Integrated Machine Learning and Stochastic Modeling for Understanding Disease-related Cell Signaling and Regulation

Juan Cui, Assistant Professor
System Biology and Biomedical Informatics (SBBI) Laboratory
Computer Science and Engineering, UNL

Outline

- Introduction to Bioinformatics
- Big data challenges
- Example projects
- Future perspectives
- Bioinformatics program at CSE

A Deep Connection Between Biology and Information Science

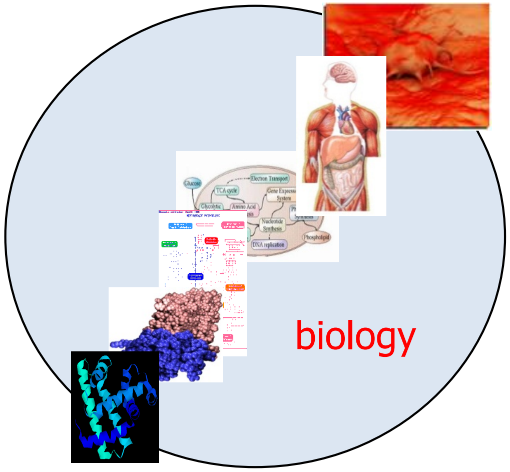
molecular biology is information science ....

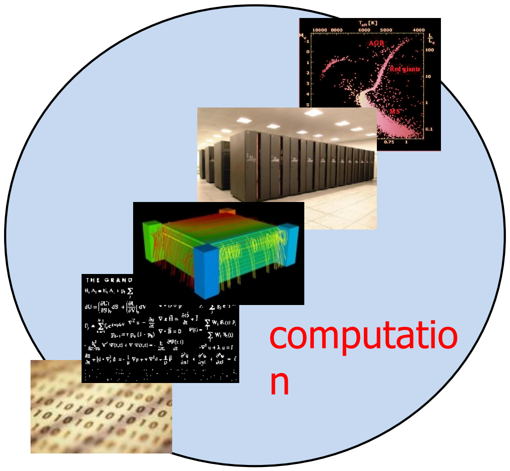
\- Leroy Hood (founder of omics science and system biology)

…since the fundamental information in life is digital!

cgtacgtacgtagagtgctagtctagtcgtagcgccgtagtcgatcgtgtgggtagtagctgatatgatg
cgaggtaggggataggatagcaacagatgagcggatgctgagtgcagtggcatgcgatgtcgatga
tagcggtaggtagacttcgcgcataaagctgcgcgagatgattgcaaagragttagatgagctgatgc
tagaggtcagtgactgatgatcgatgcatgcatggatgatgcagctgatcgatgtagatgcaataagtc
gatgatcgatgatgatgctagatgatagctagatgtgatcgatggtaggtaggatggtaggtaaattgat
agatgctagatcgtaggta…………………………………

Bioinformatics





- This interdisciplinary science … is about *providing computational support to studies on linking the behavior of cells, organisms and populations to the information encoded in the genomes*.

-- Temple Smith, *Current Topics in Computational Molecular Biology (2002)*

Bioinformatics (Computational Biology)

data management; data mining; modeling; prediction; theory formulation

bioinformatics

genes, proteins, protein complexes, pathways, cells, organisms, ecosystem

an indispensable part of biological science

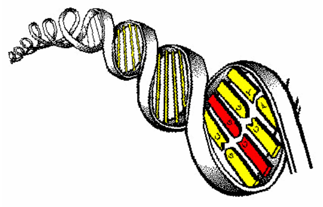
engineering
aspect

scientific
aspect

computer science, biology, medicine, statistics,
mathematics, physics, chemistry, engineering,…

What Bioinformatics Has Done

- Human genome project (1986 – 2003)

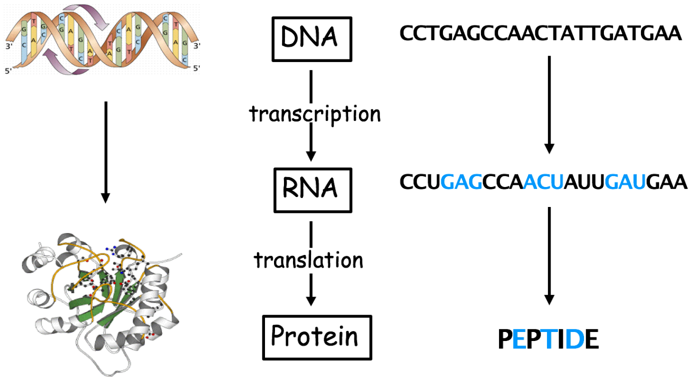


cgtacgtacgtagagtgctagtctagtcgtagcgccgtagtcgatcgtgtgggtagtagctgatatgat
gcgaggtaggggataggatagcaacagatgagcggatgctgagtgcagtggcatgcgatgtcgat
gatagcggtaggtagacttcgcgcataaagctgcgcgagatgattgcaaagragttagatgagctga
tgctagaggtcagtgactgatgatcgatgcatgcatggatgatgcagctgatcgatgtagatgcaata
agtcgatgatcgatgatgatgctagatgatagctagatgtgatcgatggtaggtaggatggtaggtaa
attgatagatgctagatcgtaggta……………………………….

What does this "text" say?

What Bioinformatics Has Done

- From the basic biology, we know the human genome encodes genes, their regulatory systems;
- More specifically, it encodes
  - The color of our eyes
  - The disease we may develop at certain age
  - How tall we will become
  - …



DNA → transcription → RNA → translation → Protein

CCTGAGCCAACTATTGATGAA

CCUGAGCCAACUAUUGAUGAA

PEPTIDE

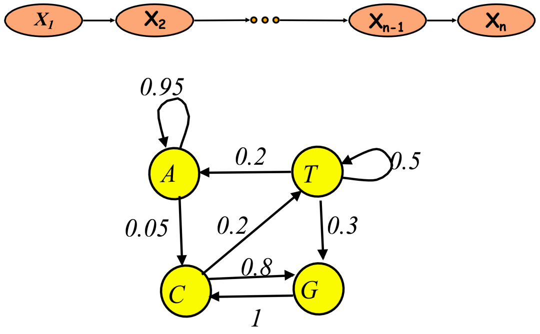
- … but how can I read this "text"？

- Gene: discrete units of hereditary information located on the chromosomes and consisting of DNA.

What Bioinformatics Has Done

- Through statistical analysis, scientists learned that protein-encoding genes follow 5th order Markov chain models while other regions do not.

- This observation laid the foundation for computational scientists to develop computer programs to find genes in a genome.


1st order Markov chain

Find Genes in A Genome

- A simplified version: the six-letter words (e.g., AAGTGC) have different frequencies in genes from non-gene regions

Frequency in genes (AAA ATT) = 1.4%;
Frequency in genes (AAA GAC) = 1.9%;
Frequency in genes (AAA TAG) = 0.0%;

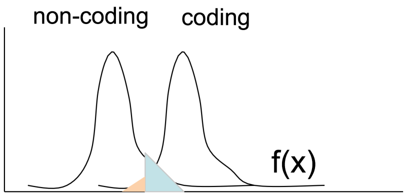
Frequency in non-genes (AAA ATT) = 5.2%
Frequency in non-genes (AAA GAC) = 4.8%
Frequency in non-genes (AAA TAG) =6.3%

AAAATTAAAATTAAAGACAAAATTAAAGACAAACACAAAATTAAATAGAAATAGAAAATT …..

Is this a gene or non-gene region if you have to make a bet?

Find Genes in A Genome

- Preference model:
  - for each 6-letter word X (e.g., AAA AAA), calculate its frequencies in gene and non-gene regions, FC(X), FN(X)
  - calculate X's preference value P(X) = log (FC(X)/FN(X))



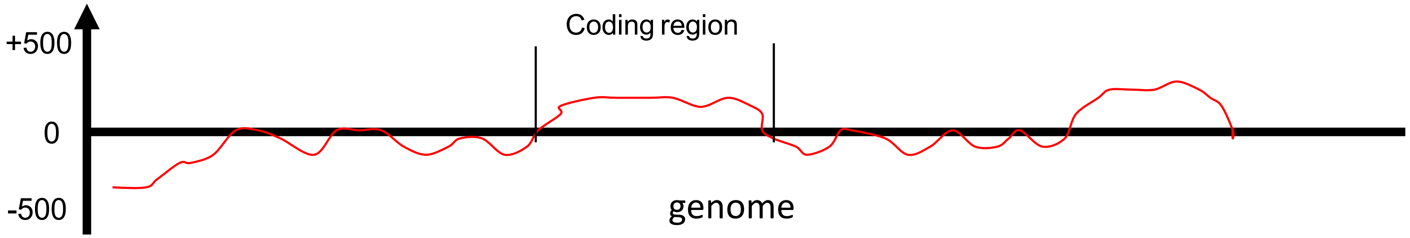
- Properties:
  - P(X) is 0 if X has the same frequencies in gene and non-gene regions
  - P(X) has positive score if X has higher frequency in gene than in non-gene region; the larger the difference, the more positive the score is
  - P(X) has negative score if X has higher frequency in non-gene than in gene region; the larger the difference, the more negative the score is

Find Genes in A Genome

Gene prediction: given a DNA region, calculate the sum of P(X) values for all 6-letter words X in the region;

- if the sum is larger than zero, predict "gene"
- otherwise predict non-gene

AAAATTAAAATTAAAGACAAAATTAAAGACAAACACAAAATTAAATAGAAATAGAAAATT .....

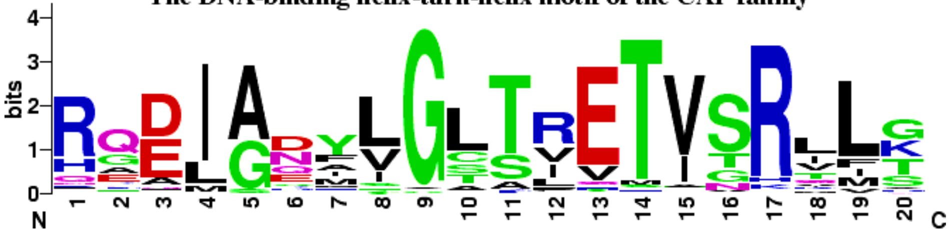


Regulatory Elements in A Genome

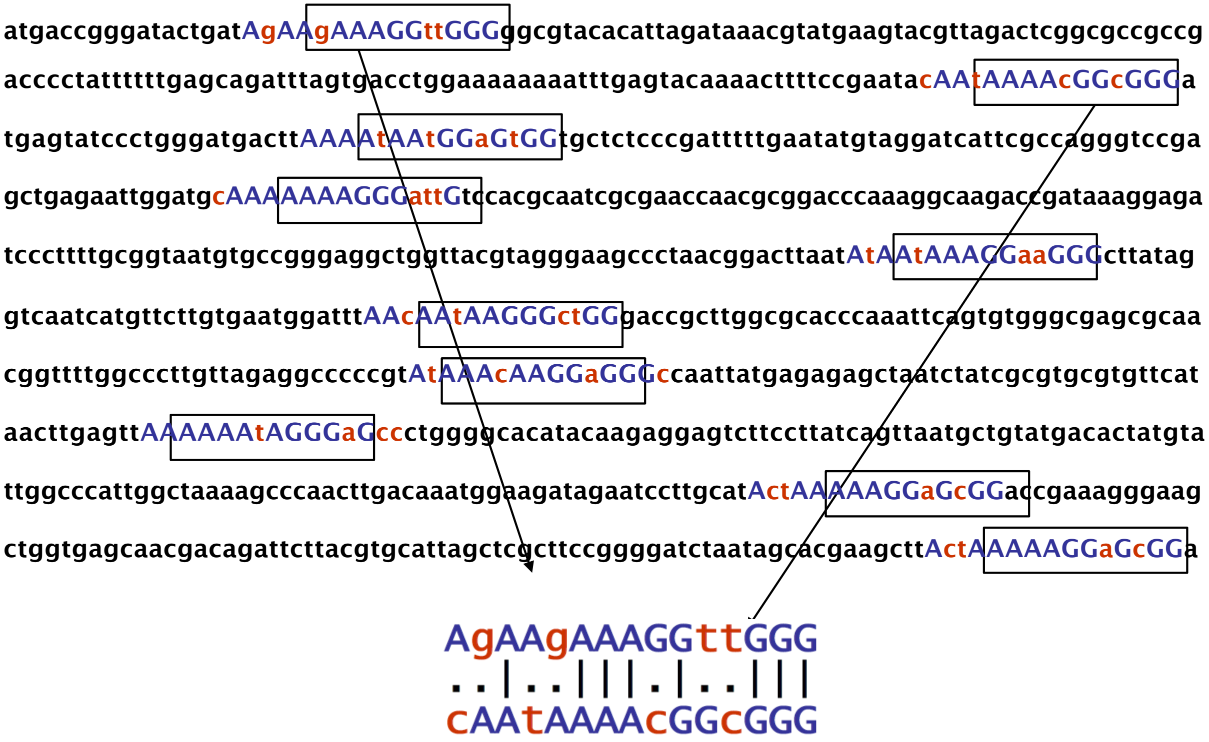
- Motif finding among aligned DNA sequence of the same length

- Often represented as positional frequency matrix/profile

```
LTMTRGDIGNYLGLTVETISRLLGRFQKSGML
LTMTRGDIGNYLGLTIETISRLLGRFQKSGMI
LTMTRGDIGNYLGLTVETISRLLGRFQKSEIL
LTMTRGDIGNYLGLTVETISRLLGRLQKMGIL
LAMSRNEIGNYLGLAVETVSRVFSRFQQNELI
LAMSRNEIGNYLGLAVETVSRVFTRFQQNGLI
LPMSRNEIGNYLGLAVETVSRVFTRFQQNGLL
VRMSREEIGNYLGLTLETVSRLFSRFGREGLI
LRMSREEIGSYLGLKLETVSRTLSKFHQEGLI
LPMCRRDIGDYLGLTLETVSRALSQLHTQGIL
LPMSRRDIADYLGLTVETVSRAVSQLHTDGVL
LPMSRQDIADYLGLTIETVSRTFTKLERHGAI
```


The DNA-binding helix-turn-helix motif of the CAP family

Motif Finding Based on Unaligned Sequence

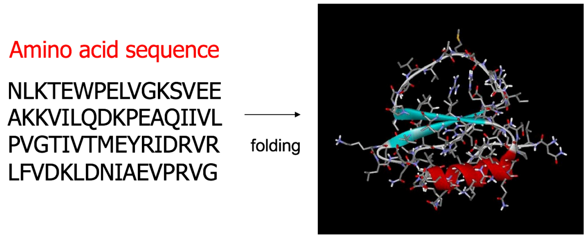


atgaccgggatactgatAgAAgAAAGGttGGGggcgtacacattagataaacgtatgaagtacgttagactcggcgccgccg
acccctattttttgagcagatttagtgacctggaaaaaaaatttgagtacaaaacttttccgaatacAAtAAAAcGGcGGGa
tgagtatccctgggatgacttAAAAtAAtGGaGtGGtgctctcccgatttttgaatatgtaggatcattcgccagggtccga
gctgagaattggatgcAAAAAAAGGGattGtccacgcaatcgcgaaccaacgcggacccaaaggcaagaccgataaaggaga
tcccttttgcggtaatgtgccgggaggctggttacgtagggaagccctaacggacttaatAtAAtAAAGGaaGGGcttatag
gtcaatcatgttcttgtgaatggatttAAcAAtAAGGGctGGgaccgcttggcgcacccaaattcagtgtgggcgagcgcaa
cggttttggcccttgttagaggcccccgtAtAAAcAAGGaGGGccaattatgagagagctaatctatcgcgtgcgtgttcat
aacttgagttAAAAAAtAGGGaGccctggggcacatacaagaggagtcttccttatcagttaatgctgtatgacactatgta
ttggcccattggctaaaagcccaacttgacaatggaagatagaatccttgcatActAAAAAGGaGcGGaccgaaagggaag
ctggtgagcaacgacagattcttacgtgcattagctcgcttccggggatctaatagcacgaagcttActAAAAAGGaGcGGa

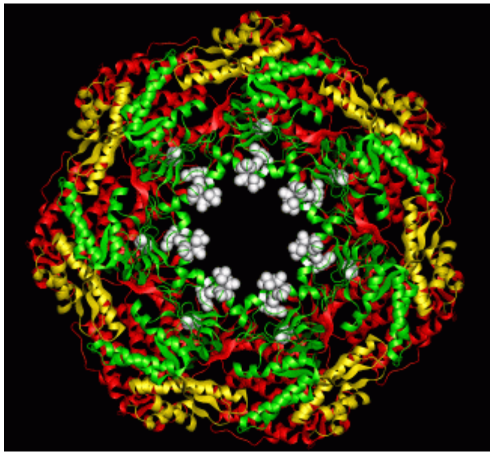
```
AgAAgAAAGGttGGG
..|..|||.|..|||
cAAtAAAAcGGcGGG
```

What Bioinformatics Has Done

- Now scientists have developed computer programs for prediction of
  - Structure of protein
  - Functions of proteins
  - Protein-protein interaction partners
  - Regulatory systems of various proteins and RNAs
  - Higher-level functional elements such as biochemical reaction systems such as how cancers may have developed
  - ……


Amino acid sequence

NLKTEWPELVGKSVEE
AKKVILQDKPEAQIIVL
PVGTIVTMEYRIDRVR
LFVDKLDNIAEVPRVG



What Bioinformatics Can Do

- The list is LONG
  - ☐ applications in biomedical industry
  - ☐ applications in energy industry
  - ☐ applications in agriculture
  - ☐ applications in environmental issues
  - ☐ …….

Big Data (3Vs)

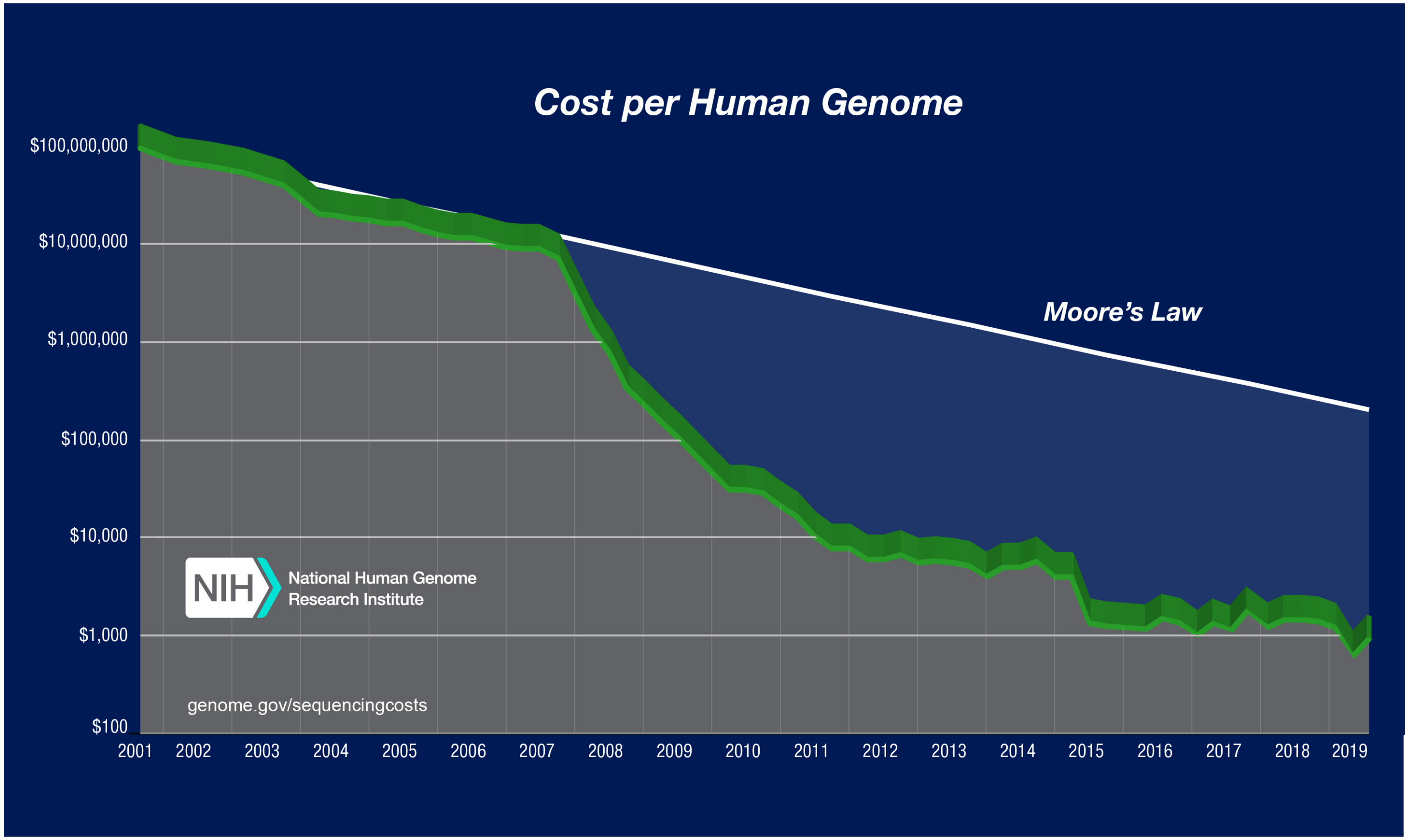
Variety:	**Complexity of data in many different structures**
Velocity:	**Streaming data and large volume data movement**
Volume:	**Scale from Terabytes to Petabytes (1K TBs) to Zetabytes (1B TBs)**

Too big,
too unstructured,
too many different sources

NSF: http://www.nsf.gov/news/news_summ.jsp?cntn_id=123607

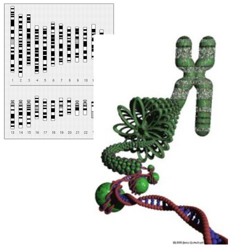
health-related data is expected to double every 73 days by 2020

Sequencing Is Becoming Much Faster and Cheaper

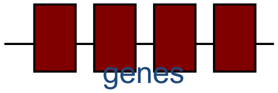

Cost per Human Genome

genome.gov/sequencingcosts

Other Types of Biological Data



chromosomes

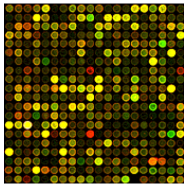


genes

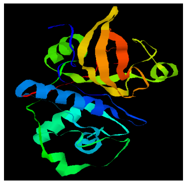
Genomics

(Transcriptomics)

atgcatgcatggatgatgcagctgatcgatgtag
atgcaataagtcgatgatcgatgatgatgctag
atgatagctagatgtgatcgatggtaggtaggat
ggtaggtaaattgatagatgctagatcgtaggta
……

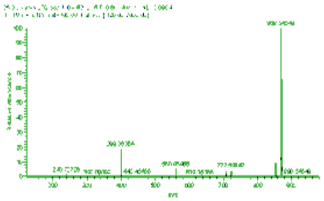


Others:
Image omics
Phenomics
…

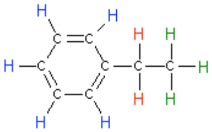


proteins

Proteomics

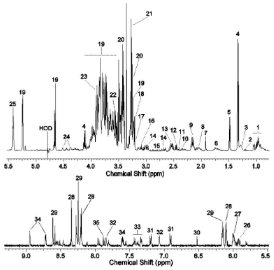


Molecular changes and cellular states can be measured through High throughput Omics techniques



metabolites

Metabolomics



Translational Research & Systems Biology

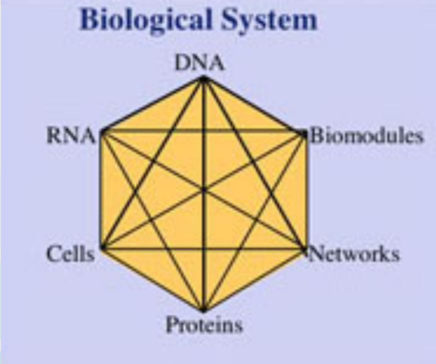
Genomics

single elements

Transcriptomics

Proteomics

Metabolomics



Epigenomics

Systems Biology

1990 1995 2000 2005 2010 2015 2020

Decoding genomes

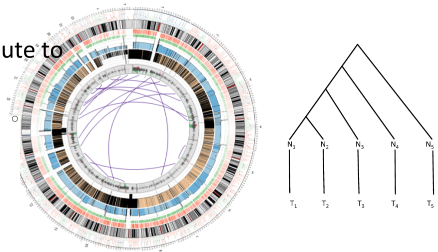
integrating elements with a systems approach

systems approach for translational research
(computational systems biology)

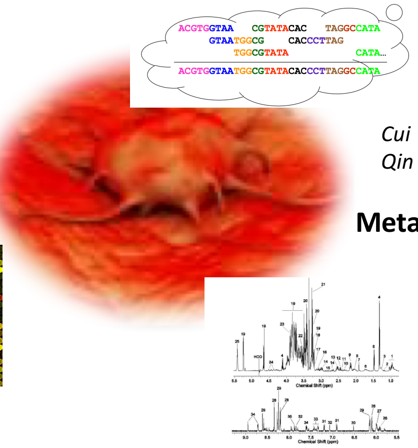
Data-Driven Disease Studies: Successes and Promises

Genome Study

What types of genetic mutation can contribute to cancer formation and progression?
How cancer genomes evovle?

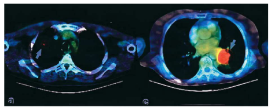


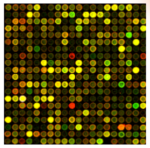
Cui J. et al, International J. Cancer 2014
Qin Ma et al. Nucleic Acids Research 2013



Biomarker Discovery

Can we find a gene or protein with altered expression in cancer versus control?

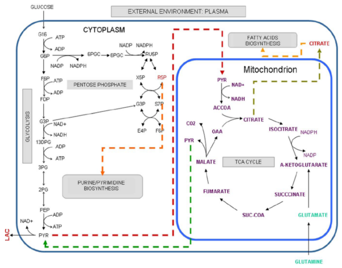




Cui J., et al, Nucleic Acids Res. 2010
Hong S., Cui J., et al, PLoS ONE, 2011
Dong X., et al, Diagnostic pathology , 2013
Cui J., et al, Bioinformatics, 2008
Q Liu, J Cui et. al, BMC bioinformatics, 2009

Metabolic Network

How ATP-production works in cancer?



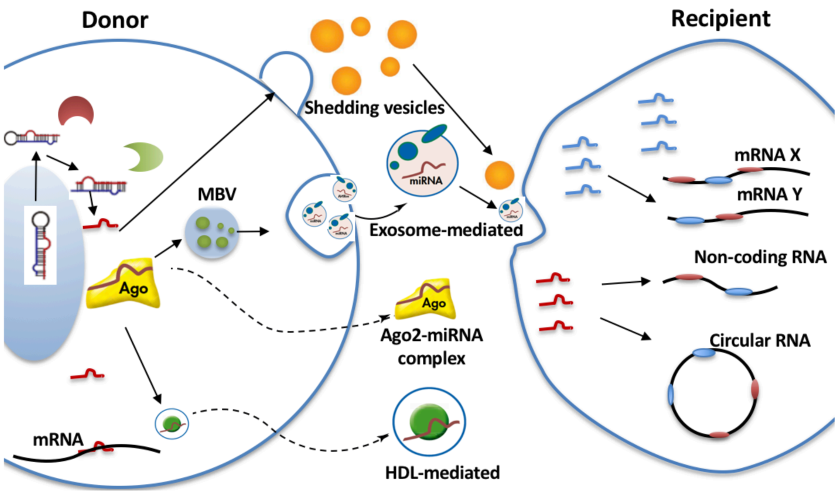
Cui, et. al., J. Molecular Cell Biology, 2012

“Utilizing the molecular changes we observed in cancer to make discoveries towards understanding cancer behavior”

Integrated Machine Learning and Stochastic Modeling for Understanding Disease-related Cell Signaling and Regulation

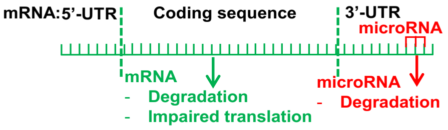
MicroRNAs and Gene Regulation

- Small non-coding RNAs, ~21nt long.
- Post-transcriptional gene silencing in eukaryotes; within or cross cells.
- In human, ~2K microRNAs regulate 70% human genes.
- Functional study is largely depending on the reliable identification of gene targets.
  - Computational predictions based on sequence and structural features
  - Cooperative binding and competitive binding should be integrated



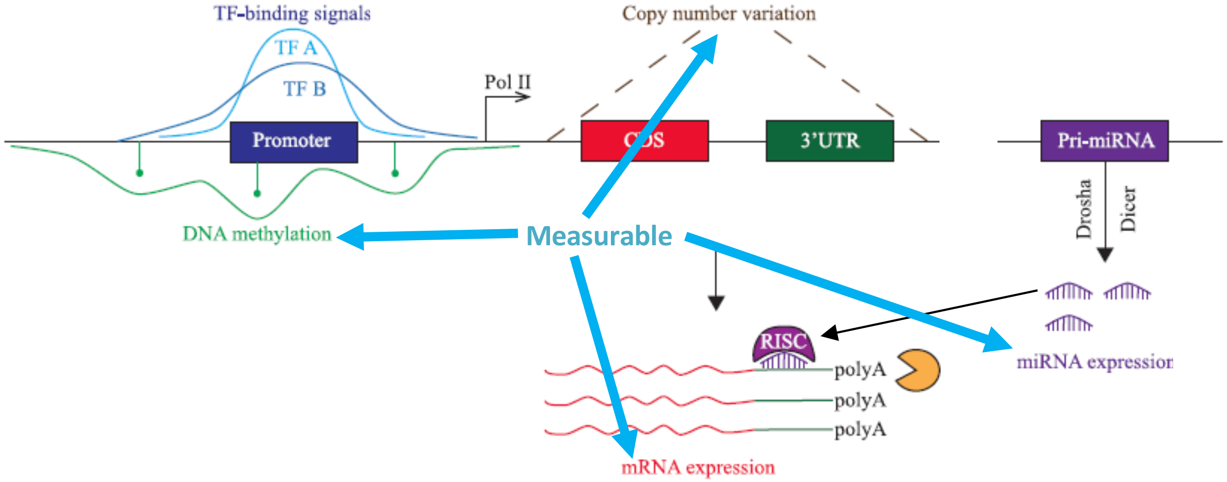
Schematic diagram of miRNA transfer between cells and competitive miRNA binding

Static



Dynamic

Regulatory Mechanisms on Gene Expression



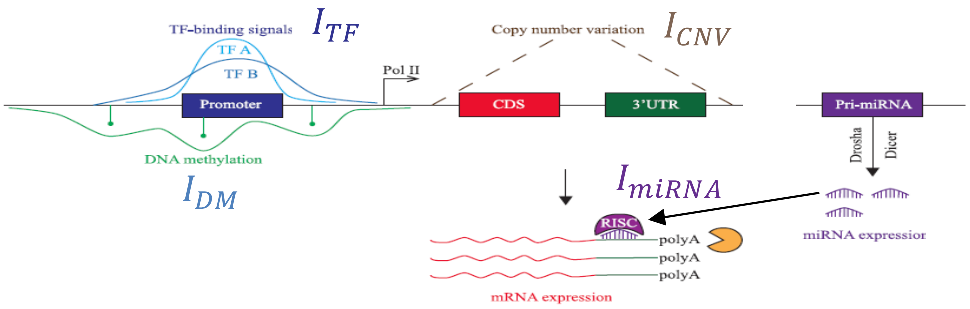
Measurable

The Model

Measurable inputs

$$\boxed{\Delta E_{mRNA} \propto I_{CNV} + I_{DM}} + \boxed{I_{TF} + I_{miRNA}}$$

Interaction based factors

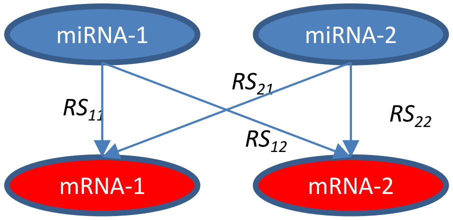


The Model

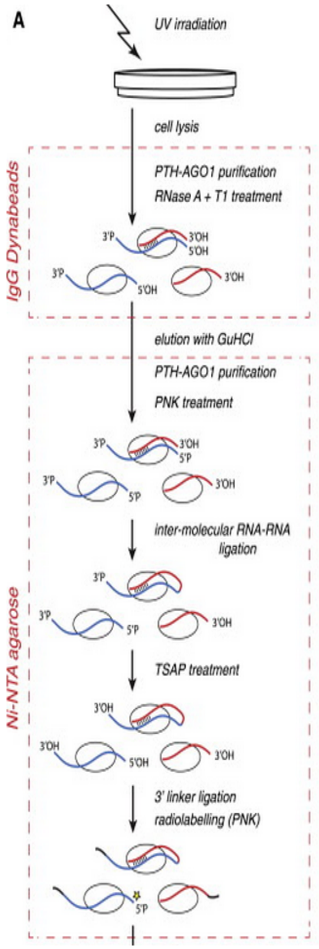
$$\overset{\text{Measurable inputs}}{\boxed{\Delta E_{mRNA} \propto I_{CNV} + I_{DM}}} + \underset{\text{Interaction based factors}}{\boxed{I_{TF} + I_{miRNA}}}$$

To estimate I_{miRNA}

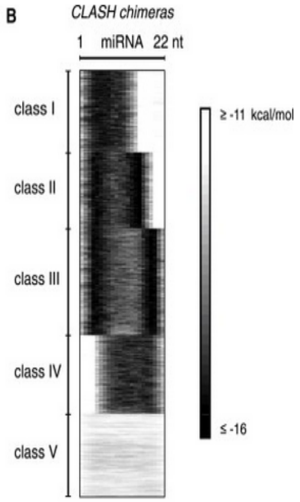
- ☐ Quantification of regulatory potentials between miRNA and its targets
- ☐ Integration of competitive regulation among regulators and targets

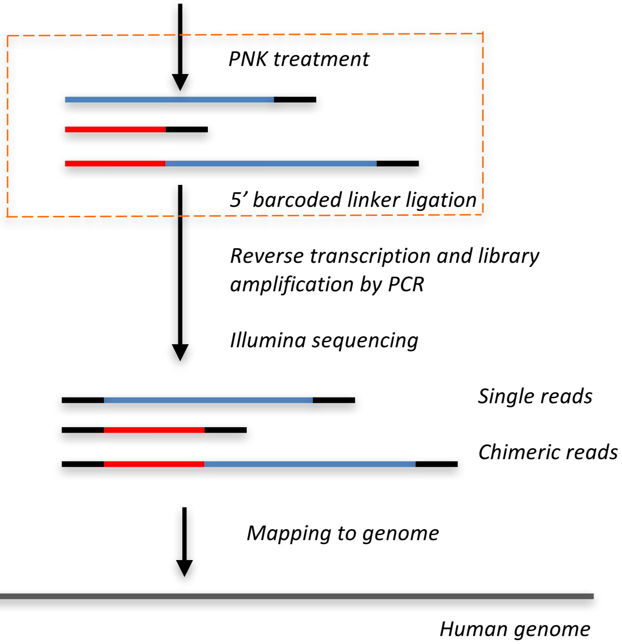


Sequencing-Derived MicroRNA Interactions



- Capture AGO-associated RNAs
  - Free microRNAs and mRNAs
  - microRNA-mRNA interaction




PNK treatment

5’ barcoded linker ligation

Reverse transcription and library amplification by PCR

Illumina sequencing

Single reads

Chimeric reads

Mapping to genome

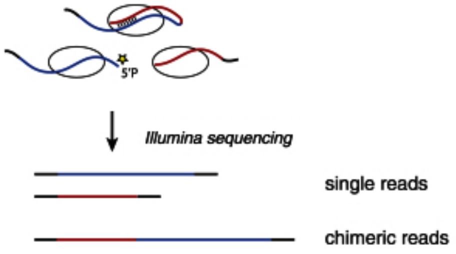
Human genome

Helwak, et al. , Cell. 2013

CLASH: crosslinking, ligation, and sequencing of hybrids

Estimation of Regulatory Potentials

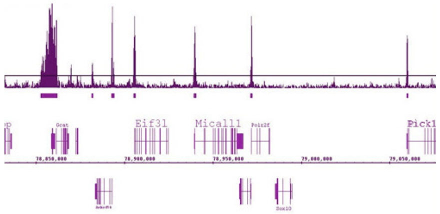
- Re-analysis of the CLASH data to identify significant interactions based on Binomial test
- Calculation of regulatory scores based on the interactions detected for miRNA *i* and mRNA *j*



Helwak *et al.*, Cell, 2013

$$RS_{ij} = \frac{1}{K}\sum_{k=1}^{K}\left(|MFE_k| \times P_{ijk}\right) \qquad RS_{ij}' = RS_{ij}/RS_{i}.$$

- *Integration of TF regulation*
  - ☐ d_{tjk} is the distance between binding site and transcription start site (TSS)

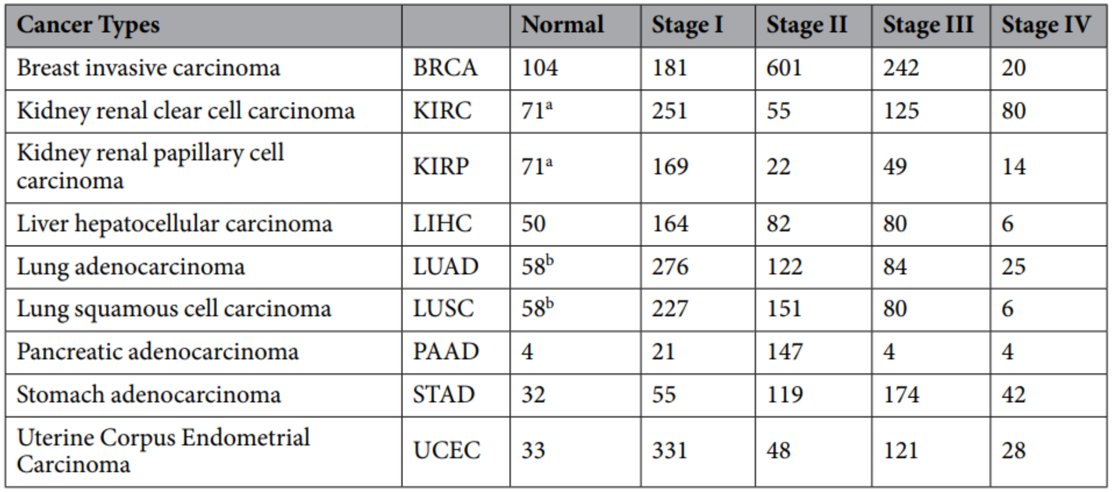


$$RS_{tjk} = \exp\left(-\left(\frac{1}{2} + \frac{4d_{tjk}}{10^5}\right)\right) \qquad RS_{tj} = 1 - \prod_{k=1}^{K}\left(1 - RS_{tjk}\right) \qquad RS_{tj}' = RS_{tj}/RS_{t}.$$

Wang *et al.,* Nat. Protoc., 2013;
Jiang *et al.*, Proc Natl Acad Sci USA, 2015.

Case Study on Various Types of Cancer

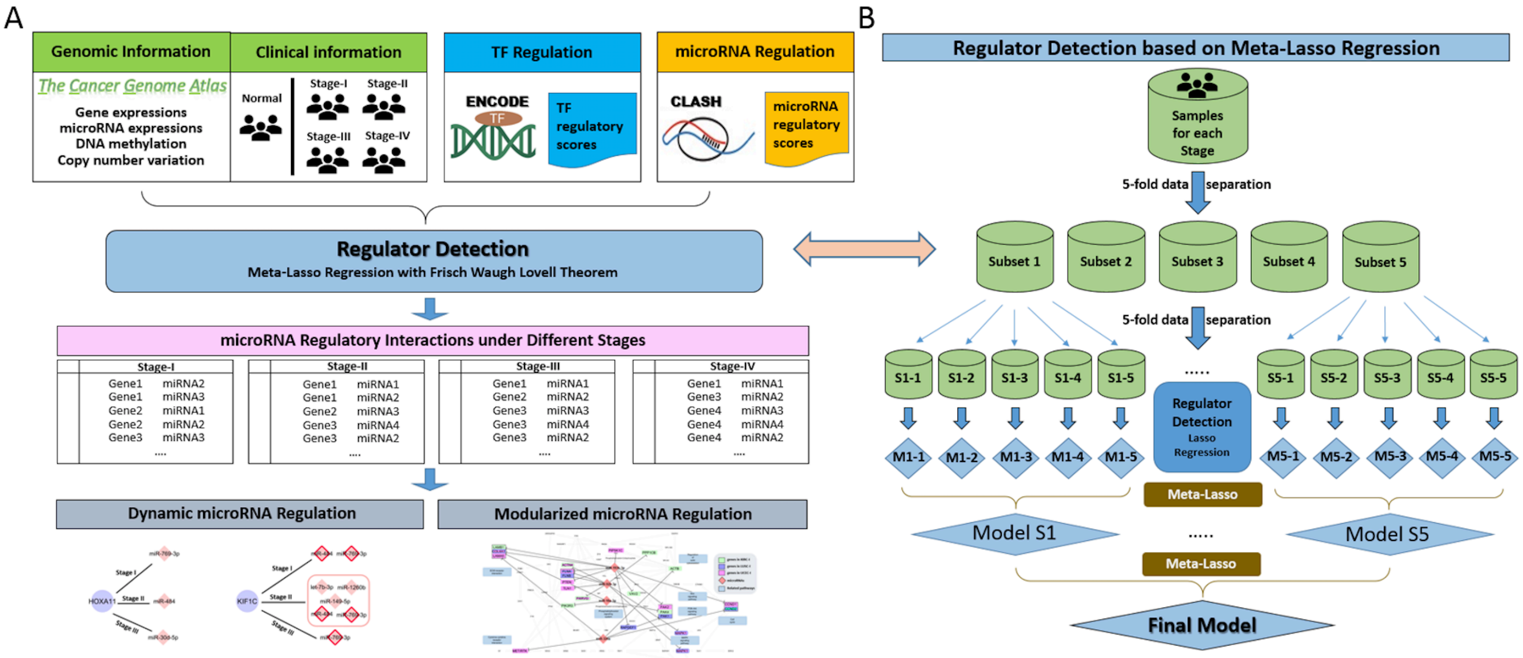
- Genomic data from >4,200 patients with 9 cancer types (from The Cancer Genome Atlas)



Cancer Types		Normal	Stage I	Stage II	Stage III	Stage IV
Breast invasive carcinoma	BRCA	104	181	601	242	20
Kidney renal clear cell carcinoma	KIRC	71[a]	251	55	125	80
Kidney renal papillary cell carcinoma	KIRP	71[a]	169	22	49	14
Liver hepatocellular carcinoma	LIHC	50	164	82	80	6
Lung adenocarcinoma	LUAD	58[b]	276	122	84	25
Lung squamous cell carcinoma	LUSC	58[b]	227	151	80	6
Pancreatic adenocarcinoma	PAAD	4	21	147	4	4
Stomach adenocarcinoma	STAD	32	55	119	174	42
Uterine Corpus Endometrial Carcinoma	UCEC	33	331	48	121	28

Shu and Cui, Sci Rep., 2017

Modeling Dynamic MicroRNA Regulation



(A) The identification pipeline of conditional miRNA regulatory interactions; (B) Meta-Lasso Regression utilized to detect the microRNA regulators of genes in each cancer stage.

$$\max_{\beta_0, \boldsymbol{g}, \boldsymbol{\zeta}} \left\{ \sum_{m=1}^{M} \ell_m(\beta_{m0}, \boldsymbol{g}, \boldsymbol{\zeta}_m) - \sum_{j=1}^{p} |g_j| - \lambda \sum_{j=1}^{p} \sum_{m=1}^{M} |\zeta_{mj}| \right\}$$

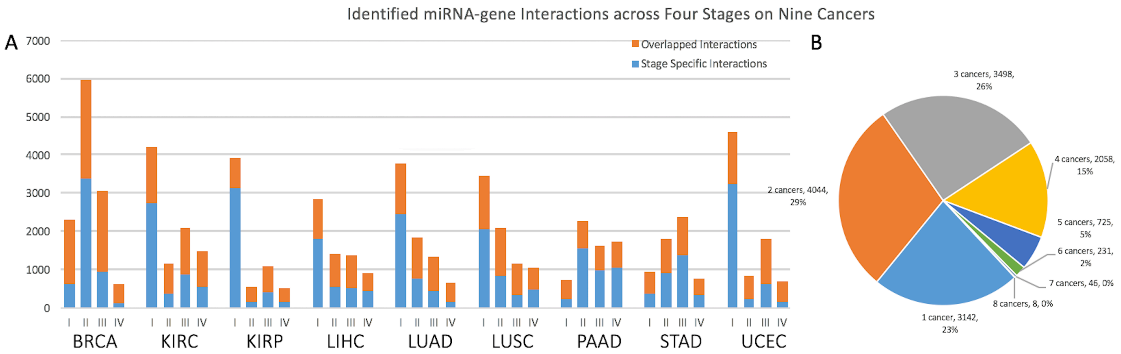
where $l_m\left(\beta_{m_0}, g, \zeta_m\right)$ is the log-likelihood function of the m-th dataset; M denotes the number of individual datasets; g_p is the effect of the p-th regulator (out of P regulators) at the overall condition; and ζ_{mp} is the effect of the p-th regulator at the m-th dataset (out of M datasets).

Shu, et. al., Scientific Reports, 2017

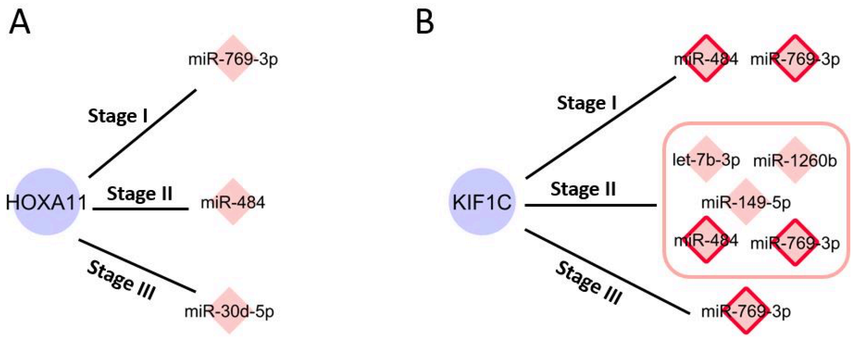
Biometrics. 2014 Dec;70(4):872-80

Conditional Gene Regulation in Human Cancers

- Overview of the miRNA-mRNA interactions identified in nine cancers

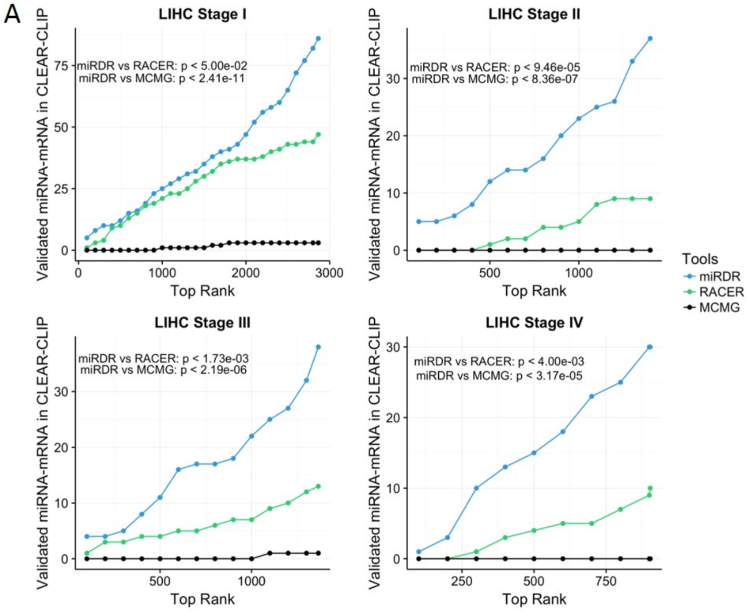


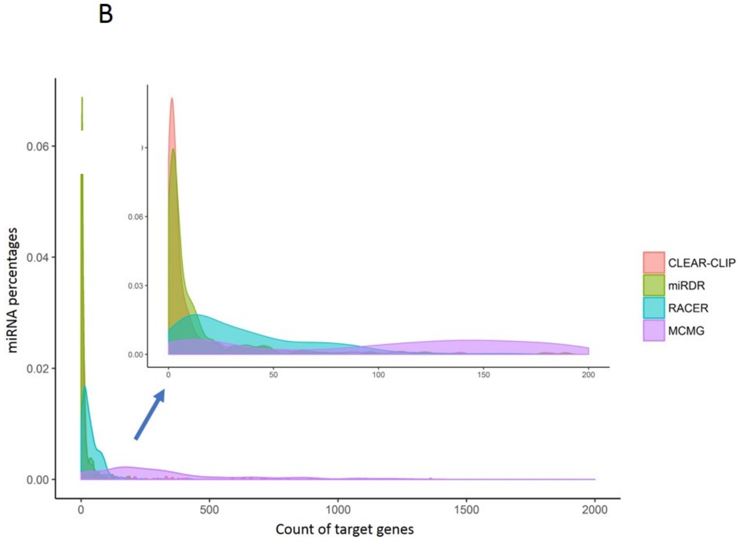
- Illustration of the dynamic miRNA-mediated gene regulation (e.g., kidney cancer)



Shu, et. al. Scientific Reports, 2017

Improved Interaction Prediction





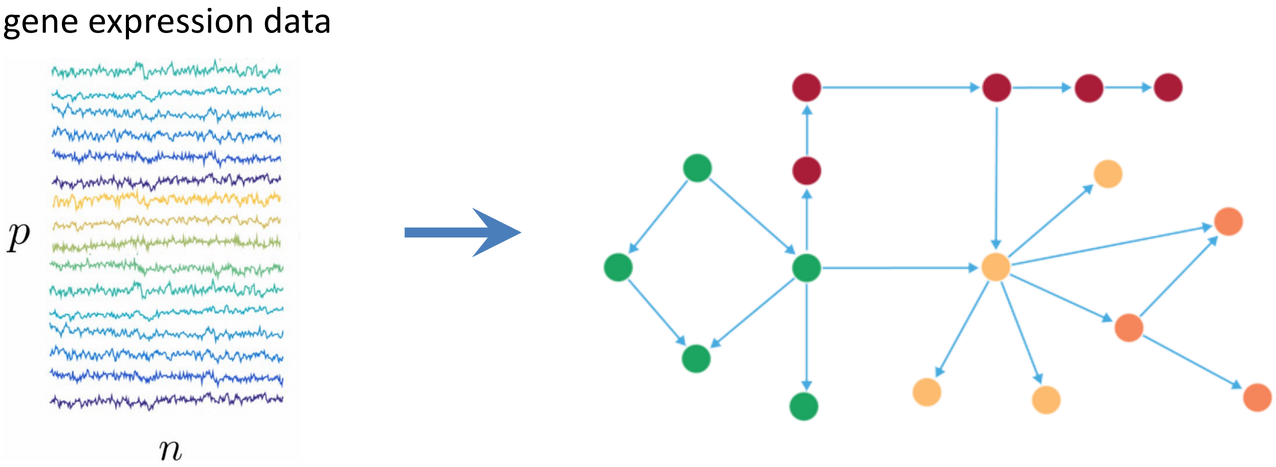
miRDR: *et al. Sci Rep., 2017.;*
MCMG*: Chen et al., Bioinformatics, 2013;*
RACER*: Li et al. Plos Comp Biol, 2014.*

Shu, et. al. Scientific Reports, 2017

Summary

- Our computational studies provide the microRNA community with novel insights in miRNA regulation, particularly:
  - ☐ new solution for significantly-improved identification of conditional miRNA regulatory interactions.
  - ☐ *which is, however, highly focused on microRNA regulation on each individual gene and ignores gene-gene interaction.*

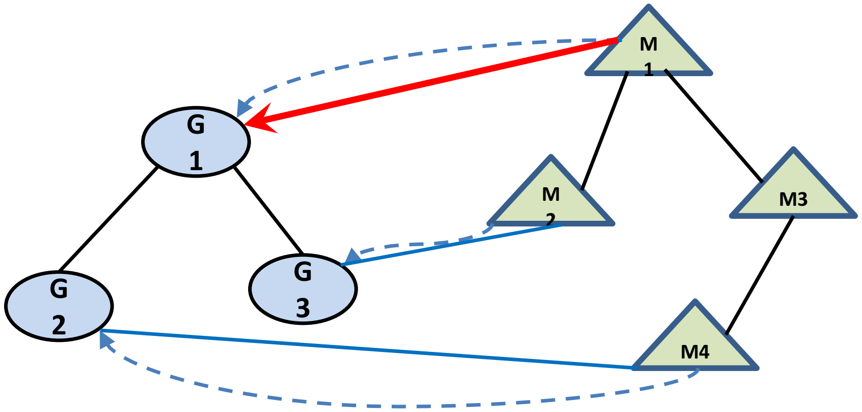
Dynamic Gene Regulation Networks (GRNs)



- **Information fusion**
  - integration of heterogeneous data analyses that reflect distinct regulatory mechanisms.
- **Model fusion**
  - integration of heterogeneous interactions inferred from different models

Connecting Models

- DAGs are converted to its Markov equivalent undirected graphical models (moralized graph)
  - ☐ Adding an undirected edge if there is a directed edge between nodes in original graph or to the parents of same node
- To explain conditional impact of
  microRNA, we test each miRNA and its dependencies to see if they make significant group difference (C/N).
  - ☐ Probabilistic distance metric to measure the data similarity with and without testing miRNAs and their dependencies (e.g., using Entropic Gromov Wasserstein)



MicroRNA and Gene Interaction Networks

- Gaussian Graphical Model (GGN)s were constructed to represent dependency networks.
- Bayesian Networks (BNs) were constructed to represent microRNA-gene binding relationship.
    - Evidence matrix is from starBase

      (e.g., 166,669 microRNA-gene interaction)

	G_1	G_2	. . .	G_n
m_1	1/0			
m_2				
.				
m_p				

    - Greedy Hill Climbing (GHC)+Tabu search based on Bayesian Dirichlet equivalence uniform (BDeu) scoring
    - Intervention to a model is a lot easier

$$\boldsymbol{X} = [X_1, X_2, \ldots, X_p]$$

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad x_1, x_2, \ldots, x_n$$

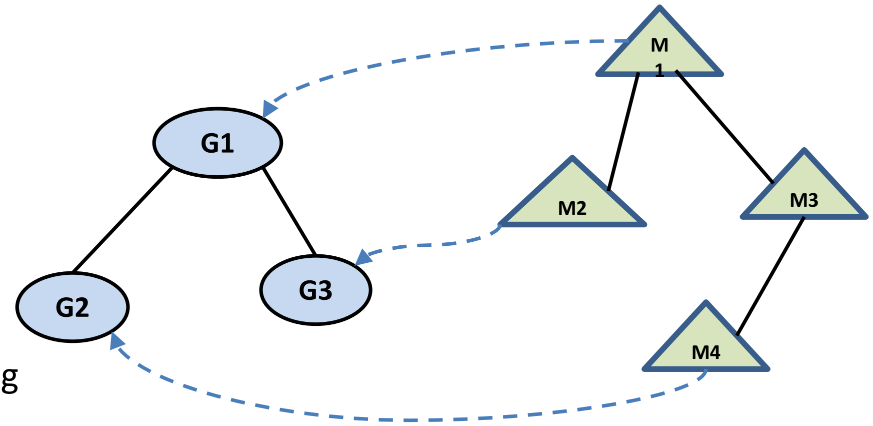
$$p(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

Precision matrix: $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$

Optimization function of fused graphical lasso:

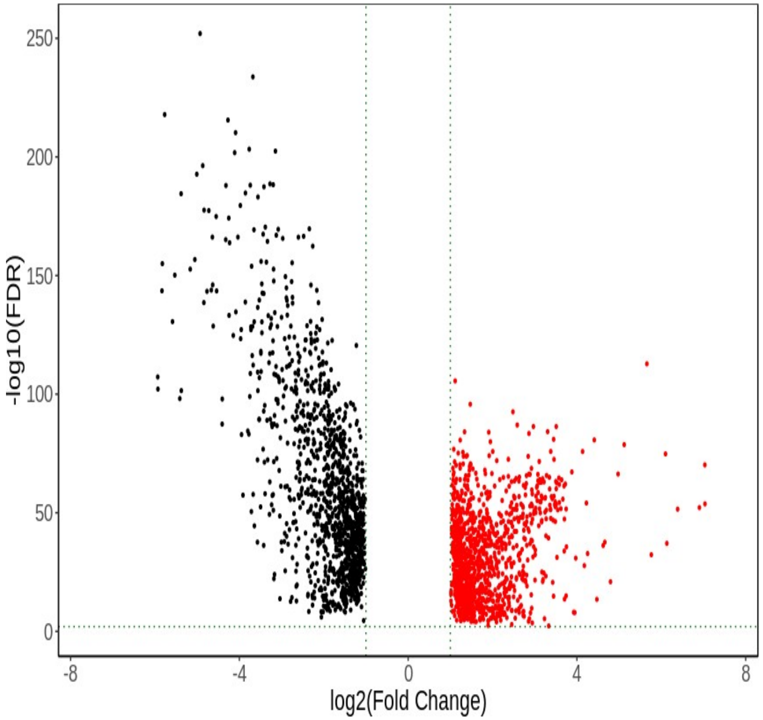
$$\min_{\{\boldsymbol{\Omega}>0\}} \mathcal{L}(\{\boldsymbol{\Omega}\}) := \textstyle\sum_{k=1}^{K} \operatorname{tr}(\widehat{\boldsymbol{\Sigma}}^{(k)}\boldsymbol{\Omega}^{(k)}) - \log\det \boldsymbol{\Omega}^{(k)} + P(\{\boldsymbol{\Omega}\})$$

where $$P(\{\boldsymbol{\Omega}\}) = \lambda_1 \sum_{k=1}^{K} \sum_{i\neq j} |\omega_{ij}^{(k)}| + \lambda_2 \sum_{i\neq j} \sqrt{\sum_{k=1}^{K} {\omega_{ij}^{(k)}}^2}$$



Data

- Data: TCGA-BRC
- Samples:
  - Normal: 104
  - Stage 1: 179
  - Stage 2: 608
  - Stage 3: 242
  - Stage 4: 20
- DEGs (fold-change > 2):
  - Up: 1218; Down: 1236
- 617 microRNAs
  - 1,678 binding relationships (DAG)
  - 3,137 interactions in undirected graph make significant group difference (C/N)

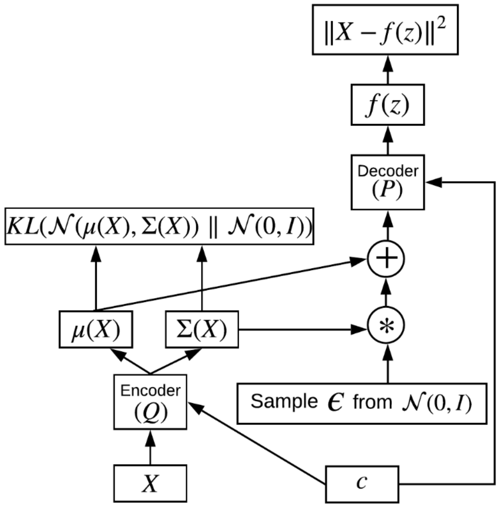


Dogan and Cui, manuscript under review at Bioinformatics

Data Augmentation

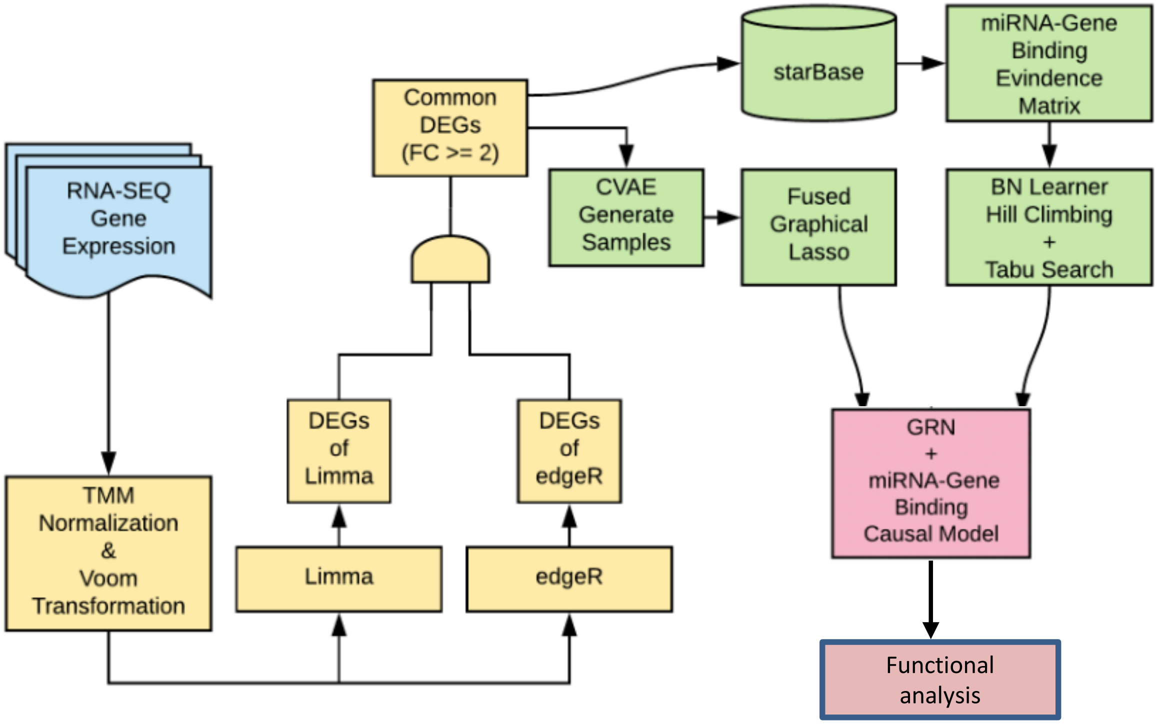
We used deep learning to generate more in-vitro samples:

- number of samples in groups is imbalanced
- unbiased estimator for a learner



The Conditional Variational Autoencoder (CVAE) architecture

The Workflow



RNA-SEQ Gene Expression

TMM Normalization & Voom Transformation

Limma

edgeR

DEGs of Limma

DEGs of edgeR

Common DEGs (FC >= 2)

starBase

miRNA-Gene Binding Evindence Matrix

CVAE Generate Samples

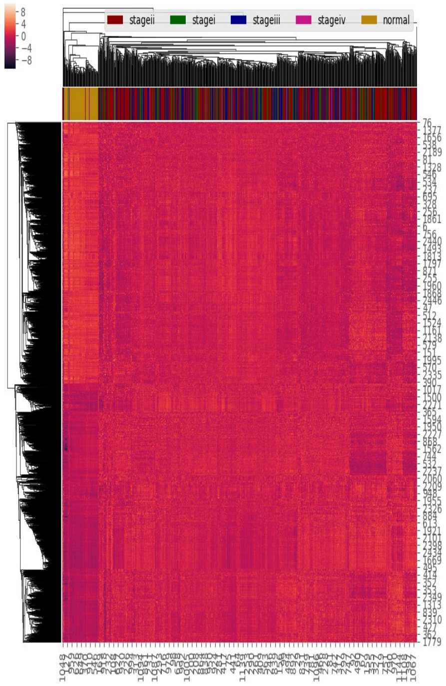
Fused Graphical Lasso

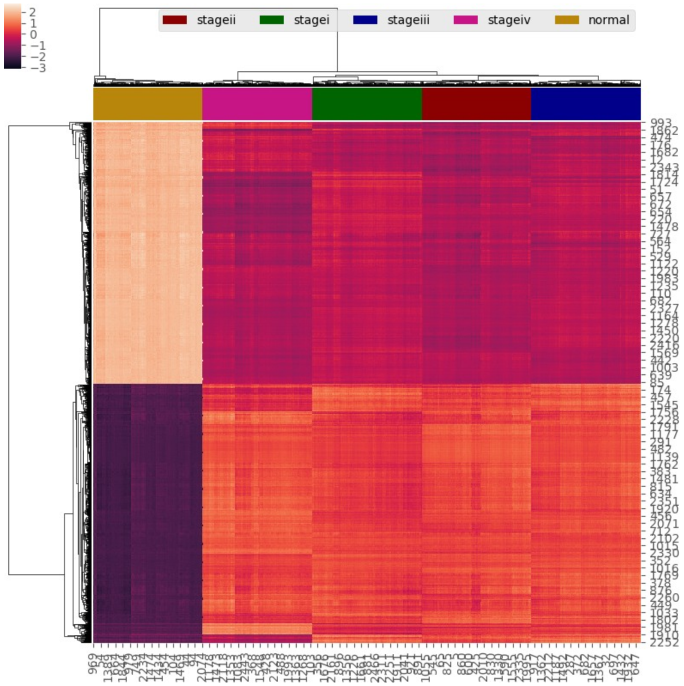
BN Learner Hill Climbing + Tabu Search

GRN + miRNA-Gene Binding Causal Model

Functional analysis

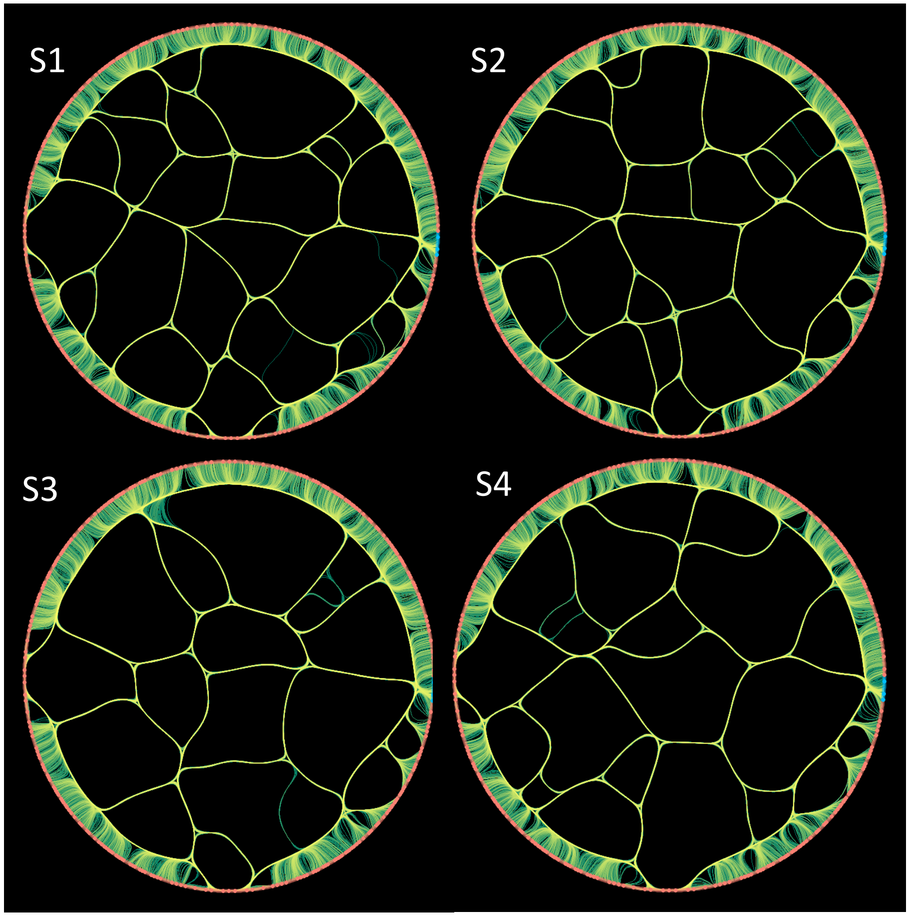
After Data Augmentation



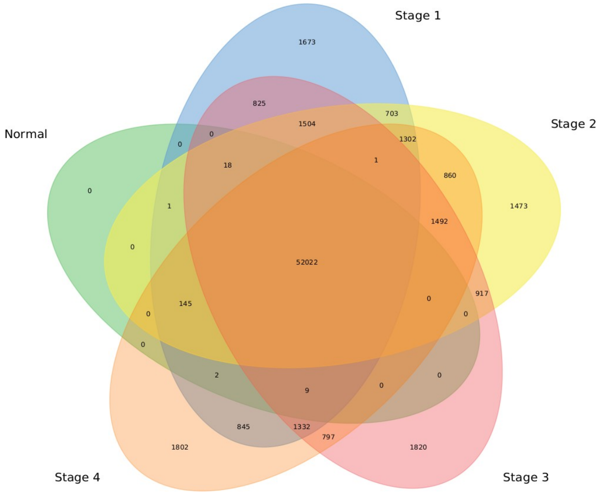


Dogan and Cui, manuscript under preparation

Results: Conditional Networks during Progression

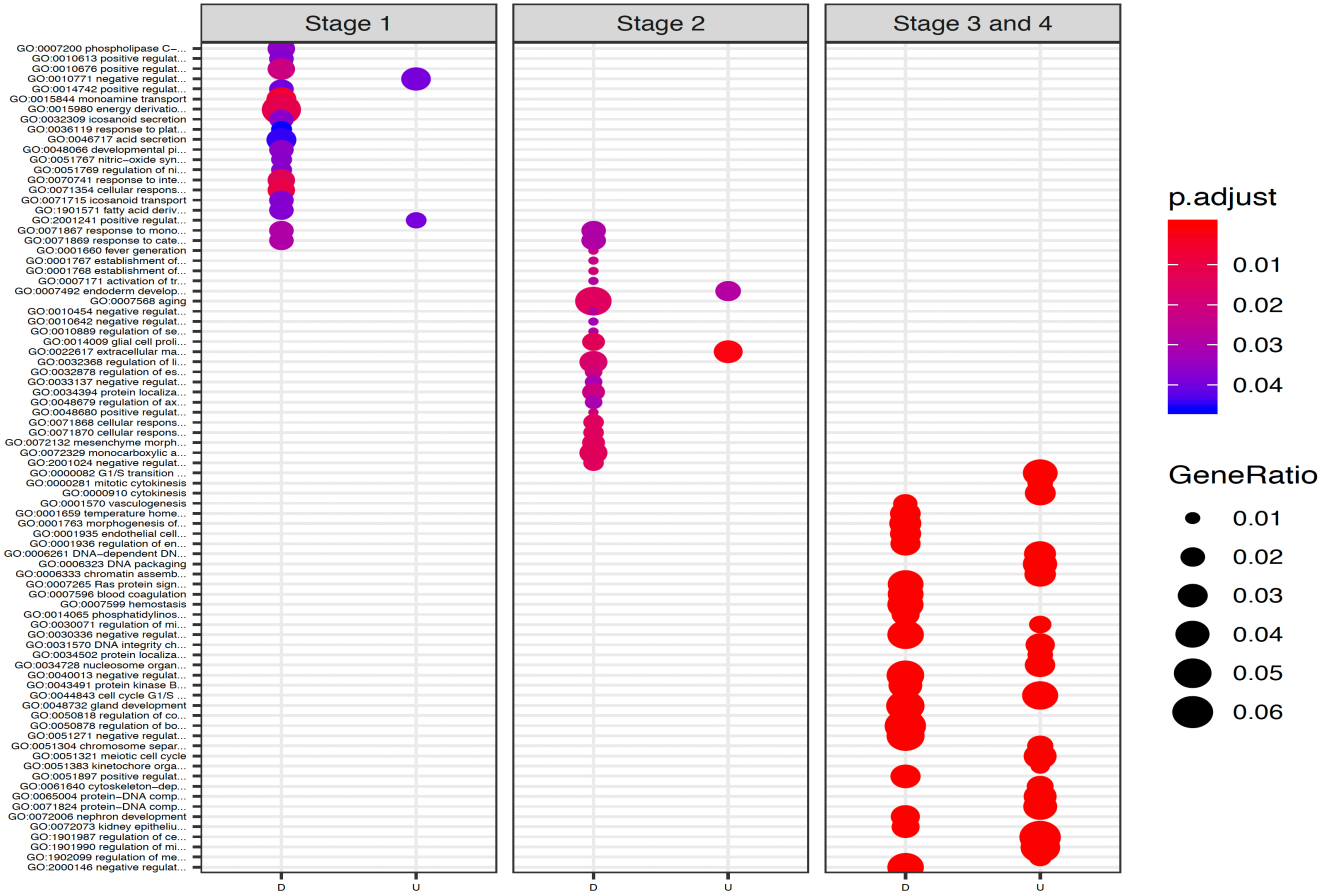


Red dots: gene
Blue dots: microRNA

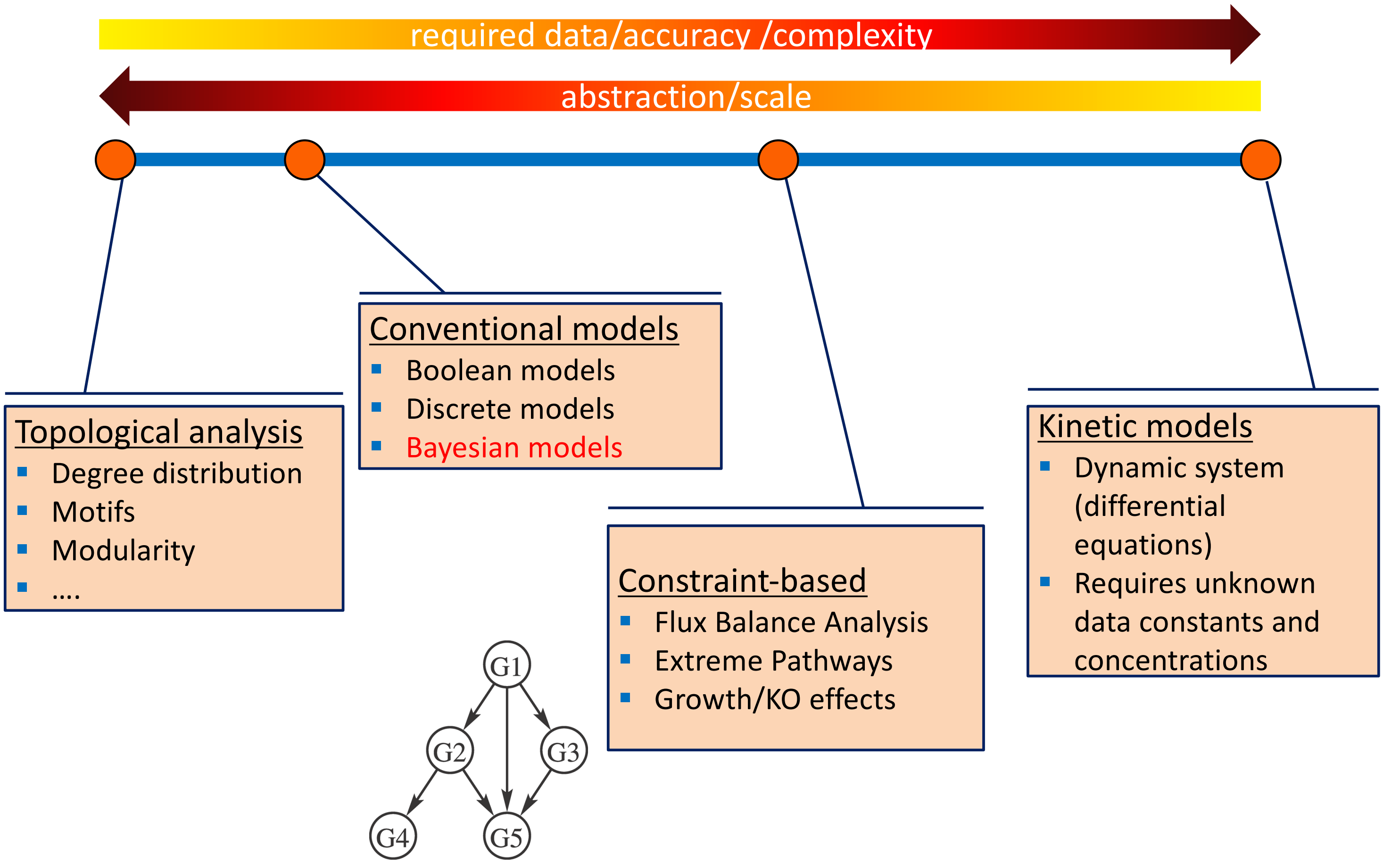


Dogan and Cui, manuscript under preparation

Results: Functional Transition During Progression



Network Models



required data/accuracy /complexity

abstraction/scale

Topological analysis
- Degree distribution
- Motifs
- Modularity
- ....

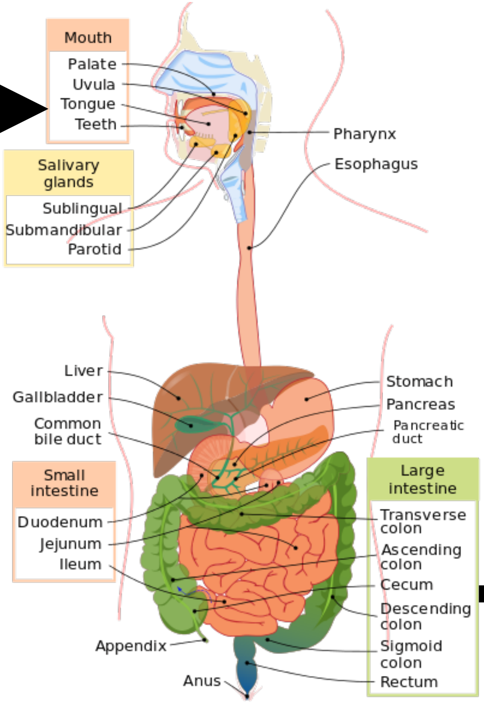
Conventional models
- Boolean models
- Discrete models
- Bayesian models

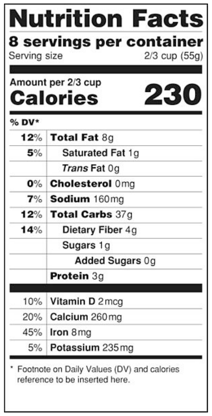
Constraint-based
- Flux Balance Analysis
- Extreme Pathways
- Growth/KO effects

Kinetic models
- Dynamic system (differential equations)
- Requires unknown data constants and concentrations

Stochastic Modeling of Glucose Metabolism

Food Intake





Nutrition Facts

8 servings per container

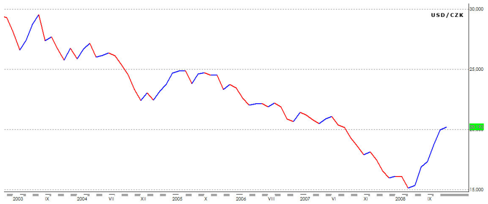
Serving size 2/3 cup (55g)

Amount per 2/3 cup

Calories 230

% DV*	
12%	**Total Fat** 8g
5%	**Saturated Fat** 1g
	Trans **Fat** 0g
0%	**Cholesterol** 0mg
7%	**Sodium** 160mg
12%	**Total Carbs** 37g
14%	**Dietary Fiber** 4g
	Sugars 1g
	Added Sugars 0g
	Protein 3g
10%	**Vitamin D** 2mcg
20%	**Calcium** 260mg
45%	**Iron** 8mg
5%	**Potassium** 235mg

* Footnote on Daily Values (DV) and calories reference to be inserted here.



Cellular Glucose, ATP

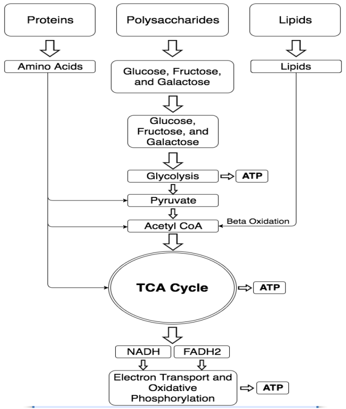
- As per the CDC there are 30.3 Million (9.4% of US population) cases of diabetes
- Diabetes is a disease where the blood glucose reaches abnormal levels
- Insulin plays a key role in the regulation of glucose uptake from blood by cells

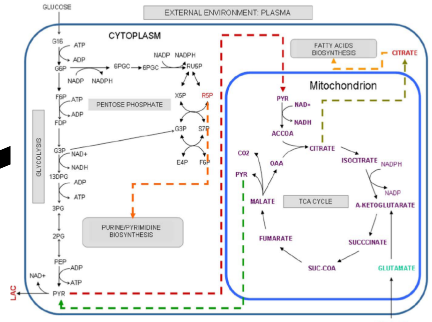
Model

Reaction Rates and Constants

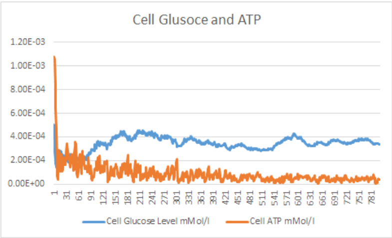
$$A+B \underset{k}{\rightleftharpoons} C+D$$

Initial Concentrations





Abbreviation	Glycolytic reactions	Genes
HEX1	[c]GLC + ATP ⟶ G6P + ADP + H	*glk*
PGI	[c]G6P ⟷ F6P	*pgi*
PFK	[c]ATP + F6P ⟶ ADP + FDP + H	*pfkA, pfkB*
FBA	[c]FDP ⟷ DHAP + G3P	*fbaA, fbaB*
TPI	[c]DHAP ⟷ G3P	*tpiA*
GAPD	[c]G3P + NAD + PI ⟷ 13DPG + H + NADH	*gapA, gapC1, gapC2*
PGK	[c]13DPG + ADP ⟷ 3PG + ATP	*pgk*
PGM	[c]3PG ⟷ 2PG	*gpmA, gpmB*
ENO	[c]2PG ⟷ H_2O + PEP	*eno*
PYK	[c]ADP + H + PEP ⟶ ATP + PYR	*pykA, pykF*



Step-Wise Concentrations

An enzymatic reaction

$$E+S \underset{k_{-1}}{\overset{k_1}{\longleftrightarrow}} ES \xrightarrow{k_2} E+P$$

The ODEs of this reaction:

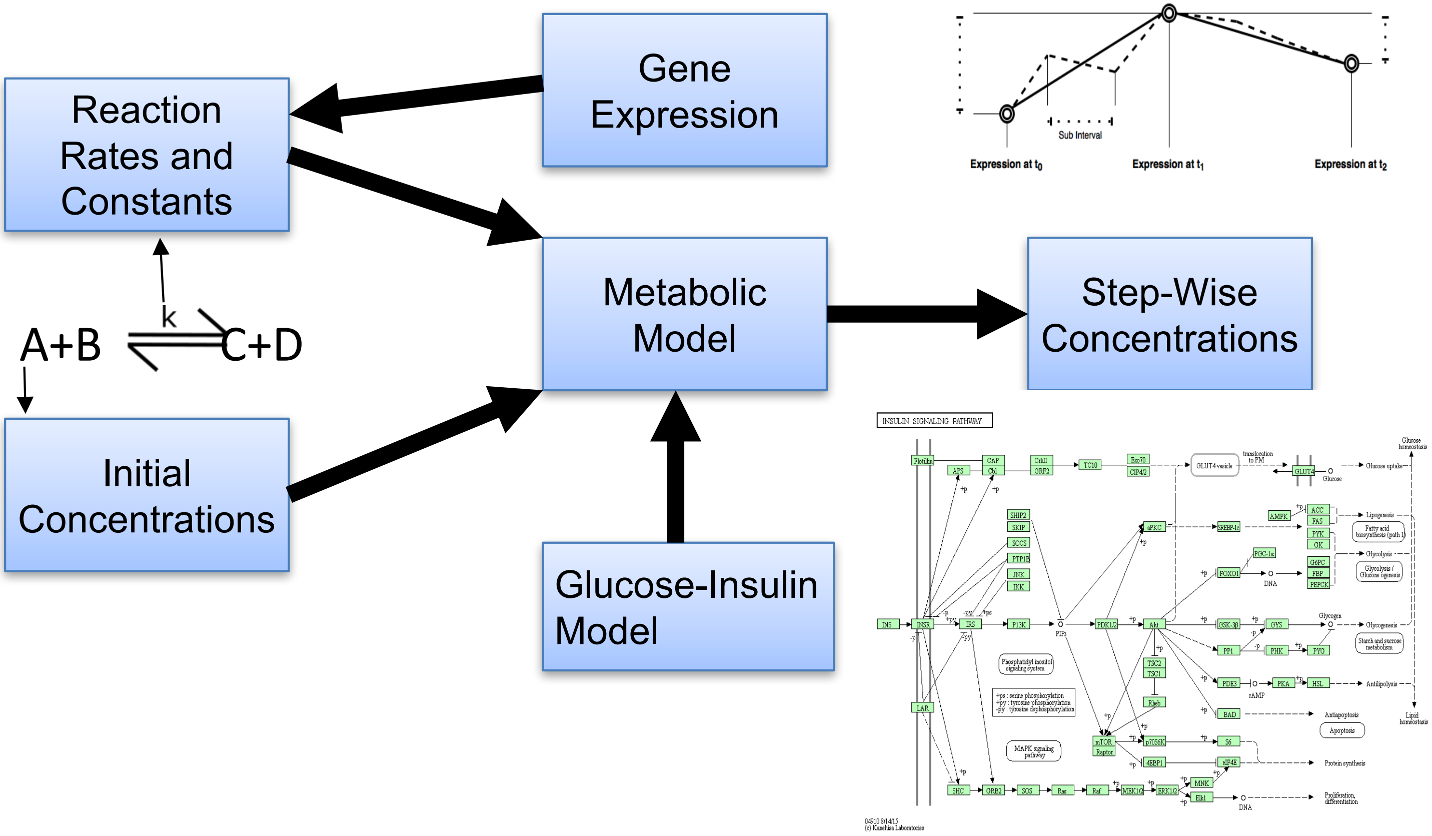
$$\frac{dS}{dt} = -k_1 E \bullet S + k_{-1} ES$$

$$\frac{dES}{dt} = k_1 E \bullet S - (k_{-1} + k_2) ES$$

$$\frac{dE}{dt} = -k_1 E \bullet S + (k_{-1} + k_2) ES$$

$$\frac{dP}{dt} = k_2 ES$$

Model



Gene Expression

Reaction Rates and Constants

Metabolic Model

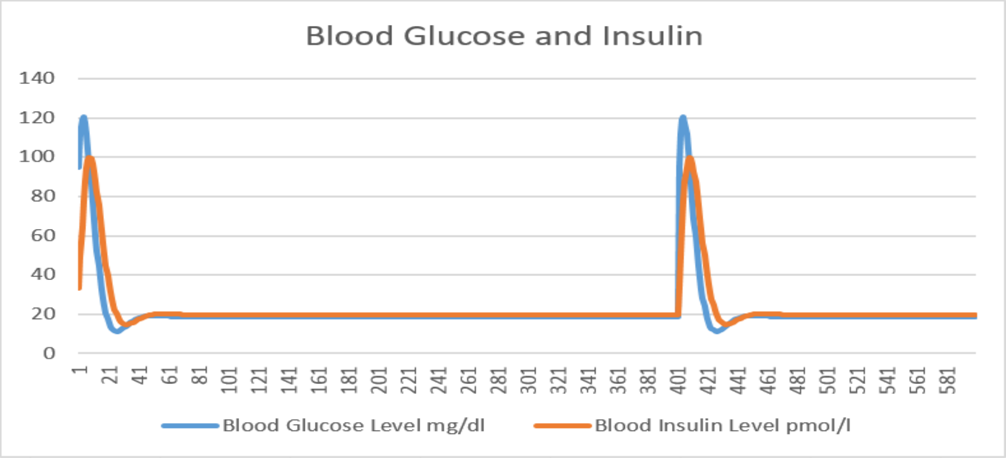
Step-Wise Concentrations

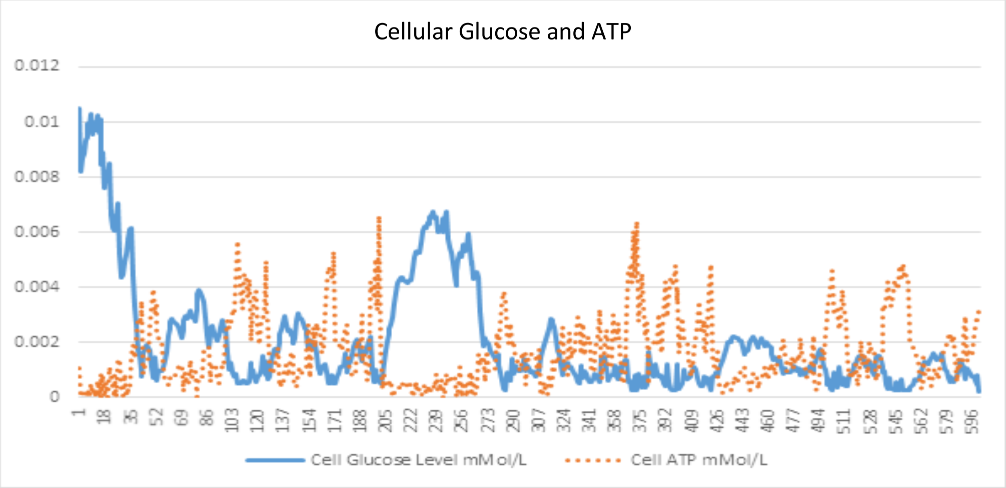
Initial Concentrations

Glucose-Insulin Model

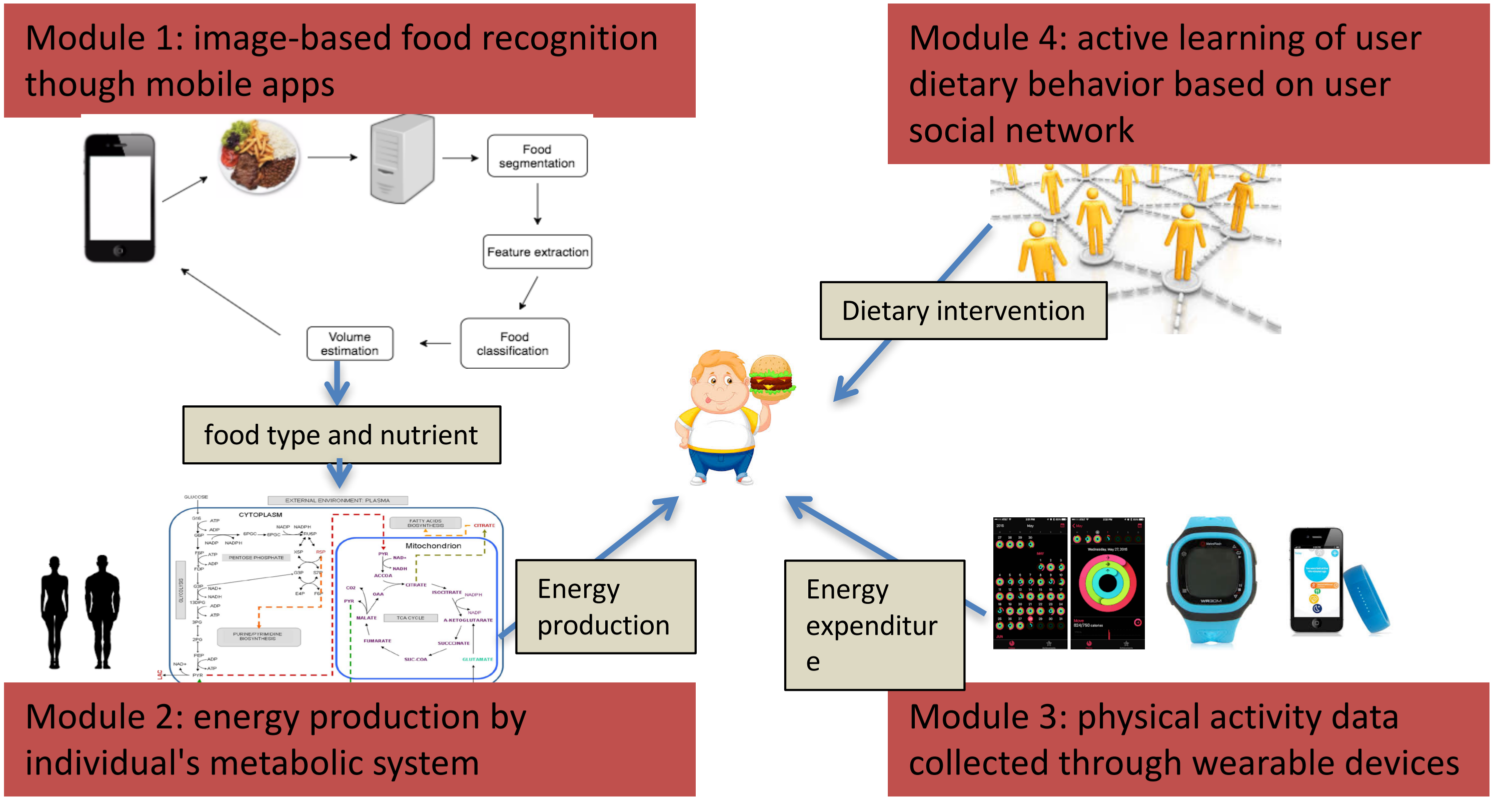
$A+B \overset{k}{\rightleftharpoons} C+D$

Results: Glucose, ATP, and Insulin





A Smart Health System



Module 1: image-based food recognition though mobile apps

food type and nutrient

Module 2: energy production by individual's metabolic system

Energy production

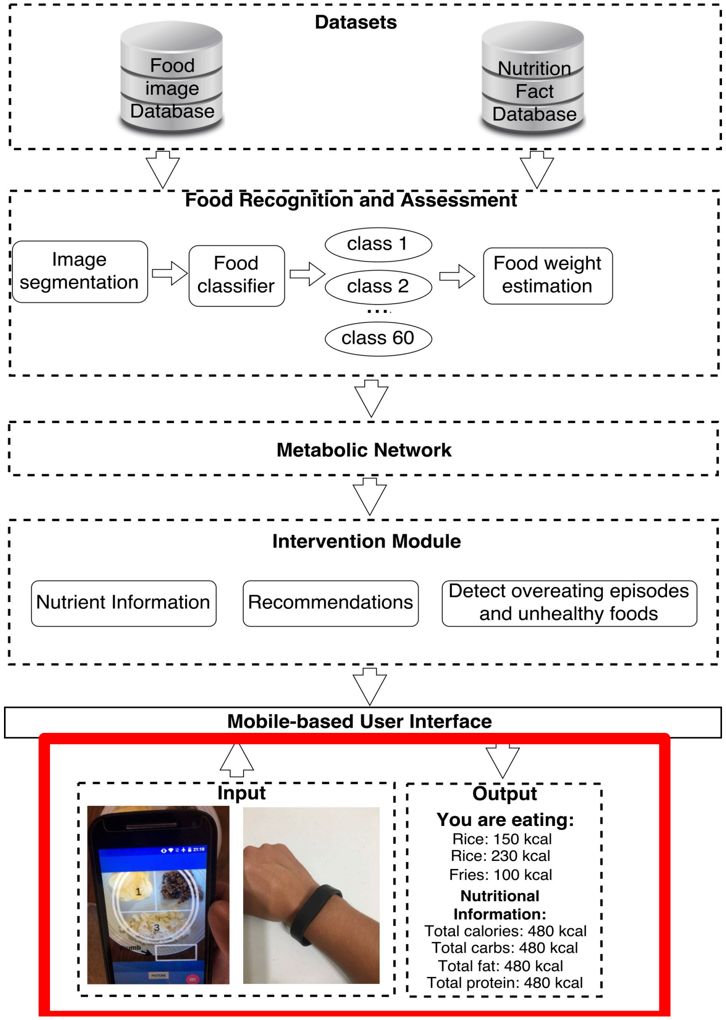
Module 4: active learning of user dietary behavior based on user social network

Dietary intervention

Energy expenditure

Module 3: physical activity data collected through wearable devices

Image-Based Food Recognition



Datasets

Food image Database

Nutrition Fact Database

Food Recognition and Assessment

Image segmentation

Food classifier

class 1

class 2

…

class 60

Food weight estimation

Metabolic Network

Intervention Module

Nutrient Information

Recommendations

Detect overeating episodes and unhealthy foods

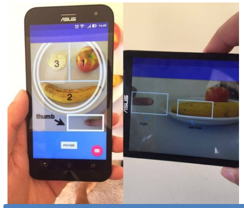
Mobile-based User Interface

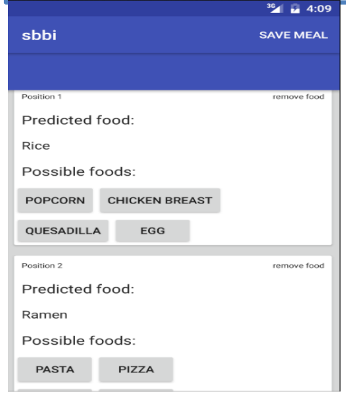
Input

Output

You are eating:
Rice: 150 kcal
Rice: 230 kcal
Fries: 100 kcal

Nutritional Information:
Total calories: 480 kcal
Total carbs: 480 kcal
Total fat: 480 kcal
Total protein: 480 kcal





Position 1 remove food

Predicted food:

Rice

Possible foods:

Position 2 remove food

Predicted food:

Ramen

Possible foods:

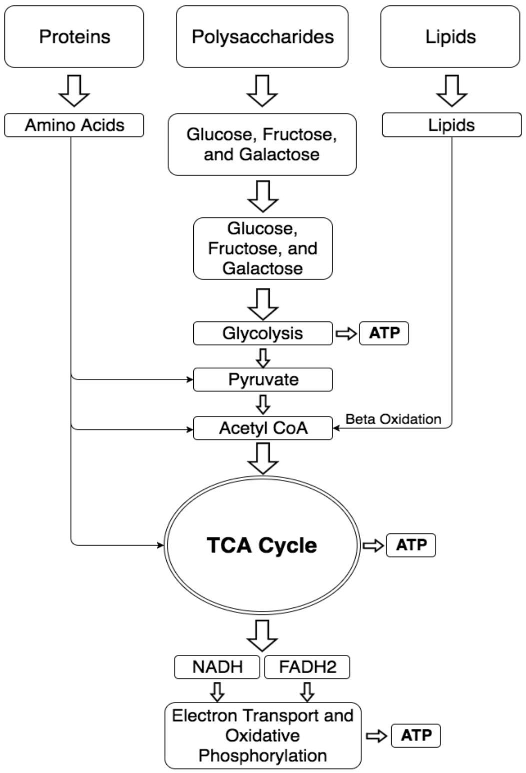
Silva, et al. , J. of Health and Medical Informatics, 2018

Estimation versus Ground Truth

Illustration of the estimated weight (est.) versus the ground truth (gtr.) based on the prediction of three sample meals

Meal	**Food item**	**Weight (in grams)**		**Calorie (in kcal)**		**Protein (in grams)**		**Lipid (in grams)**		**Carbohydrate (in grams)**		**Fiber (in grams)**		**Sugar (in grams)**	
		est	grt	**est**	grt	**est**	grt	**est**	grt	**est**	grt	**est**	grt	**est**	grt
Breakfast	Banana	**89.0**	95.0	**79.0**	85.0	**0.97**	1.04	**0.29**	0.31	**20.33**	21.70	**2.30**	2.5	**10.88**	11.62
	Apple	**134.0**	140.0	**70.0**	73.0	**0.34**	0.36	**0.23**	0.24	**18.50**	19.33	**3.25**	3.4	**13.92**	14.55
	Cookies	**52.0**	50.0	**254.80**	245.0	**2.96**	2.85	**11.44**	11.0	**35.15**	33.80	**0.83**	0.8	**17.10**	16.45
Lunch	Rice	**190.0**	180.0	**247.0**	234.0	**4.52**	4.28	**0.40**	0.38	**54.31**	51.46	**0.53**	0.5	**0.0**	0.0
	Ramen	**130.0**	150.0	**572.0**	660.0	**13.22**	15.25	**22.87**	26.38	**78.34**	90.39	**3.8**	4.3	**2.57**	2.97
	French fries	**46.0**	60.0	**133.0**	174.0	**1.52**	1.99	**6.05**	7.9	**18.17**	23.7	**1.76**	2.3	**0.0**	0.0
Dinner	Rice	**246.0**	240.0	**319.8**	312.0	**5.85**	5.71	**0.51**	0.5	**70.33**	68.62	**0.71**	0.7	**0.0**	0.0
	Chicken brest	**60.0**	55.0	**90.54**	83.0	**18.32**	16.8	**1.89**	1.74	**0.0**	0.0	**0.0**	0.0	**0.0**	0.0
	French fries	**80.0**	100.0	**232.0**	290.0	**2.65**	3.31	**10.53**	13.17	**31.6**	39.5	**3.12**	3.9	**0.0**	0.0
TOTAL		**1027.0**	1070.0	**1998.14**	2156.0	**50.35**	51.59	**54.21**	61.62	**326.73**	348.5	**16.3**	18.4	**44.47**	45.59

Simulation of Energy Production



Proteins

Polysaccharides

Lipids

Amino Acids

Glucose, Fructose, and Galactose

Lipids

Glucose, Fructose, and Galactose

Glycolysis ⇨ **ATP**

Pyruvate

Acetyl CoA

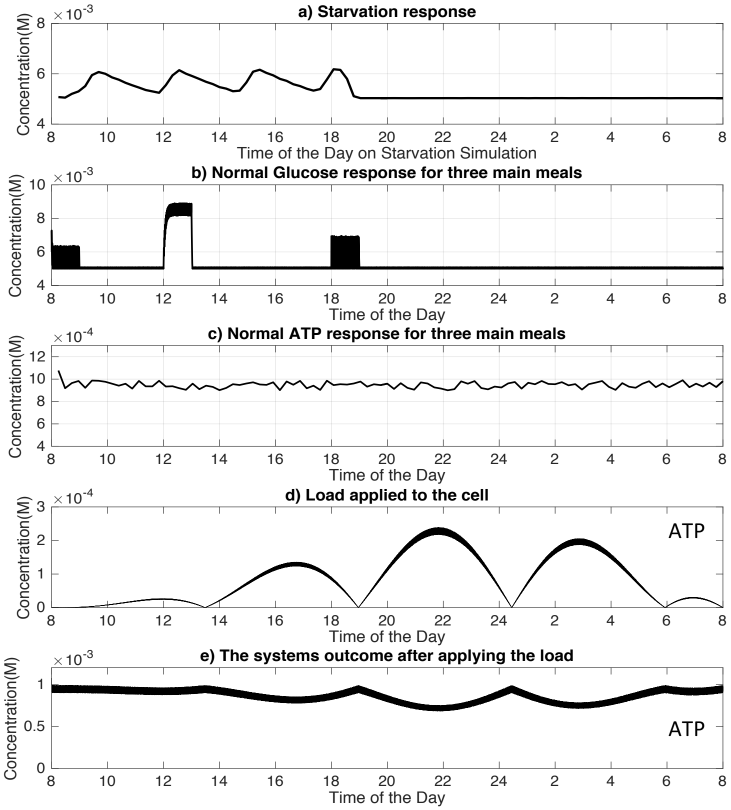
Beta Oxidation

TCA Cycle ⇨ **ATP**

NADH

FADH2

Electron Transport and Oxidative Phosphorylation ⇨ **ATP**



Silva, et al. , J. of Health and Medical Informatics, 2018

Summary

- Integration of multi-omics information from various sources is key to the understanding of important but complex processes of dynamic gene regulation.

- Various learning models facilitate the mechanistic discoveries in complex human diseases, before a more complete stochastic or kinetics model is available; or may hopefully bypass it.

- New generation of interpretable deep learning frameworks by coupling the structure of the neural network with the internal workings of cell is desired.

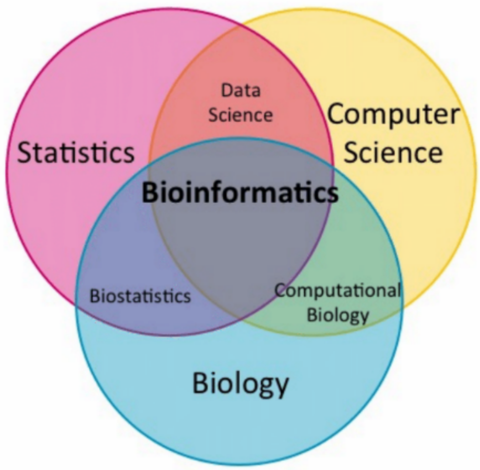
A Widely Accepted Saying

- What computational science to molecular biology will be like what mathematics has been to physics in the past 400 years

Bioinformatics Programs and Courses at CSE

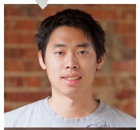
- **Computational Biology and Bioinformatics (CBB) minor**
- **PhD/MS in CS with Bioinformatics specialization**

- CSCE496/896 Computational Methods in Bioinformatics (Renamed to CSCE 471/871 Introduction to Bioinformatics)
  - A general introduction to the field of bioinformatics
  - A way of thinking -- tackling “biological problem” computationally
  - Some exposure to computational biology and bioinformatics research, covering multiple aspects of computational genomics, proteomics and systems biology
- CSCE971 Advanced Bioinformatics
  - Fundamental machine learning and state-of-the-art deep learning
  - Probabilistic modeling
- CSCE155T Programming in Python
- CSCE311 Data Structures and Algorithms for Informatics



Acknowledgements

Project team members:

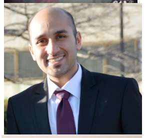


Kevin Chiang

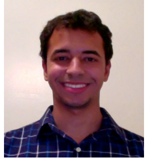
Jiang Shu

Hanyuan Zhang

Sonal Sukreet



Milad Ghiasi Rad



Bruno Vieira

Collaborators:

- UNL

Janos Zempleni, Nutrition
Jiri Adamec, Biochemistry
Stephen Scott, Computer Science
Mengna Xia, BORC @ NPOD
Zheng Xu, Statistics

- UNMC

Guoku, Hu, Pharmacology

NIH COBRE 1P20GM104320
NIH 1R01DK107264/ USDA

UNL Food for Health Seed Grant
UNL Layman grant