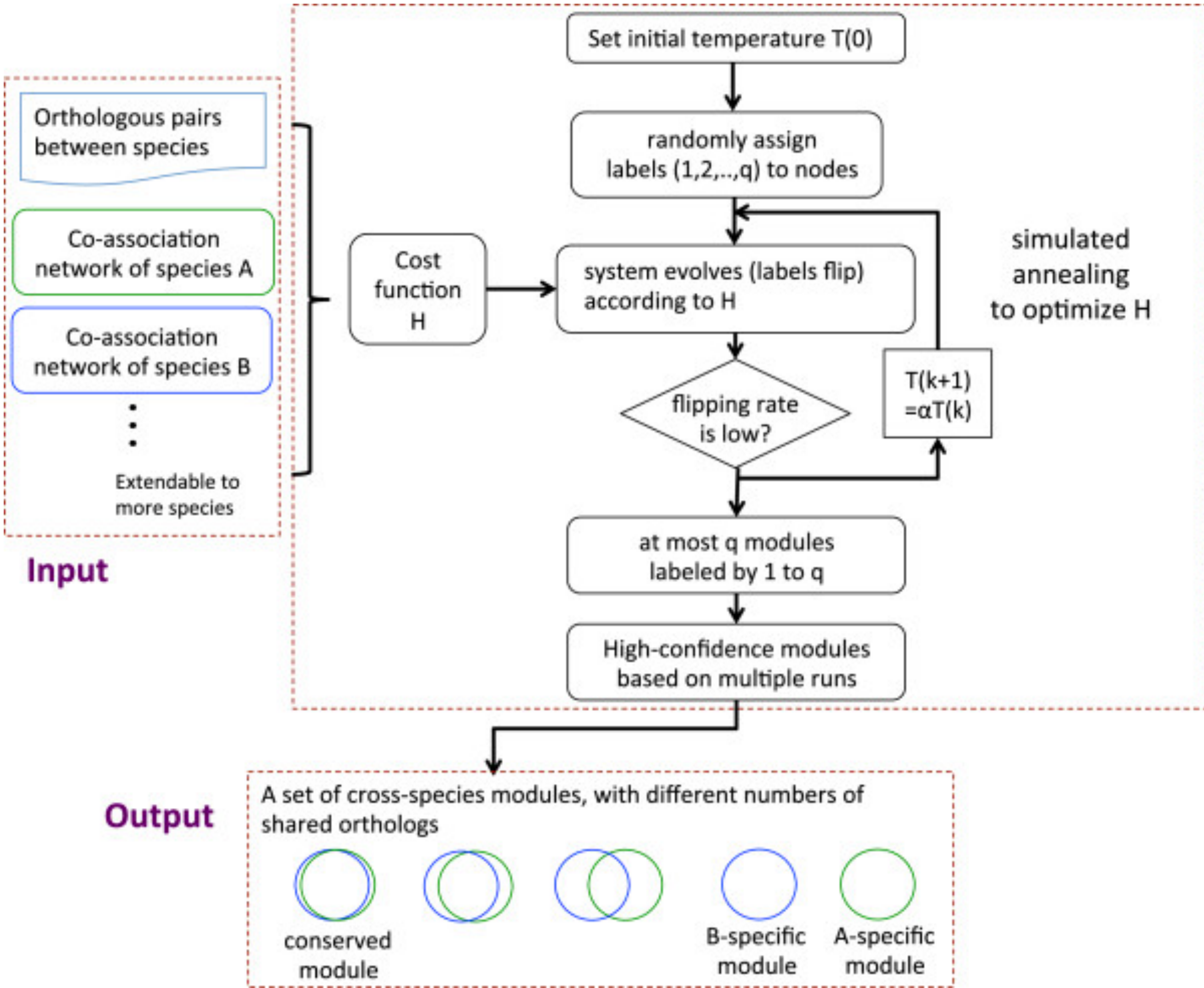
http://www.discoveryandinnovation.com/BIOL202/notes/lecture24.html

Integration of co-expressed and orthologous genes across species to transfer function information

Cross-species gene co-expression network

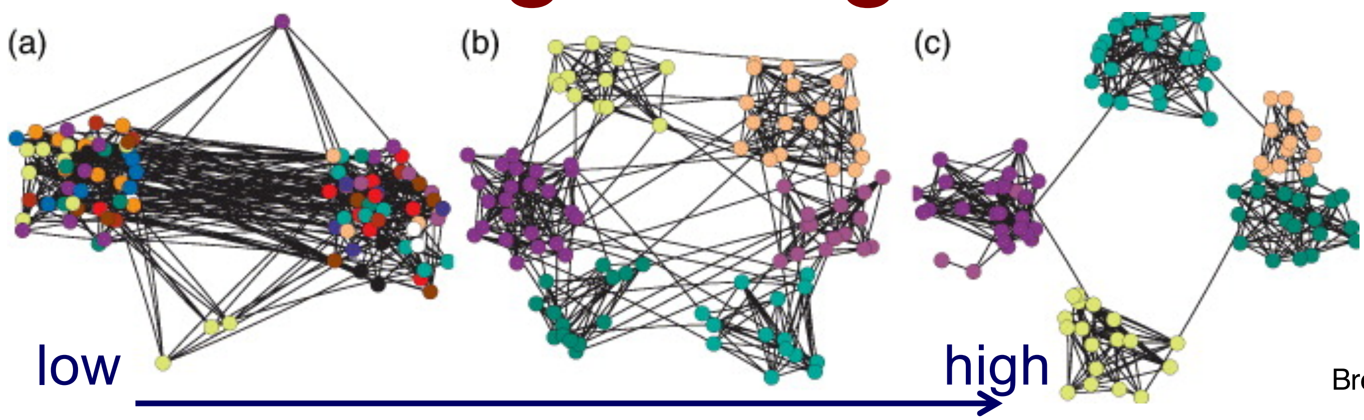


OrthoClust: an orthology-based method for clustering cross-species networks (e.g., co-expression networks)



Yan*, **Wang***, et al., *Genome Biology*, 2014

Maximize "modularity" for clustering a single network



Brede, Europhysics Letters, 2010.

Modularity Q: measurement on strength of network division

$$Q = \frac{1}{2m} \sum_{i,j} \left(W_{ij} - \frac{k_i k_j}{2m} \right) \delta_{\sigma_i \sigma_j}$$

sum over nodes within a group (module)

normalization
m: total number of edges

edge weight between nodes i and j

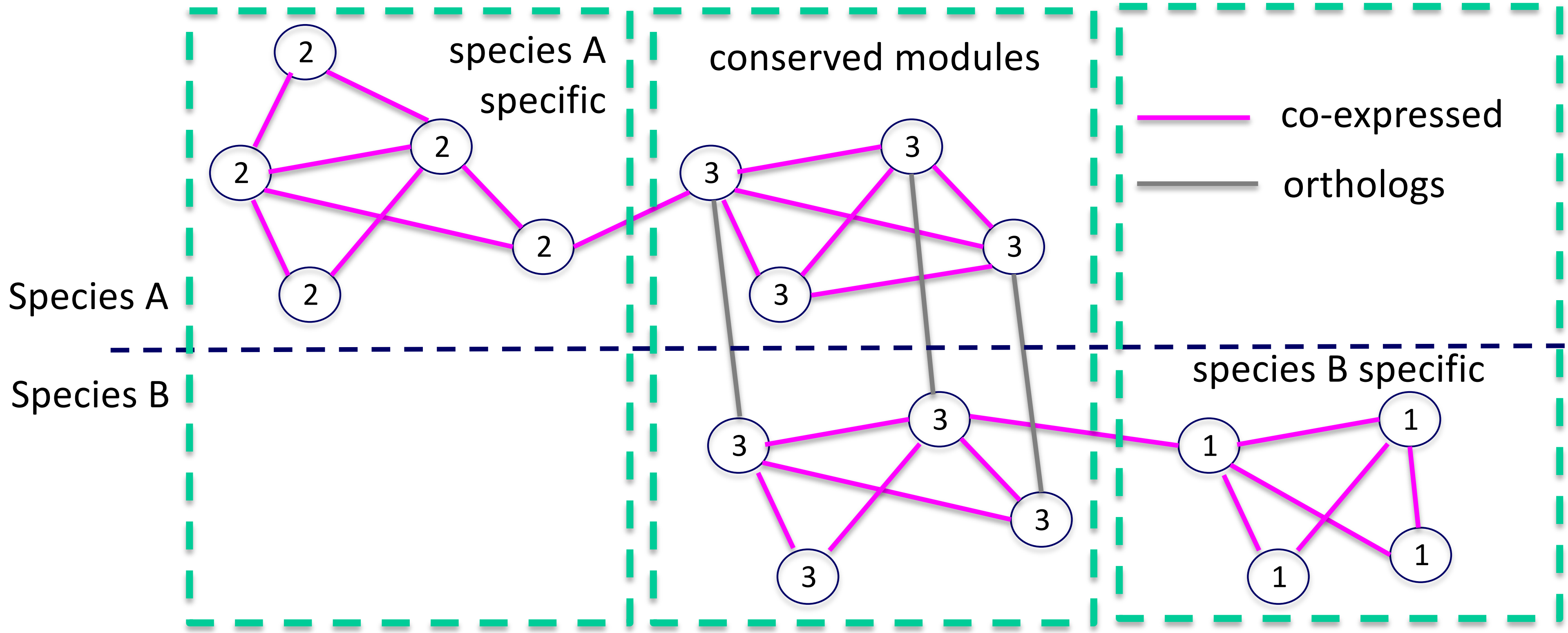
$\frac{k_i k_j}{2m} = p_{ij}$=expected edge weight that would go between i and j

Clustering goal: assign each node a module to maximize "modularity" as an objective function (module is a group of highly connected nodes)

Newman, PNAS, 2006.

OrthoClust: an orthology-based method for clustering cross-species networks

Every node i is assigned with a module number σ_i.



Objective function

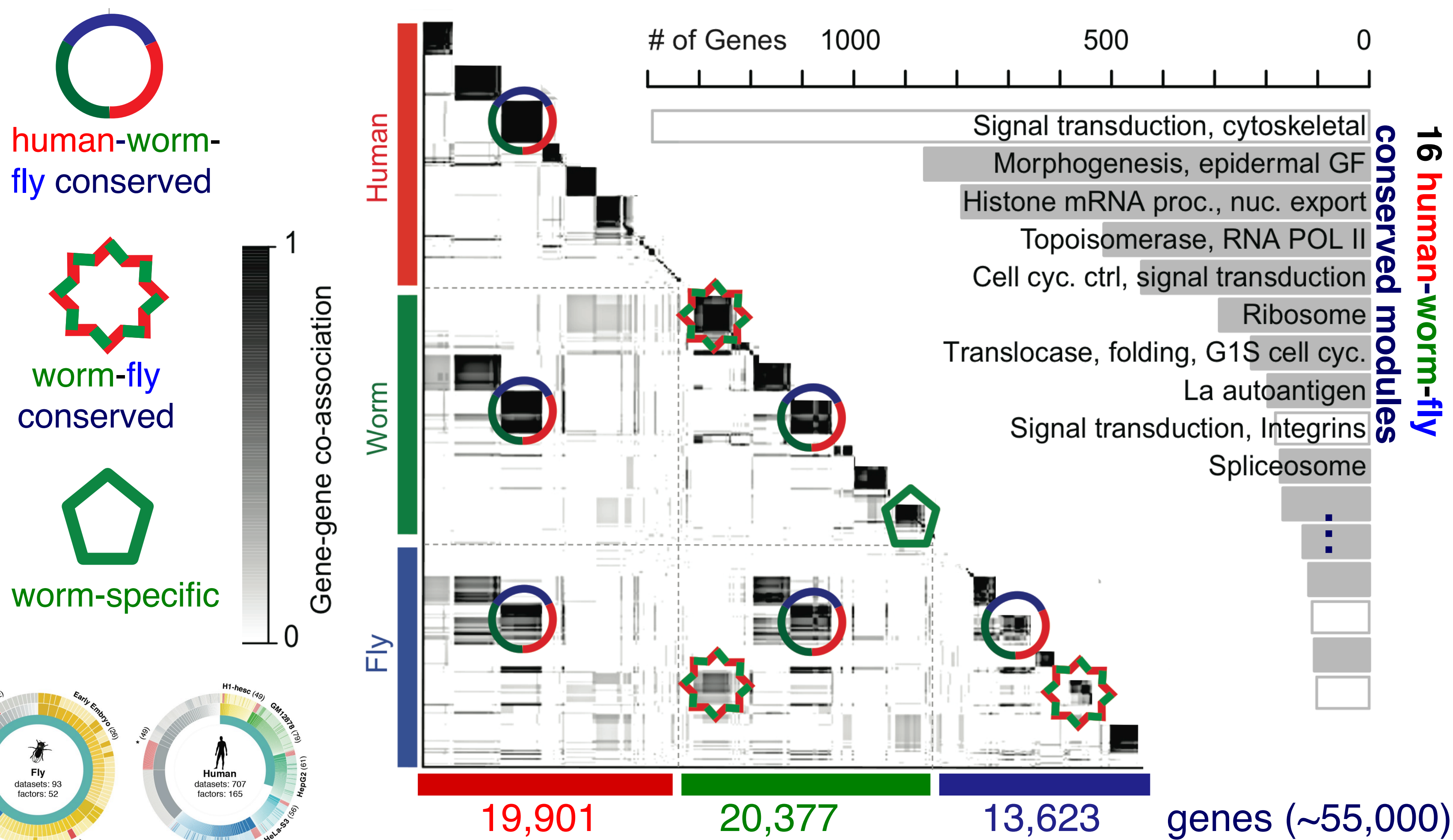
$$H = Q_A + Q_B + \kappa \sum_{(i,j')\in Ortho} \delta_{\sigma_i \sigma_{j'}}$$

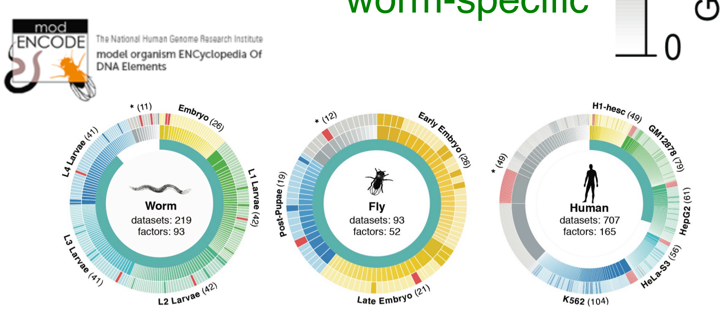
reward an orthologous pair in the same module

"**Modularity**" in species A + "**Modularity**" in species B + **consistency** between A & B

Yan*, **Wang***, et al., *Genome Biology*, 2014

Conserved gene co-expression modules discovered human genes having developmental functions





20,377 gene co-expression network across 33 developmental stages

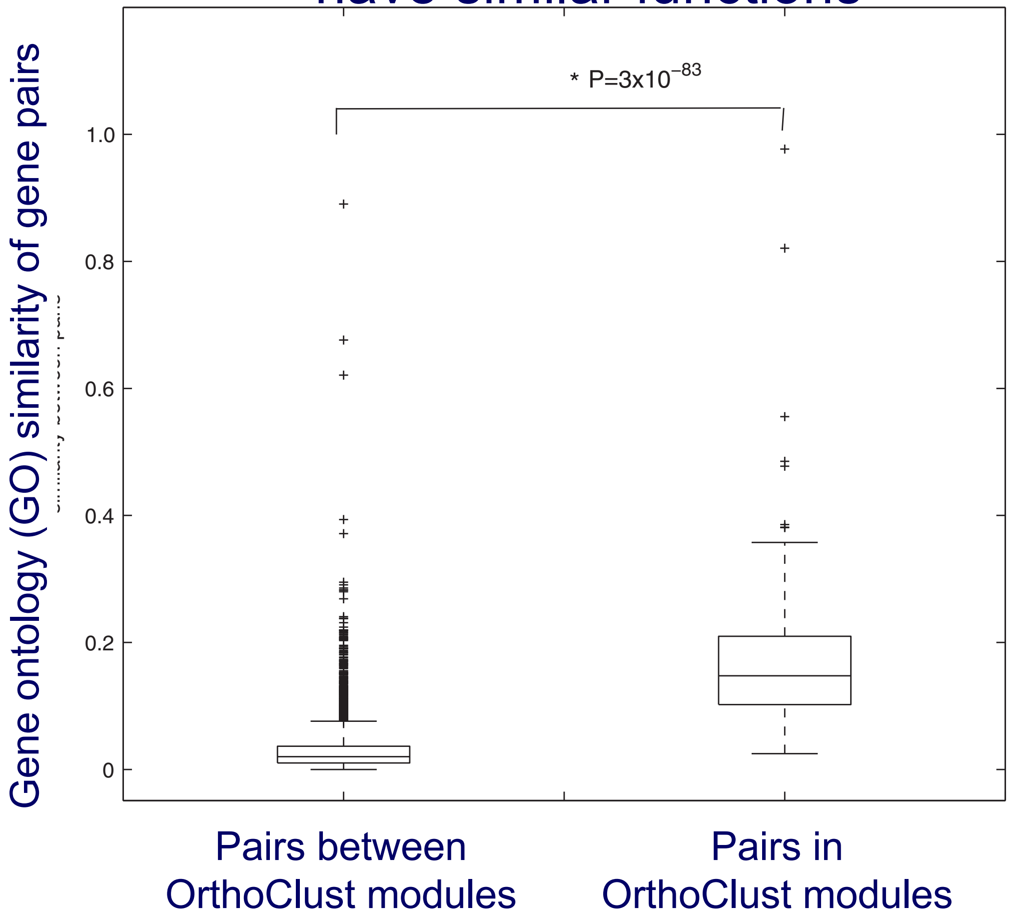
13,623 gene co-expression network across 30 developmental stages

19,901 gene co-expression network across 19 cell lines

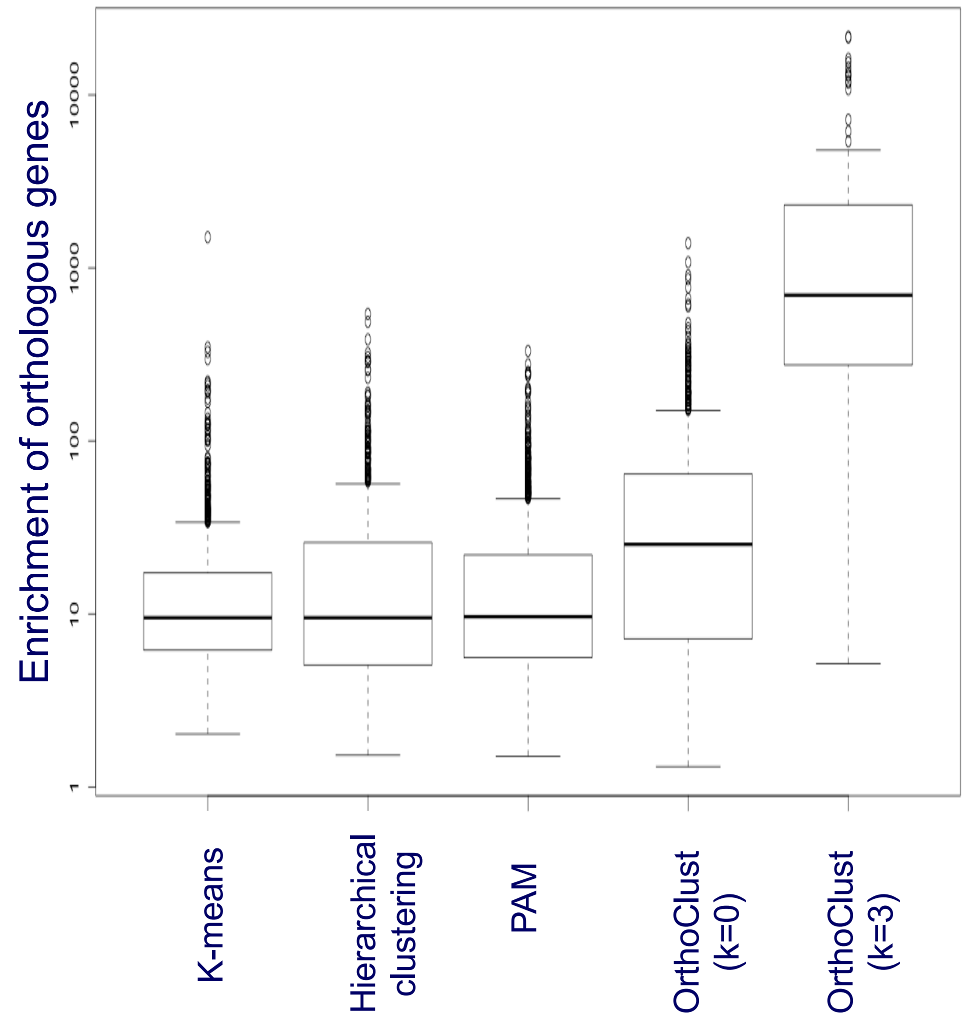
Gerstein*,…, **Wang***, et al., *Nature*, 2014

OrthoClust reveals better genomic functional groups

OrthoClust's modular genes have similar functions

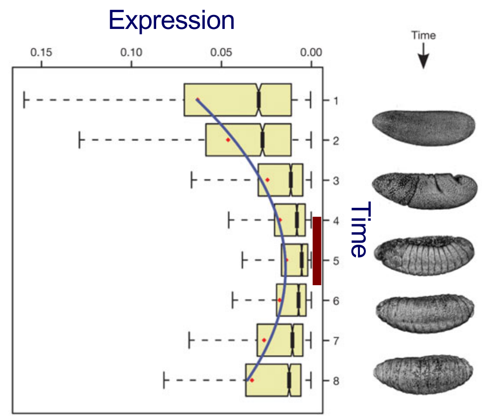


OrthoClust clusters more orthologs than other clustering methods

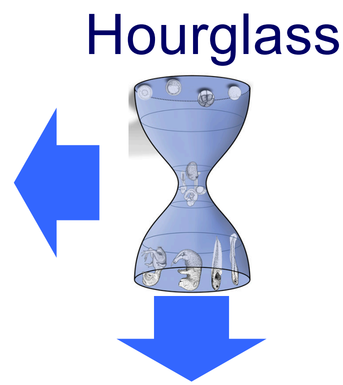


Developmental hourglass behavior across conserved modules in a species

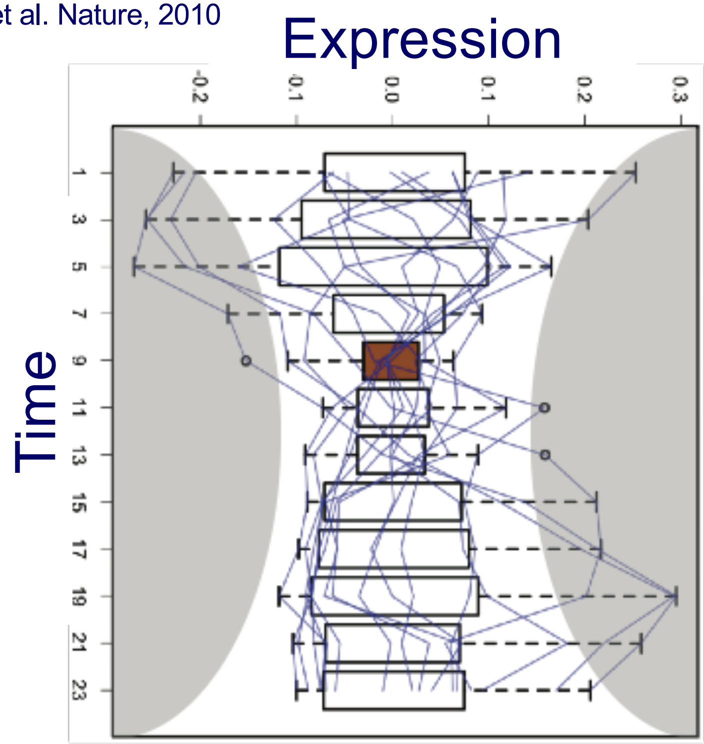
Inter-organism
Temporal differences among ortholog expression levels are minimized at phylotypic stage **across different species**.



Kalinka et al. Nature, 2010

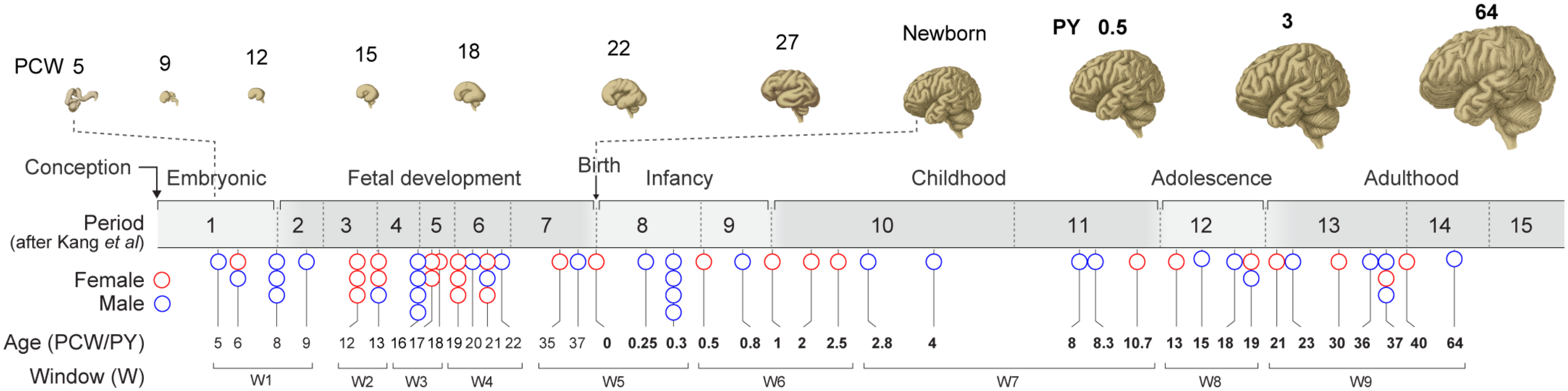


Intra-organism
Temporal differences among ortholog expression levels are minimized at phylotypic stage **across conserved modules in a species (fly)**.

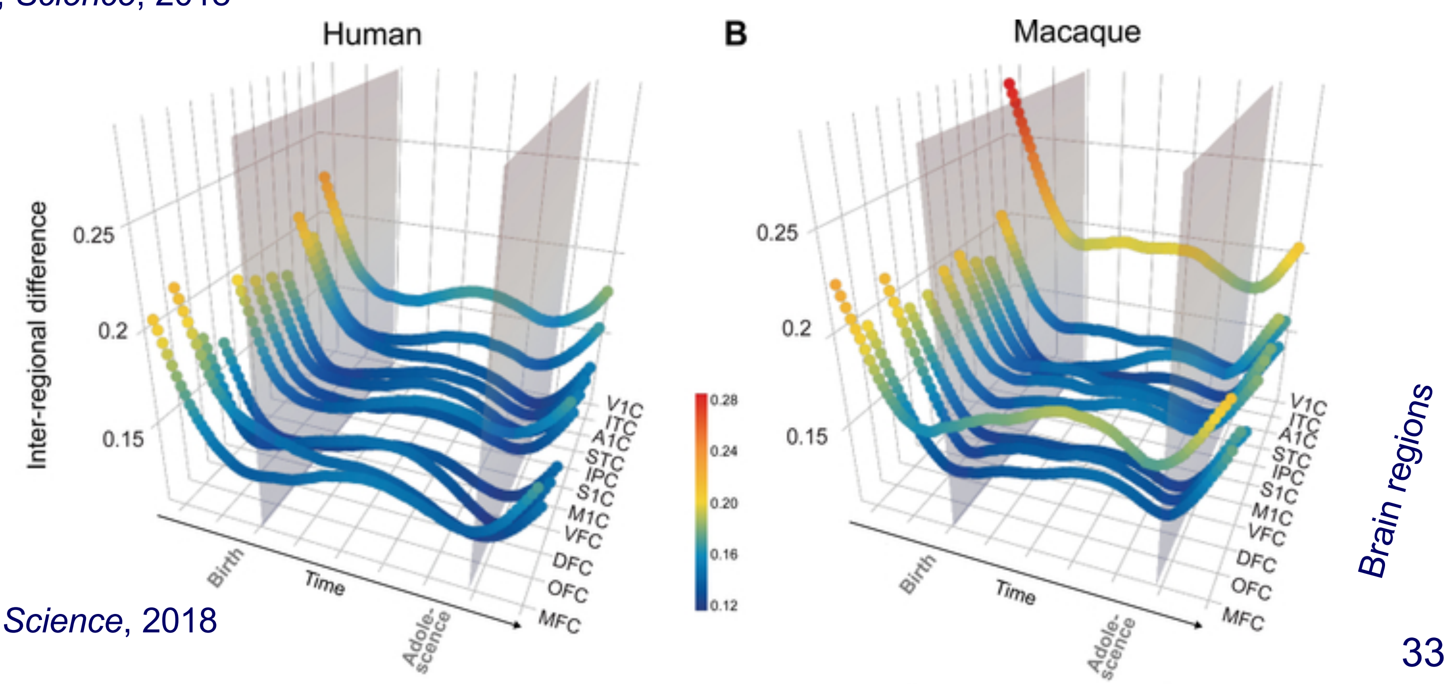


Gerstein*,…, **Wang***, et al., Nature, 2014

Human and Rhesus brain developmental “hourglass”



Li, …, **Wang**, ..., Sestan, *Science*, 2018



Ying, ..., Sestan, *Science*, 2018

Goals for lecture

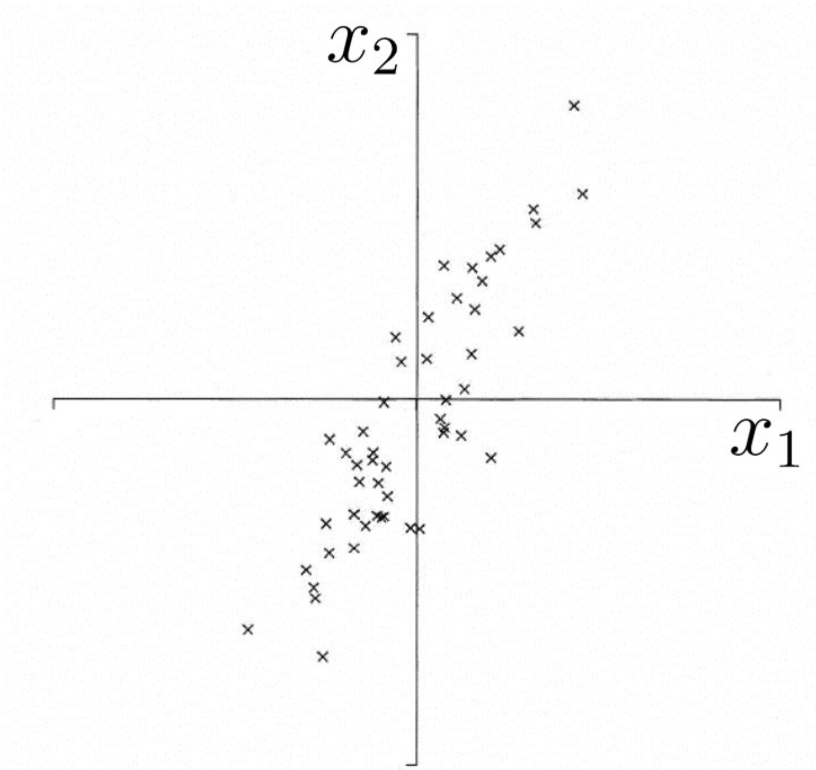
- Multi-omics data
- Machine learning modeling
  – Empirical risk minimization (ERM)
- Multi-layer network clustering
- Dimensionality reduction & Spectral methods
- Decision tree
- Neural network

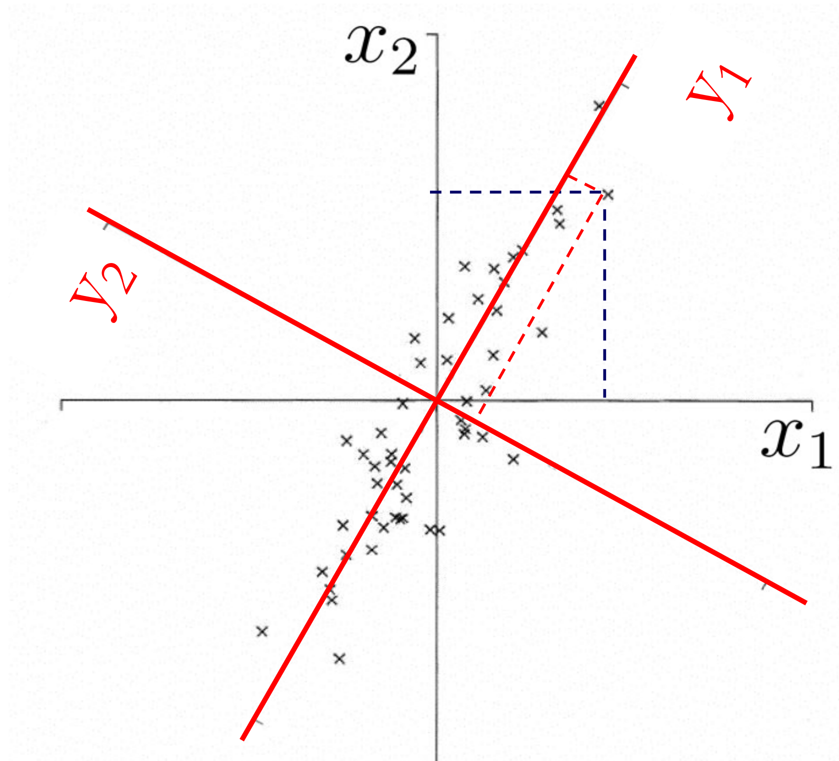
Reading list for spectral methods

- O Alter et al. (2000). "Singular value decomposition for genome-wide expression data processing and modeling." PNAS 97: 10101
- Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between co-expression modules. BMC Systems Biology 2007, 1:54
- Z Zhang et al. (2007) "Statistical analysis of the genomic distribution and correlation of regulatory elements in the ENCODE regions." Genome Res 17: 787
- TA Gianoulis et al. (2009) "Quantifying environmental adaptation of metabolic pathways in metagenomics." PNAS 106: 1374.

What is Principal component analysis (PCA) ?

- A technique used to reduce the dimensionality of a data set by finding directions of maximum variability
- Projection (typically a rotation) into new axes
- But still retains the dataset's variation



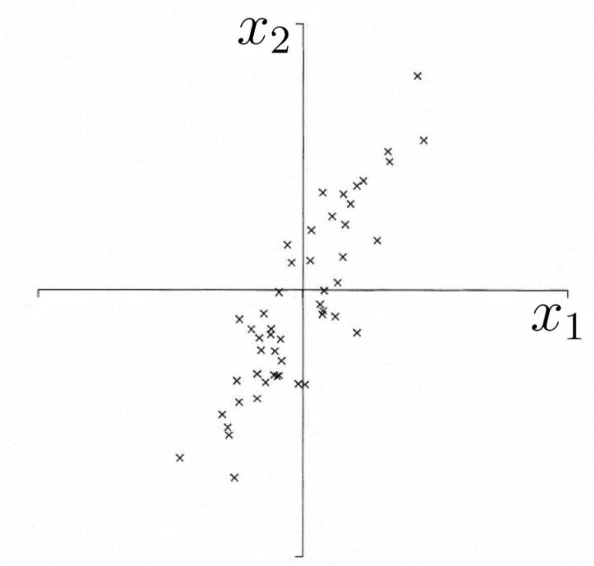


Adapted from http://www.astro.princeton.edu/~gk/A542/PCA.ppt

PCA Matrix

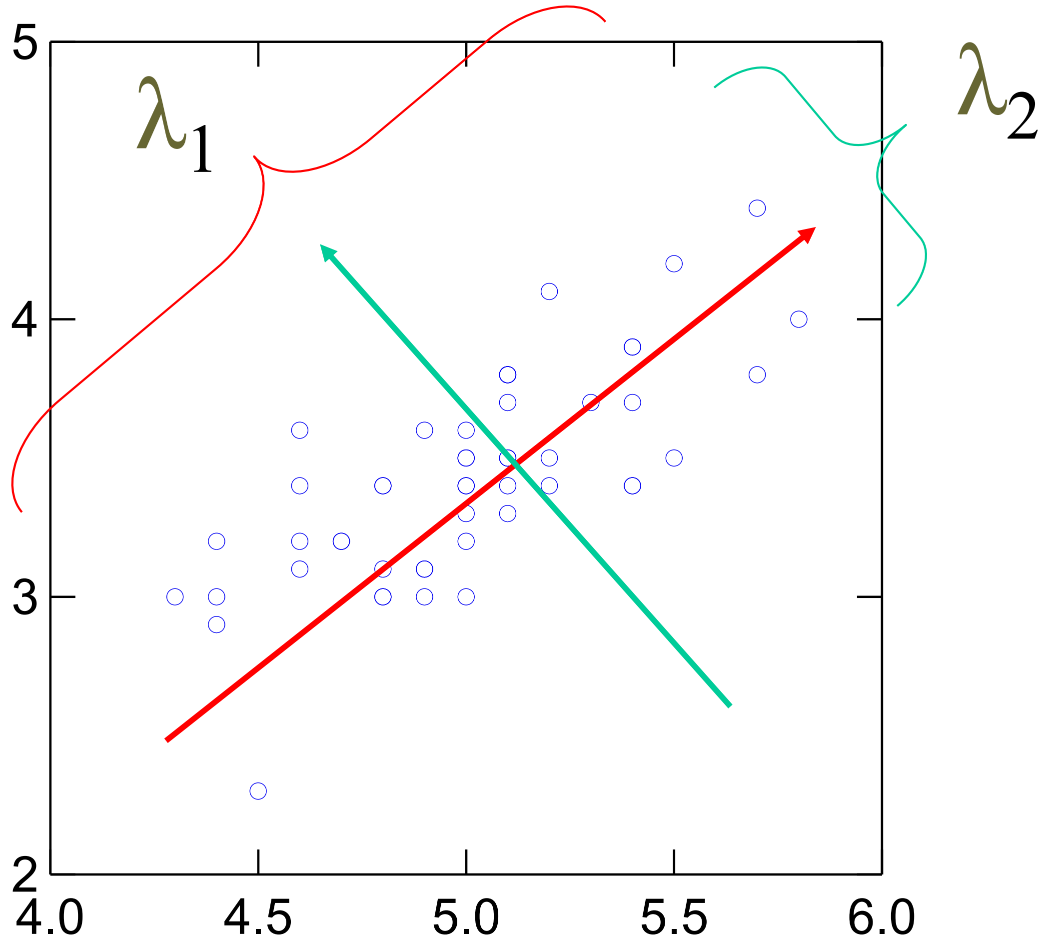
1. Start with dataset of k variables X = x_1, x_2 ... x_k and n observations.
2. Construct **covariance or correlation matrix** for variables.
3. **The Eigenvalue Problem** or Eigenanalysis: matrix diagonalization and solve for eigenvalues and eigenvectors

E.g. Start with a bunch of coordinates



Observations	X1	X2
1	2	5
2	5	6
3	4	2
4	3	7
5	9	-5
…		
n	-5	-8

Interpretation: Eigenvalues & Eigenvectors



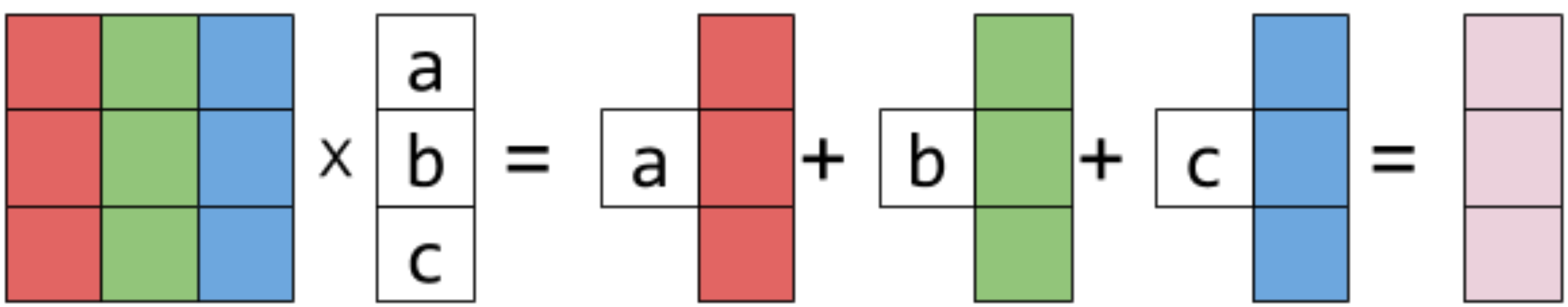
Adapted from http://myweb.dal.ca/~hwhitehe/BIOL4062/pca.ppt

Quick Refresher on Matrices

$$\begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{pmatrix} * \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} ax_1 + by_1 + cz_1 \\ ax_2 + by_2 + cz_2 \\ ax_3 + by_3 + cz_3 \end{pmatrix}$$

$$\underset{\text{Matrix } \mathbf{A} \text{ is } 3x4}{\begin{bmatrix} 8 & 3 & 0 & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}} \underset{\text{Matrix } \mathbf{B} \text{ is } 4x4}{\begin{bmatrix} 5 & \cdot & \cdot & \cdot \\ 4 & \cdot & \cdot & \cdot \\ 3 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot \end{bmatrix}} = \underset{\text{Matrix } \mathbf{C} \text{ is } 3x4}{\begin{bmatrix} 53 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}}$$

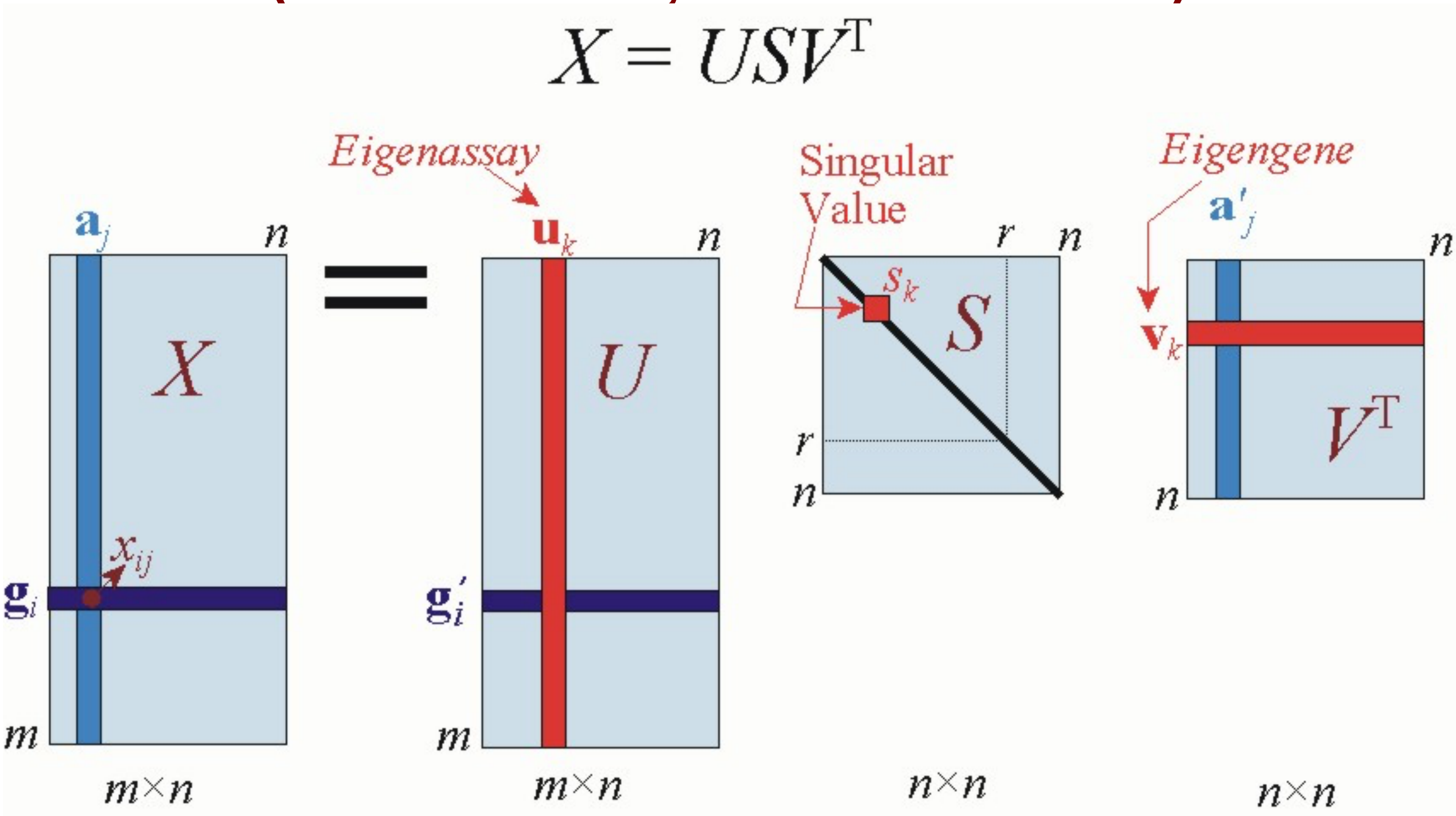
$$\text{because } c_{11} = \sum_{k=1}^{4} a_{1k} b_{k1} = 8 \cdot 5 + 3 \cdot 4 + 0 \cdot 3 + 1 \cdot 1 = 53$$



http://eli.thegreenplace.net/2015/visualizing-matrix-multiplication-as-a-linear-combination/

http://www.catonmat.net/blog/mit-linear-algebra-part-three/

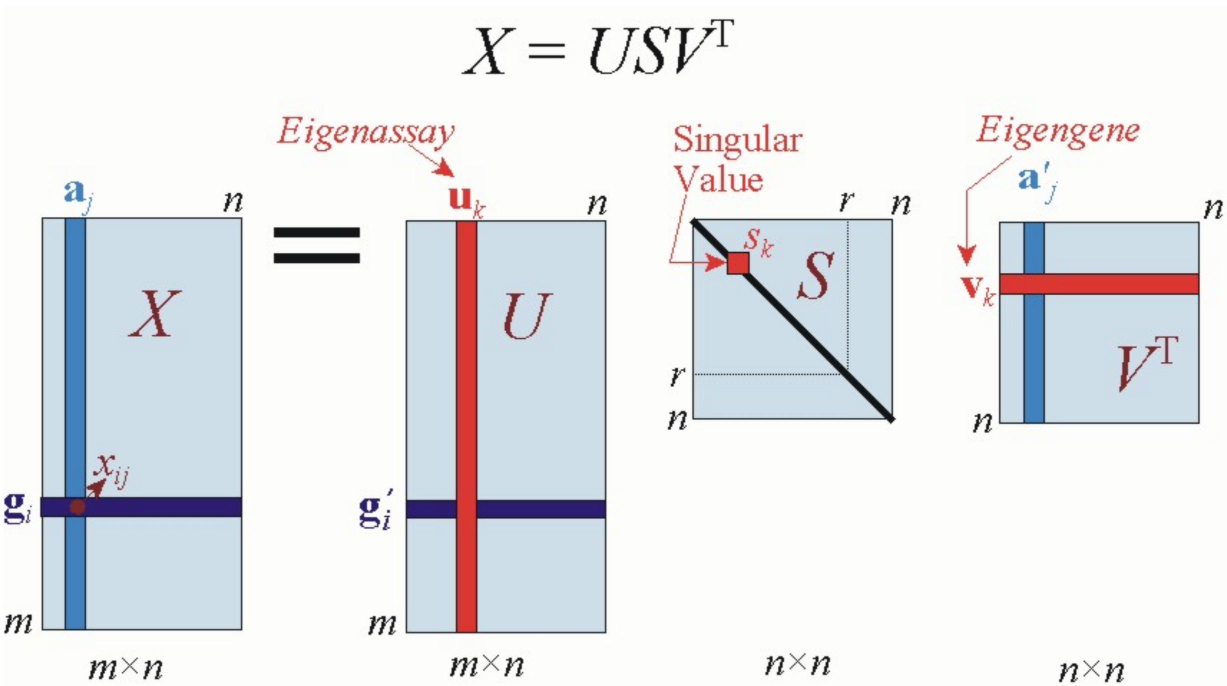
SVD for gene expression data (Alter et al, PNAS 2000)



http://www.gersteinlab.org/courses/452/

Notation

- m=1000 genes
  - row-vectors
  - 10 eigengene (v_i) of dimension 10 conditions
- n=10 conditions (assays)
  - column vectors
  - 10 eigenconditions (u_i) of dimension 1000 genes



http://www.gersteinlab.org/courses/452/

SVD as sum of rank-1 matrices

- $A = USV^T$
- $A = s_1 \mathbf{u}_1 \mathbf{v}_1^T + s_2 \mathbf{u}_2 \mathbf{v}_2^T + \ldots + s_n \mathbf{u}_n \mathbf{v}_n^T$
- $s_1 \geq s_2 \geq \ldots \geq s_n \geq 0$
- What is the rank-r matrix $\hat{A}$ that best approximates A ?
  - Minimize $\sum_{i=1}^{m} \sum_{j=1}^{n} \left(\hat{A}_{ij} - A_{ij} \right)^2$
- $\hat{A} = s_1 \mathbf{u}_1 \mathbf{v}_1^T + s_2 \mathbf{u}_2 \mathbf{v}_2^T + \ldots + s_r \mathbf{u}_r \mathbf{v}_r^T$
- Very useful for matrix approximation

an outer product (uv^T) giving a matrix rather than the scalar of the inner product

LSQ approx. If r=1, this amounts to a line fit.

Potential problems of SVD/PCA

If the dataset…

- Lacks Independence
  - **NO PROBLEM**
- Lacks Normality
  - Normality desirable but not essential
- Lacks Precision
  - Precision desirable but not essential
- Lacks Linearity
  - **Problem**: Use other non-linear (kernel) methods
- Many Zeroes in Data Matrix (Sparse)
  - **Problem**: Use Correspondence Analysis

Conclusion

- SVD is the "absolute high point of linear algebra"
- SVD is difficult to compute; but once we have it, we have many things
- SVD finds the best approximating subspace, using linear transformation
- Simple SVD cannot handle translation, non-linear transformation, separation of labeled data, etc.
- Good for exploratory analysis; but once we know what we look for, use appropriate tools and model the structure of data explicitly!
- http://genomicsclass.github.io/book/pages/pca_svd.html

Goals for lecture

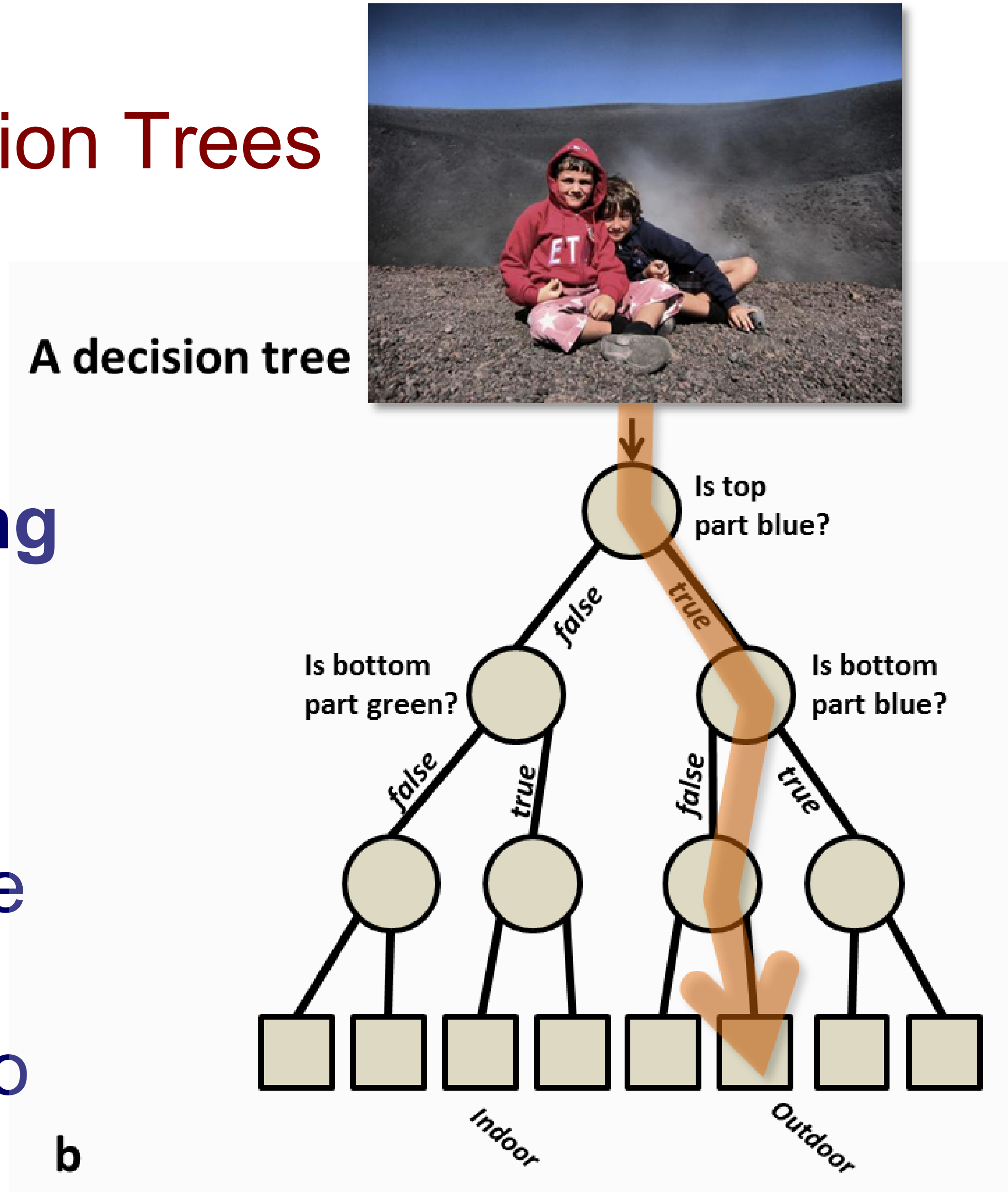
- Multi-omics data
- Machine learning modeling
  - Empirical risk minimization (ERM)
- Multi-layer network clustering
- Dimensionality reduction & Spectral methods
- Decision tree
- Neural network

Reading list

- What are decision trees?
  - Nat Biotechnol. 2008 Sep; 26(9): 1011–1013.
- Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle?
  - https://academic.oup.com/bib/article/14/3/315/255469

Decision Trees

- **Classify data by asking questions** that divide data in subgroups
- Keep asking questions until subgroups become homogenous
- Use **tree** of questions to make predictions



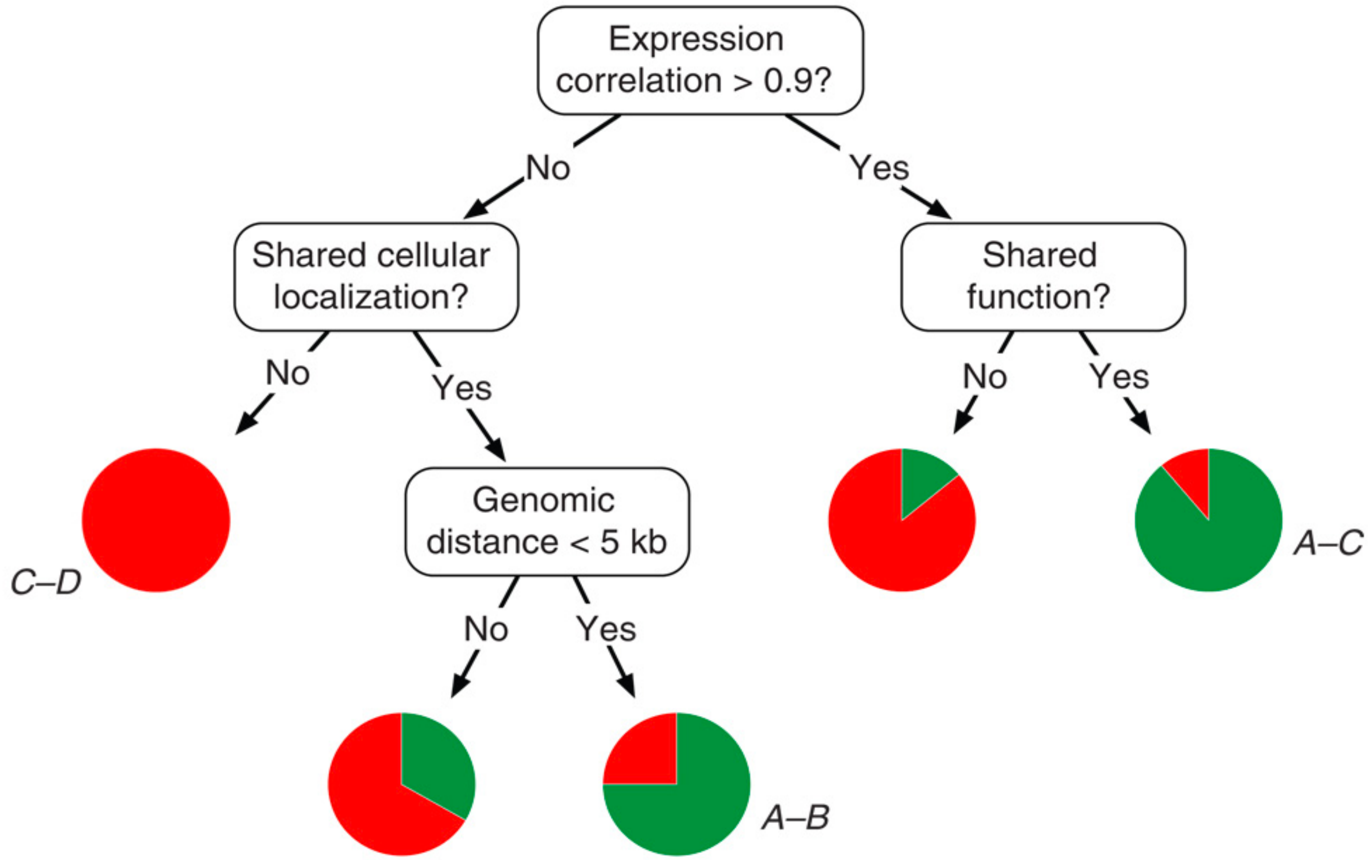
Example: Is a picture taken inside or outside?

Criminisi, Shotton, and Konukoglu *Microsoft Technical Report* 2011

a

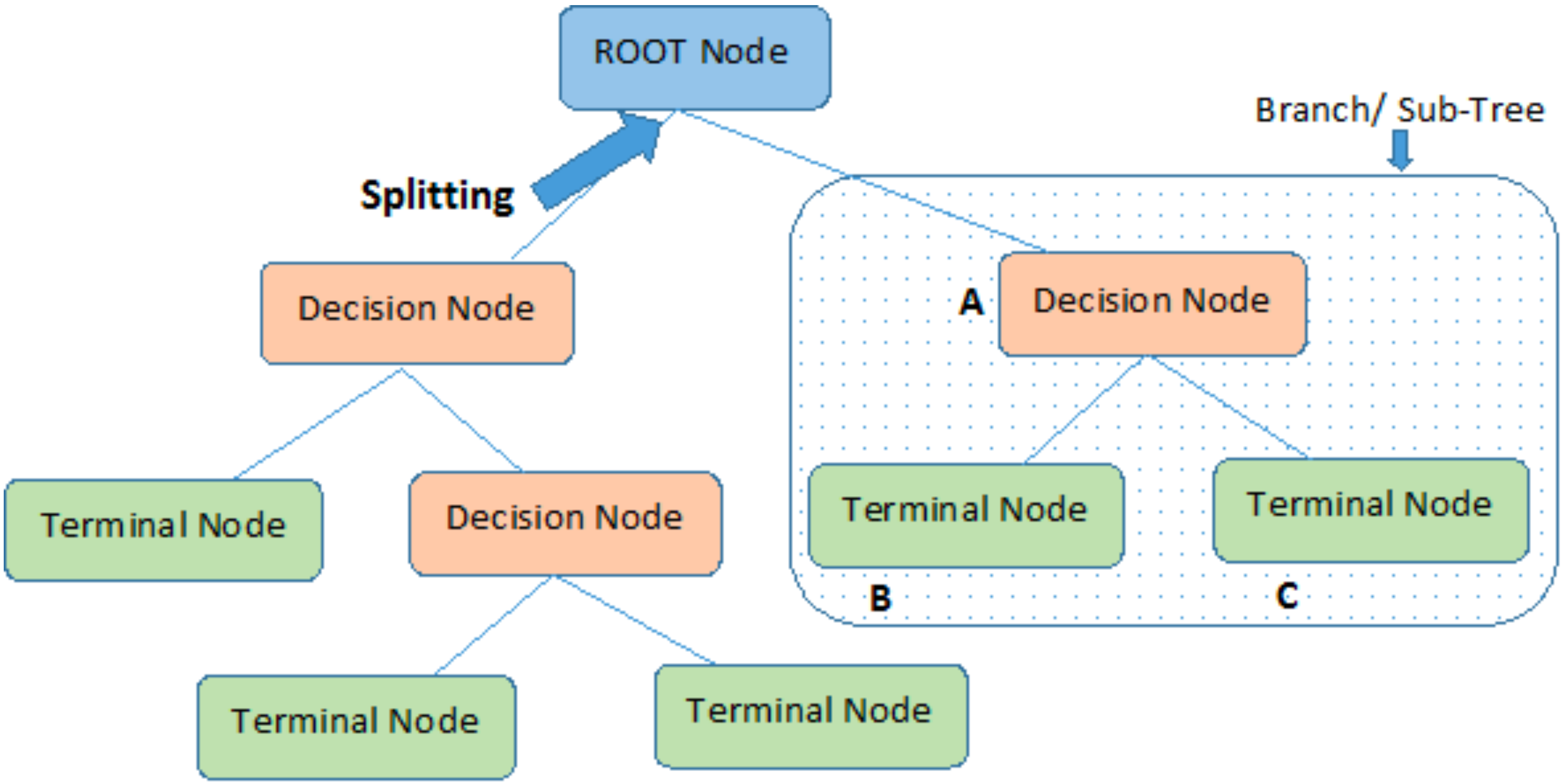
Gene Pair	Interact?	Expression correlation	Shared localization?	Shared function?	Genomic distance
A-B	Yes	0.77	Yes	No	1 kb
A-C	Yes	0.91	Yes	Yes	10 kb
C-D	No	0.1	No	No	1 Mb
⋮					

b



A hypothetical example of how a decision tree might predict protein-protein interactions
Nat Biotechnol. 2008 Sep; 26(9): 1011–1013.

Terminology related to Decision Trees

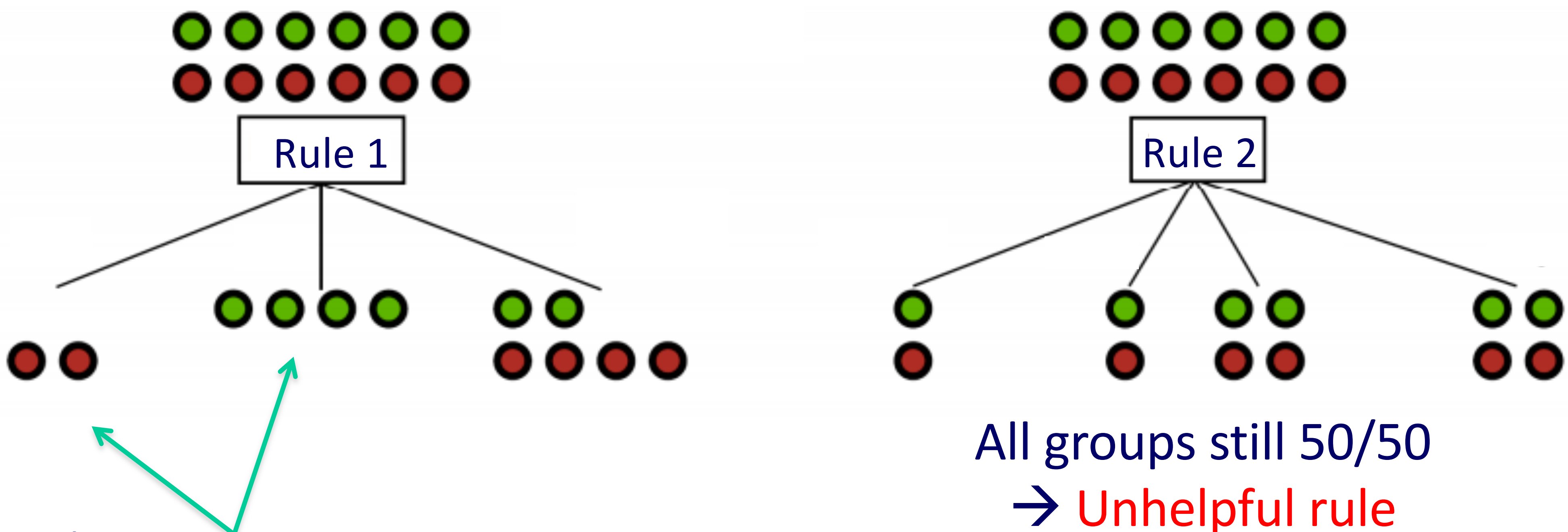


Note:- A is parent node of B and C.

https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/

What makes a good rule?

- Want resulting groups to be as homogenous as possible



2/3 Groups homogenous
→Good rule

All groups still 50/50
→ Unhelpful rule

Quantifying the value of rules

- Decrease in inhomogeneity
  - Most popular metric: Information theoretic entropy
  - Use frequency of classifier characteristic within group as probability
  - Minimize entropy to achieve homogenous group

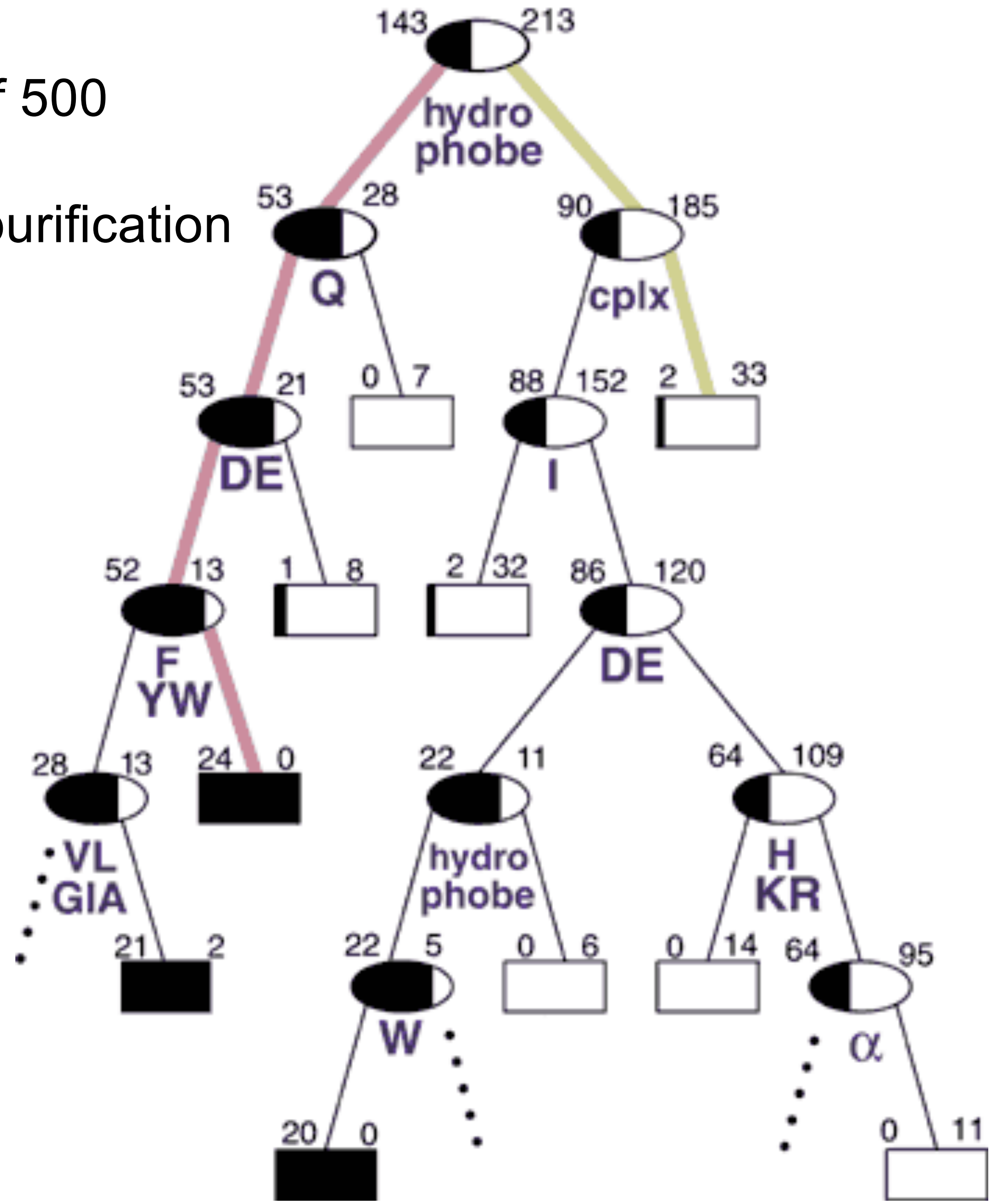
$$S = -\sum_{i=1}^{m} p_i \log p_i$$

Algorithm

- For each characteristic:
  - Split into subgroups based on each possible value of characteristic
- Choose rule from characteristic that maximizes decrease in inhomogeneity
- For each subgroup:
  - if (inhomogeneity < threshold):
    - Stop
  - else:
    - Restart rule search (recursion)

Retrospective Decision Trees

Analysis of the Suitability of 500 M.thermo. proteins to find optimal sequences purification



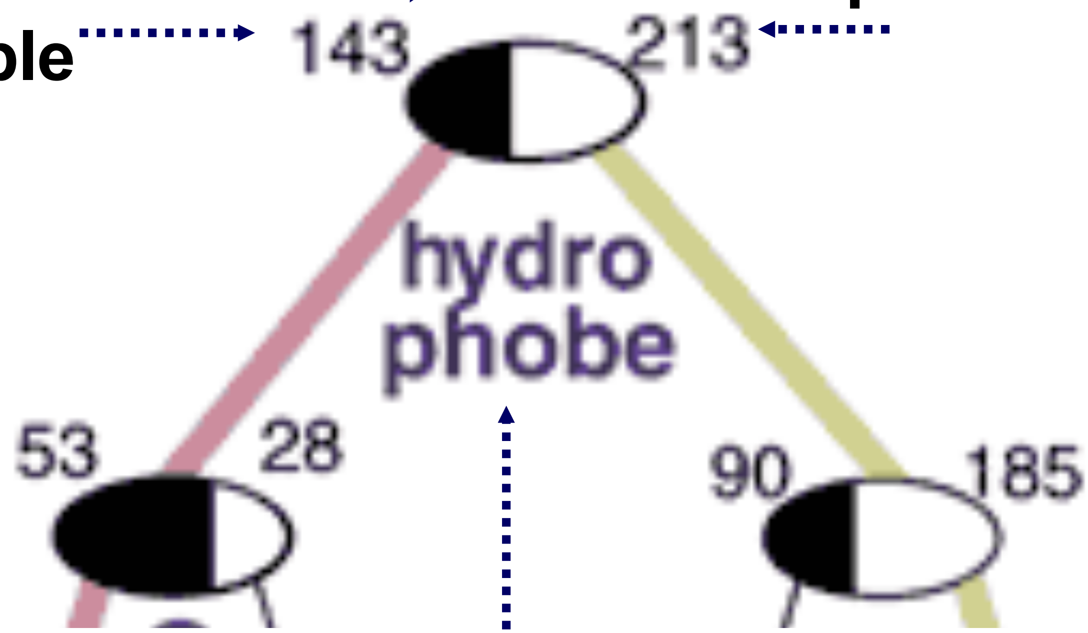
[Bertone et al. NAR ('01)]

Retrospective Decision Trees
Nomenclature

356 total

Not Expressible

Expressible

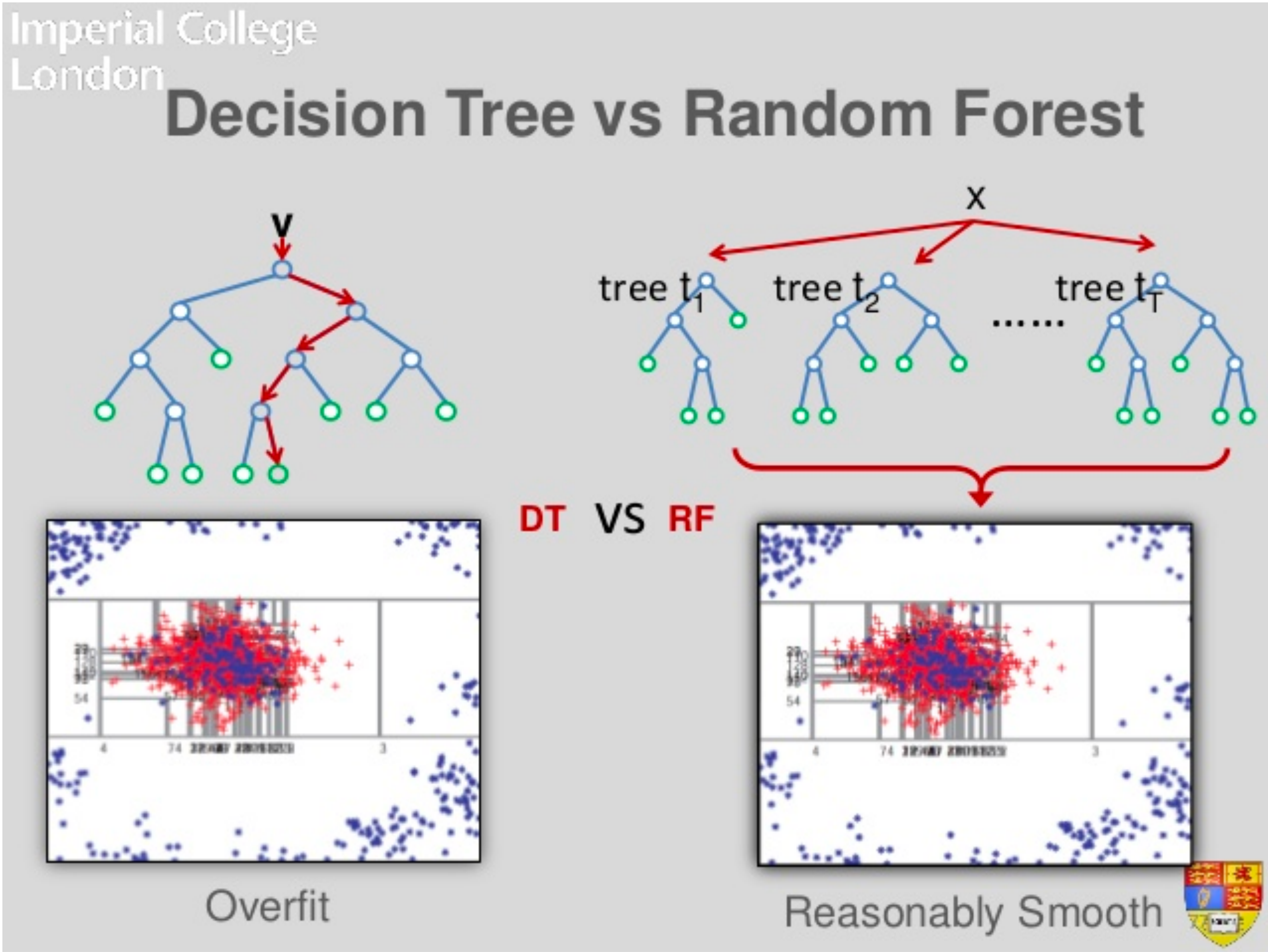


Has a hydrophobic stretch? (Y/N)

[Bertone et al. NAR ('01)]

Extensions of Decision Trees

- Decision Trees method is very sensitive to noise in data
- Random forests is an ensemble of decision trees, and is much more effective.



Exercise

- A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python)
  - https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/
- Random Forests in R
  - https://www.r-bloggers.com/random-forests-in-r/
  - http://dni-institute.in/blogs/random-forest-using-r-step-by-step-tutorial/

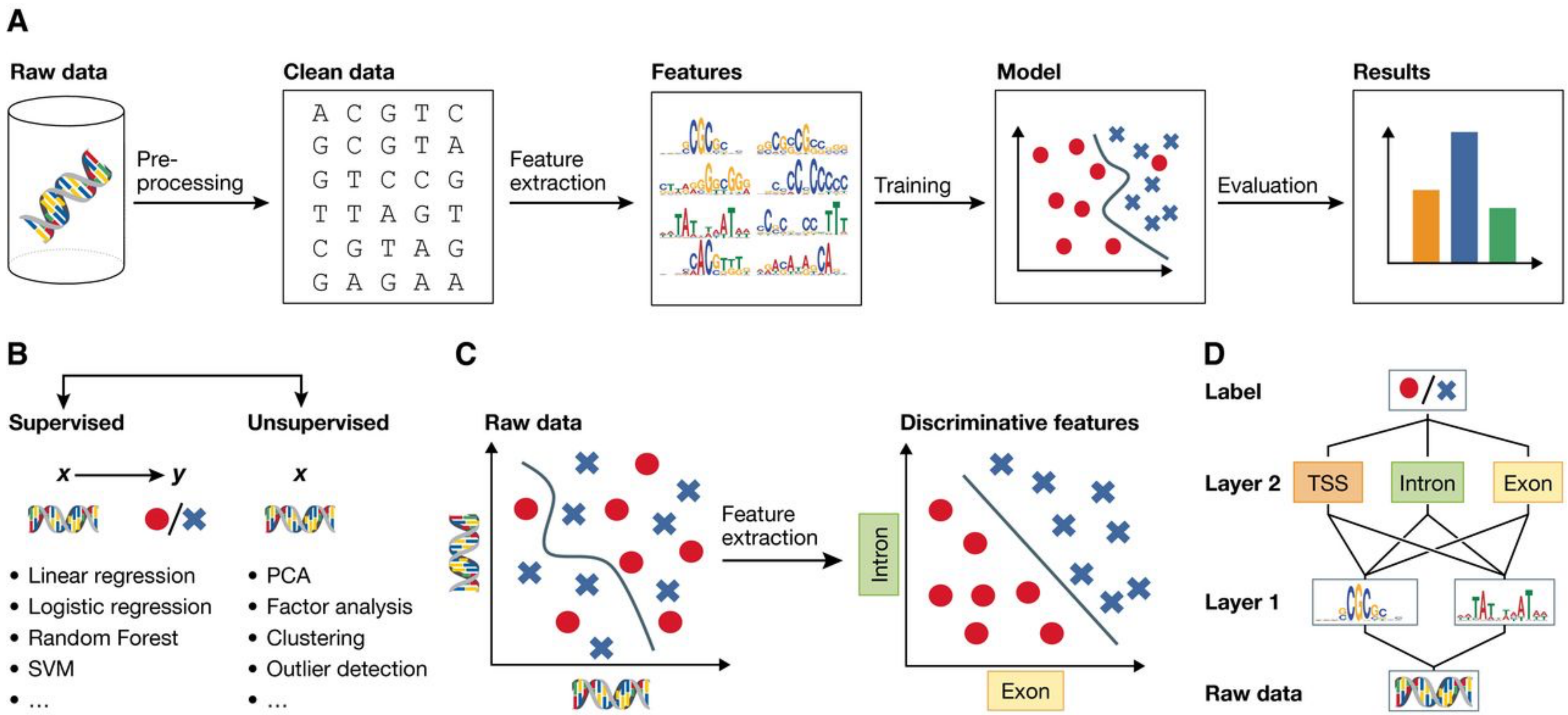
Goals for lecture

- Multi-omics data
- Machine learning modeling
  - Empirical risk minimization (ERM)
- Multi-layer network clustering
- Dimensionality reduction & Spectral methods
- Decision tree
- Neural network

Reading list

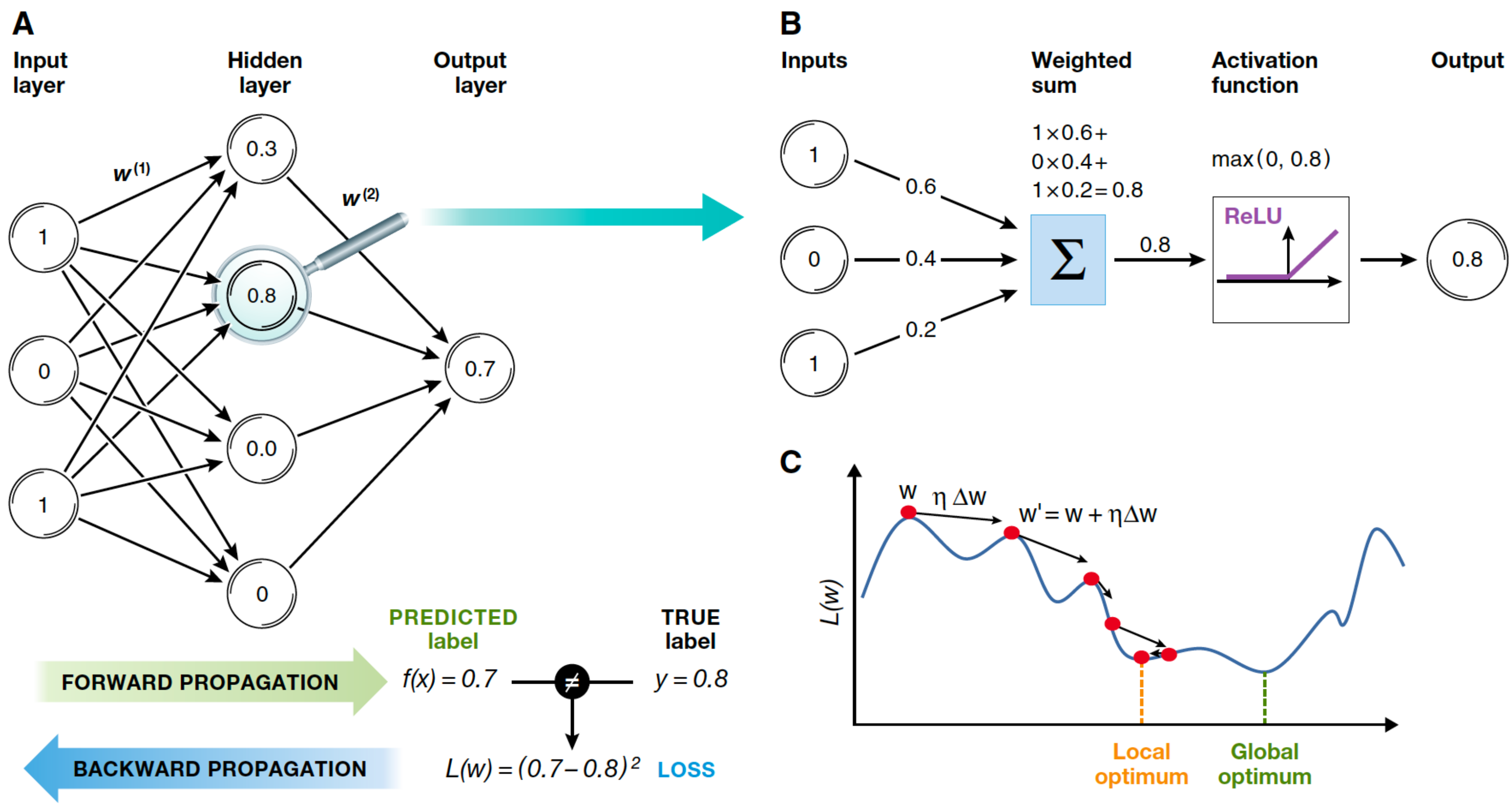
- **Deep learning for computational biology**
  - http://msb.embopress.org/content/12/7/878
- **Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning**
  - https://www.nature.com/articles/nbt.3300
- https://github.com/hussius/deeplearning-biology
- The Incredible Convergence Of Deep Learning And Genomics

Machine learning and representation learning



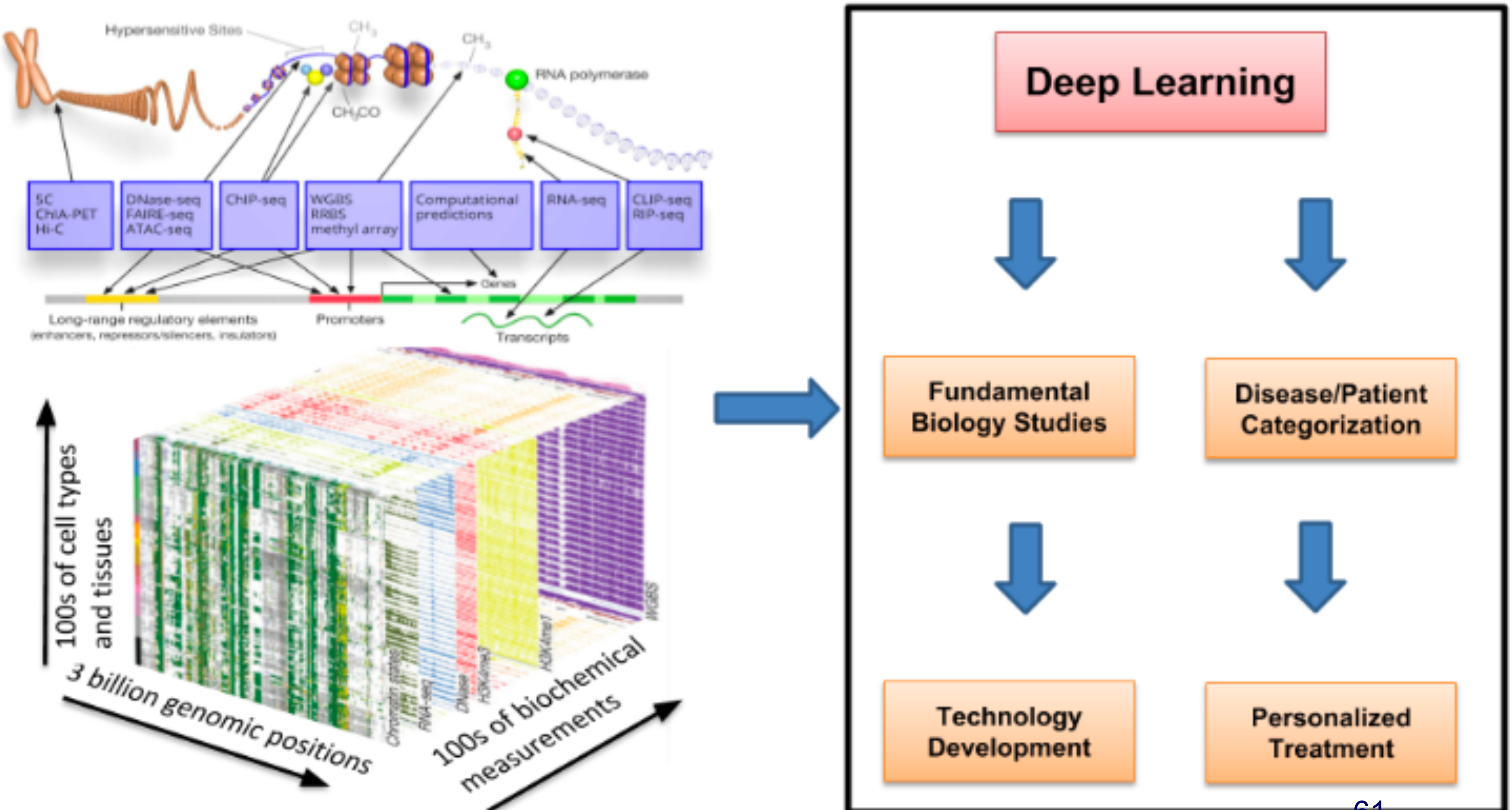
Christof Angermueller et al. Mol Syst Biol 2016;12:878

Artificial Neural Network



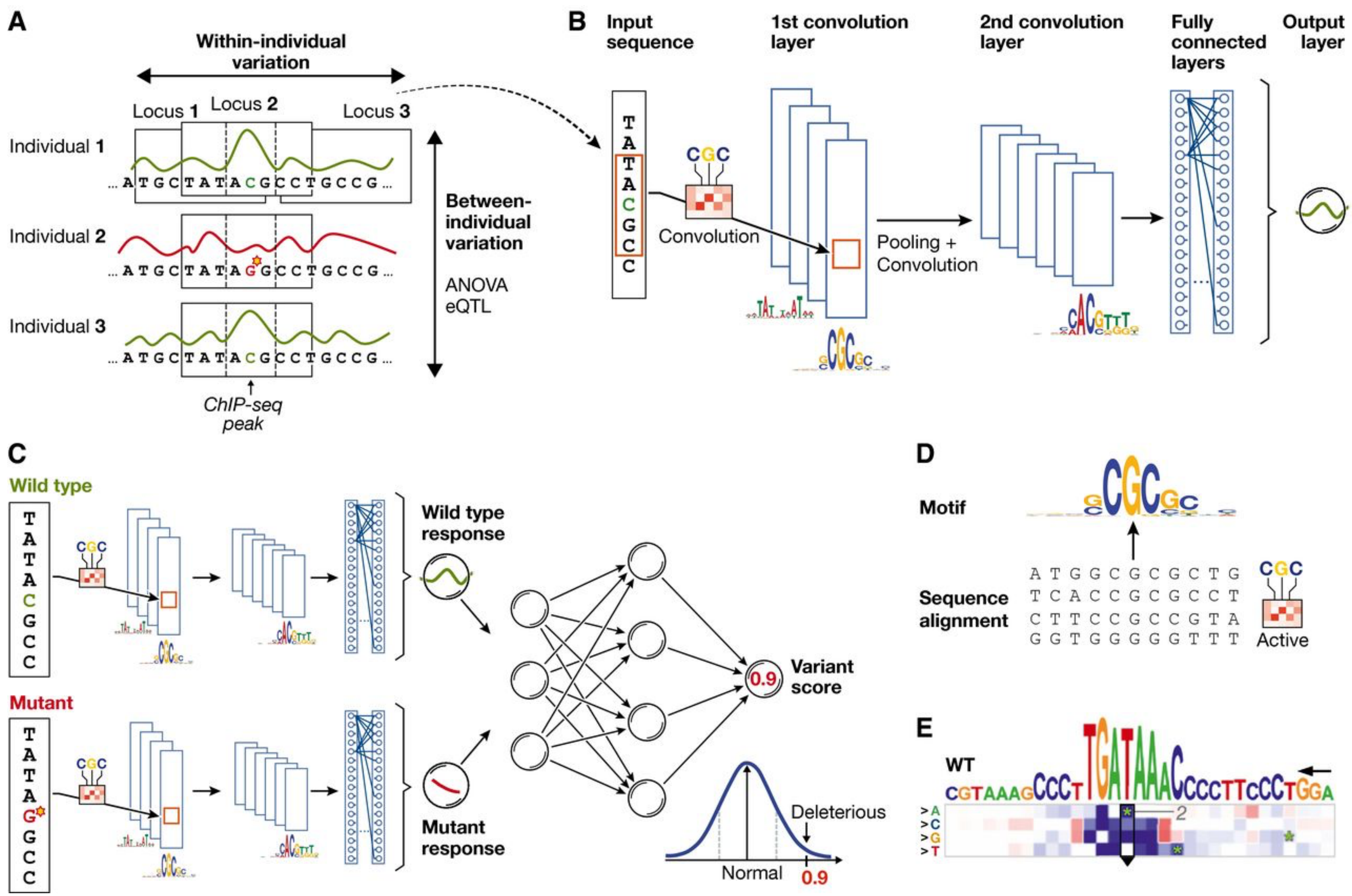
Christof Angermueller et al. Mol Syst Biol 2016;12:878

The Incredible Convergence Of Deep Learning And Genomics



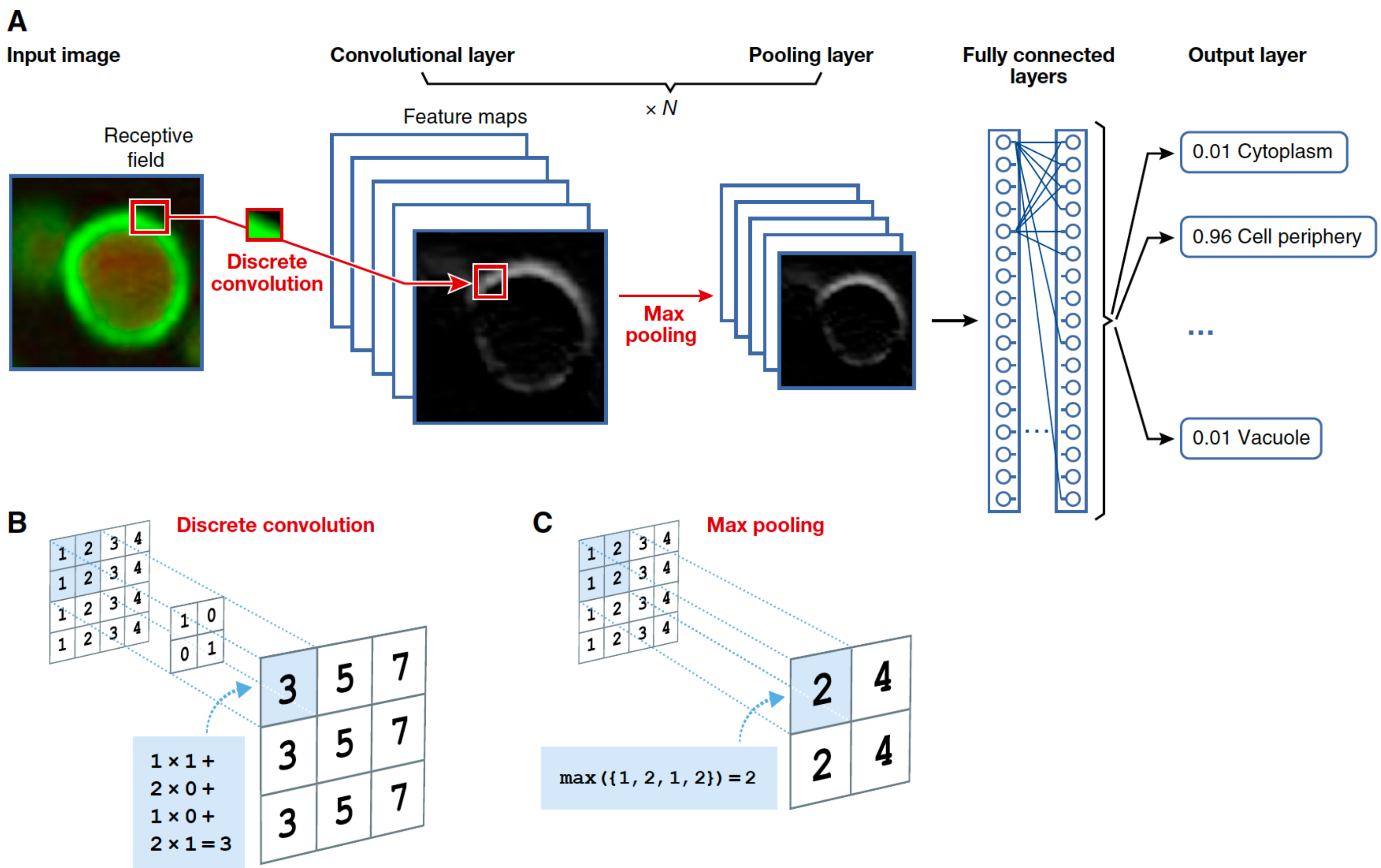
Johnny Israeli

Principles of using neural networks for predicting molecular traits from DNA sequence



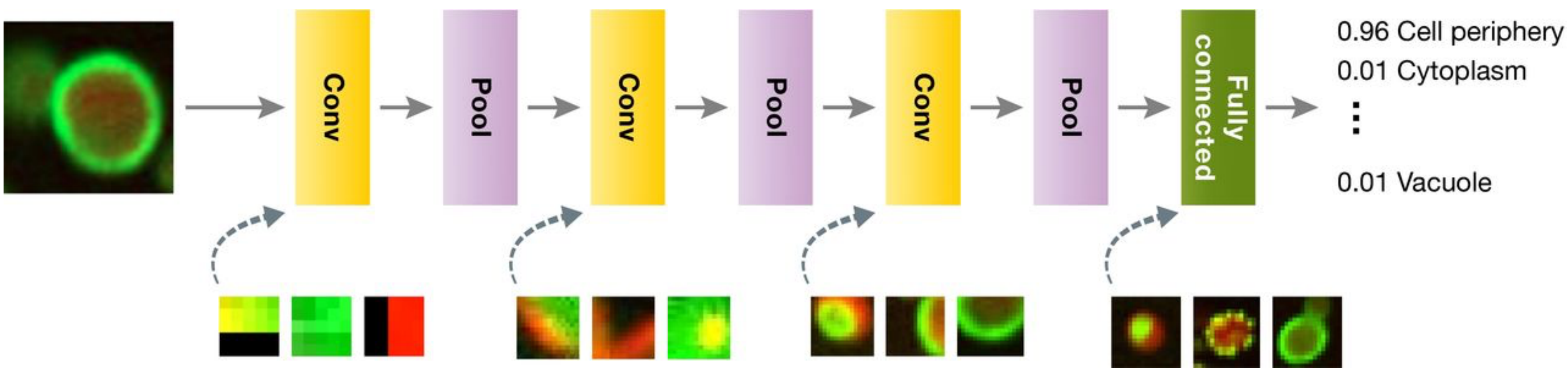
 Christof Angermueller et al. Mol Syst Biol 2016;12:878

Convolutional Neural Network



Christof Angermueller et al. Mol Syst Biol 2016;12:878

Convolution and pooling operators are stacked, thereby creating a deep network for image analysis



Christof Angermueller et al. Mol Syst Biol 2016;12:878

A pre-trained network can be used as a generic feature extractor

First layer features

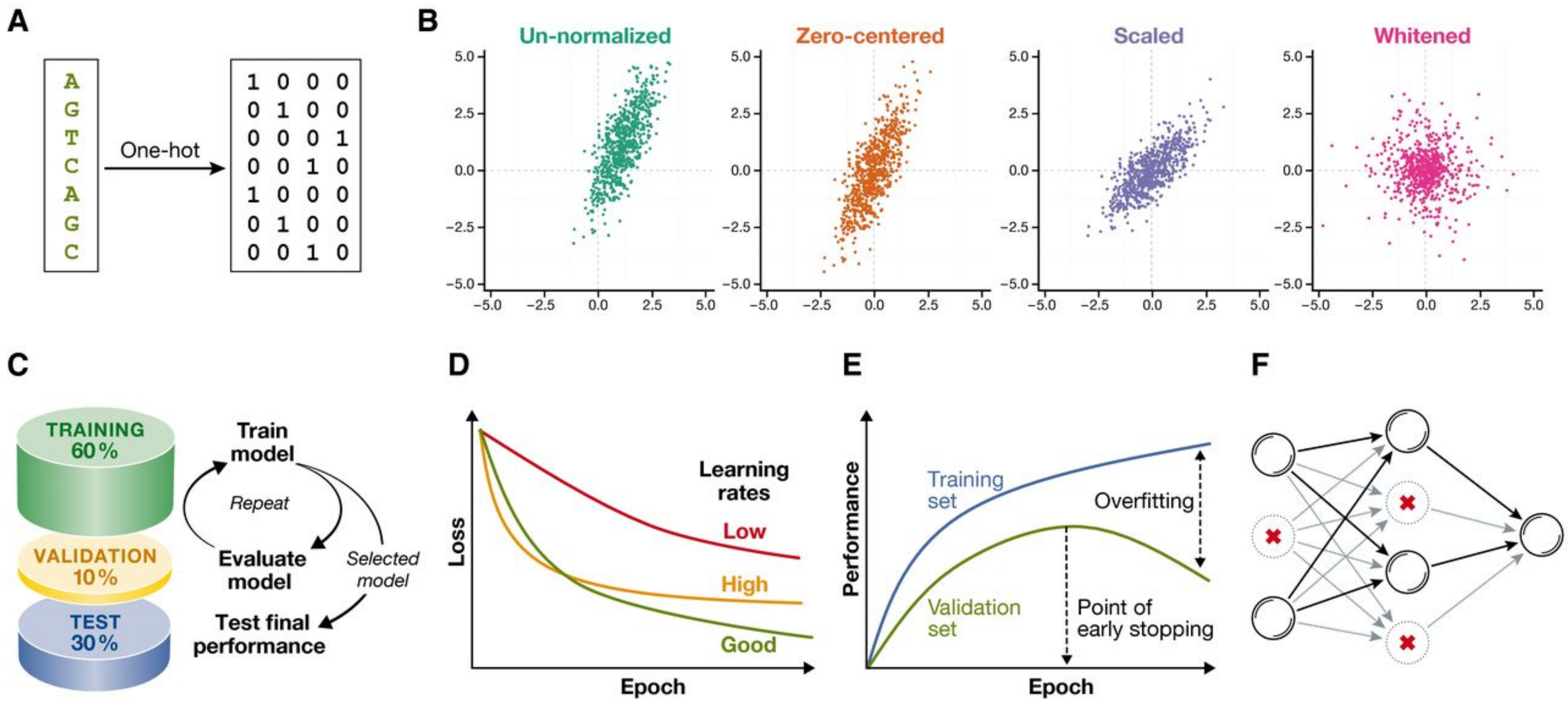
	In top left?	In top right?	…	In bottom right?
	0.21	0.24		0.01
	0.02	0.01		0.25
	0.01	0.03		0.19

Third layer features

	In left?	In right?	…	In bottom?
	2.51	0.02		2.92
	0.03	0.01		0.02
	0.02	0.01		0.01

Christof Angermueller et al. Mol Syst Biol 2016;12:878

Data normalization for and pre-processing for deep neural networks



Christof Angermueller et al. Mol Syst Biol 2016;12:878

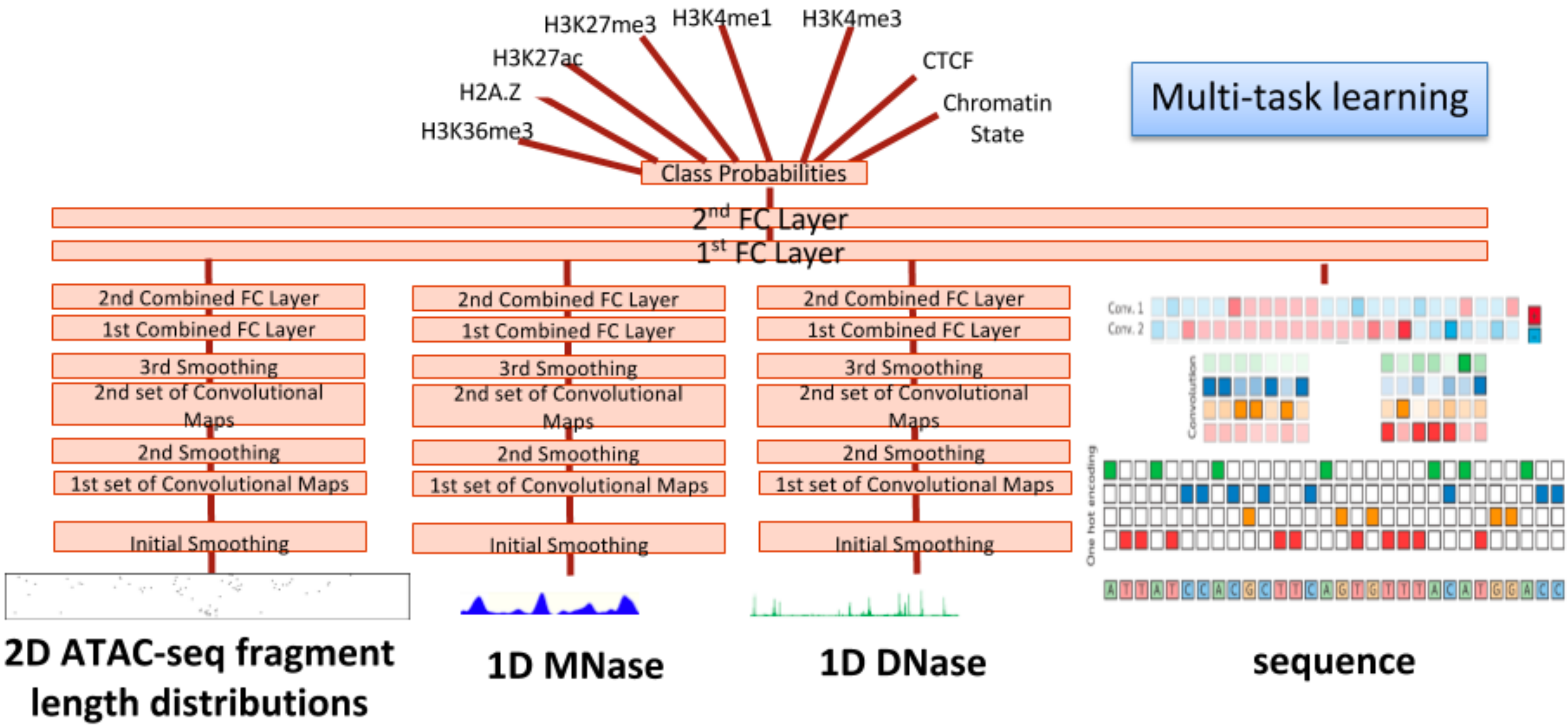
Overview of existing deep learning frameworks, comparing four widely used software solutions

	Caffe	Theano	Torch7	TensorFlow
Core language	C++	Python, C++	LuaJIT	C++
Interfaces	Python, Matlab	Python	C	Python
Wrappers		Lasagne, Keras, sklearn-theano		Keras, Pretty Tensor, Scikit Flow
Programming paradigm	Imperative	Declarative	Imperative	Declarative
Well suited for	CNNs, Reusing existing models, Computer vision	Custom models, RNNs	Custom models, CNNs, Reusing existing models	Custom models, Parallelization, RNNs

Christof Angermueller et al. Mol Syst Biol 2016;12:878

THE CHROMPUTER

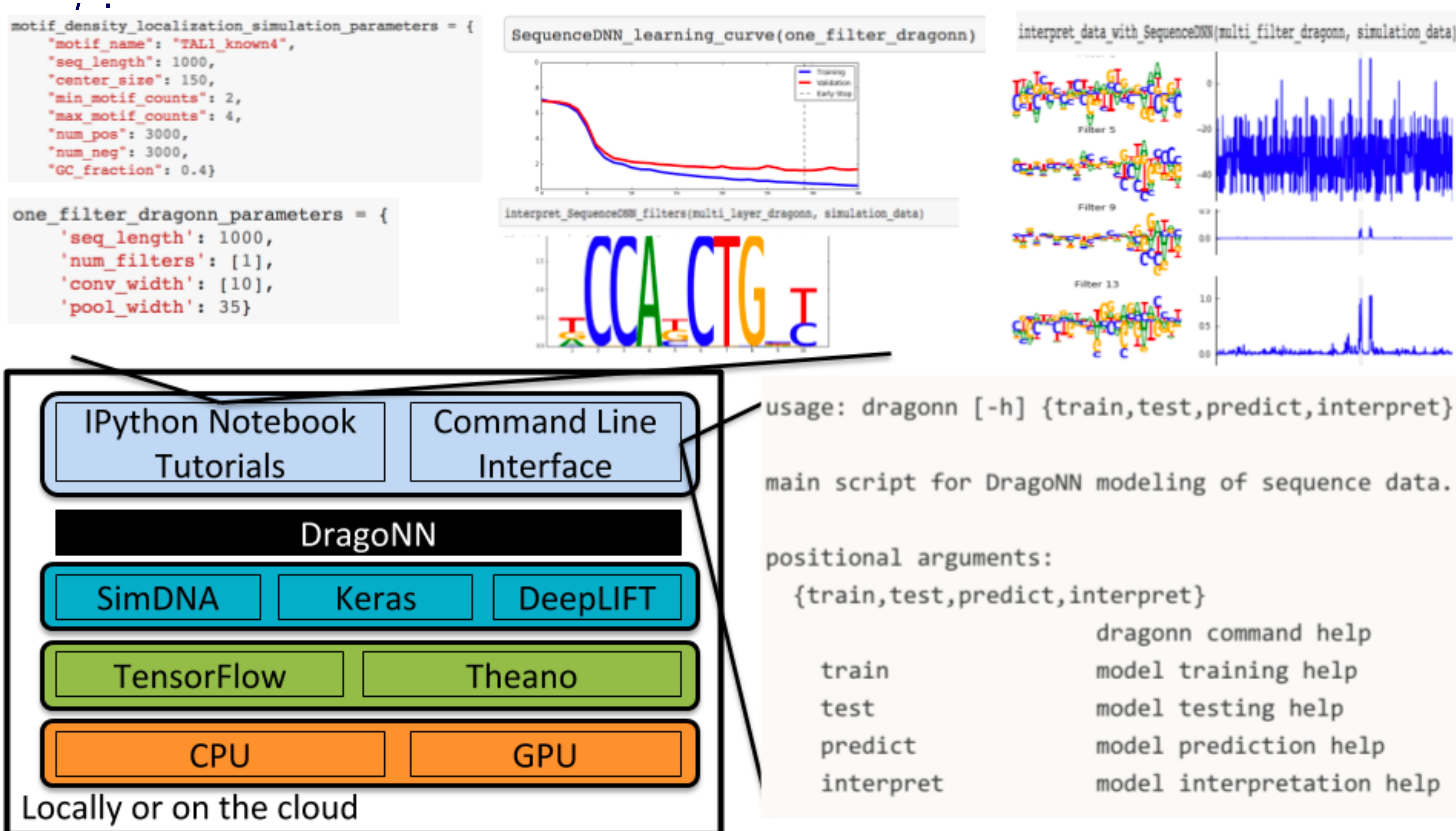
Integrating 1D, 2D signals, and sequence to **predict multiple outputs**



Johnny Israeli

How to train your DragoNN

- https://drive.google.com/file/d/0B4Yo77Kh_QeeaXZKQUtZWjNrWkE



Johnny Israeli