Information theoretic sub-network mining characterizes breast cancer subtypes in terms of cancer core mechanisms

Jinwoo Park, Benjamin Hur, Sungmin Rhee, Sangsoo Lim, Min-Su Kim, Kwangsoo Kim, Wonshik Han and Sun Kim

- Jinwoo Park(jwpark.bioinfo@snu.ac.kr)
- Bio and Health Informatics lab
- Seoul National University

Contents

Introduction

Breast cancer subtype

"Breast cancer subtype is widely used for clinical application"

- Breast cancer
  - Have various scenarios in tumor development
  - Challenging to diagnose the actual risk factors
  - PAM50 subtypes are widely used for making the clinical decision

Introduction

Mechanism of cancer

"Mechanism of cancer is characterized as cancer hallmarks"

- Cancer Hallmarks
  - Set of building blocks that represents the characteristics of the tumor
  - Usually explained and summarized in terms of biological pathways
    - Hallmark such as "resisting cell death" and "sustaining proliferative signaling" can be represented by *cell cycle*, *DNA replication*, and *p53 signaling* pathways.

Cancer hallmarks
Resisting cell death
Sustaining proliferative signaling
Evading growth suppressors
Activating invasion and metatstasis
Enabling replicative immortality
Inducing angiogenesis

Contents

Motivation

Breast cancer subtype and cancer hallmarks

"How breast cancer subtypes can be explained in terms of cancer hallmarks?"

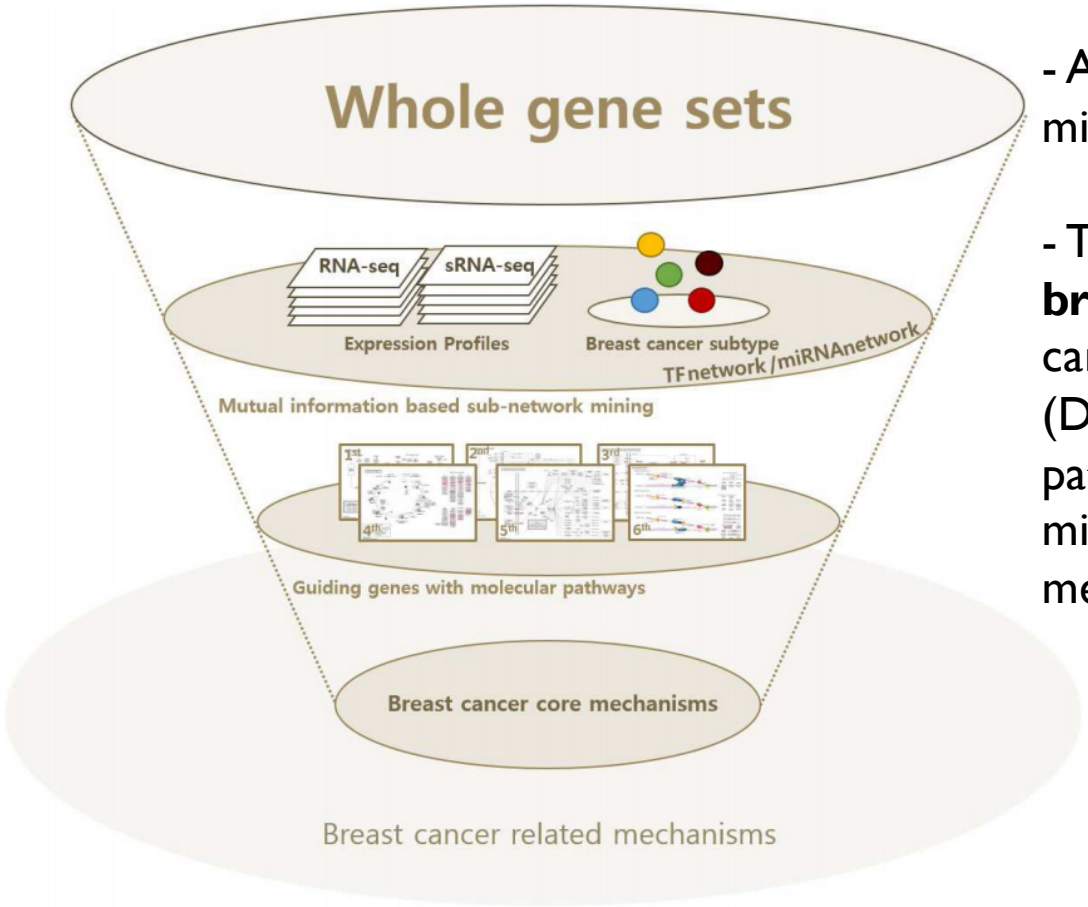
- Pam50 genes
  - Selected only by the statistical significance in terms of clinical outcomes
  - Do not explain complex relationships among genes, especially in terms of biological pathways

- Researches: Pam50 subtype ←→ Cancer Hallmarks
  - TCGA
  - Qin et al.
  - Lim et al.

Motivation

Our approach

"Our information theoretic, sub-network mining approach classify subtypes and core mechanisms of breast cancer"



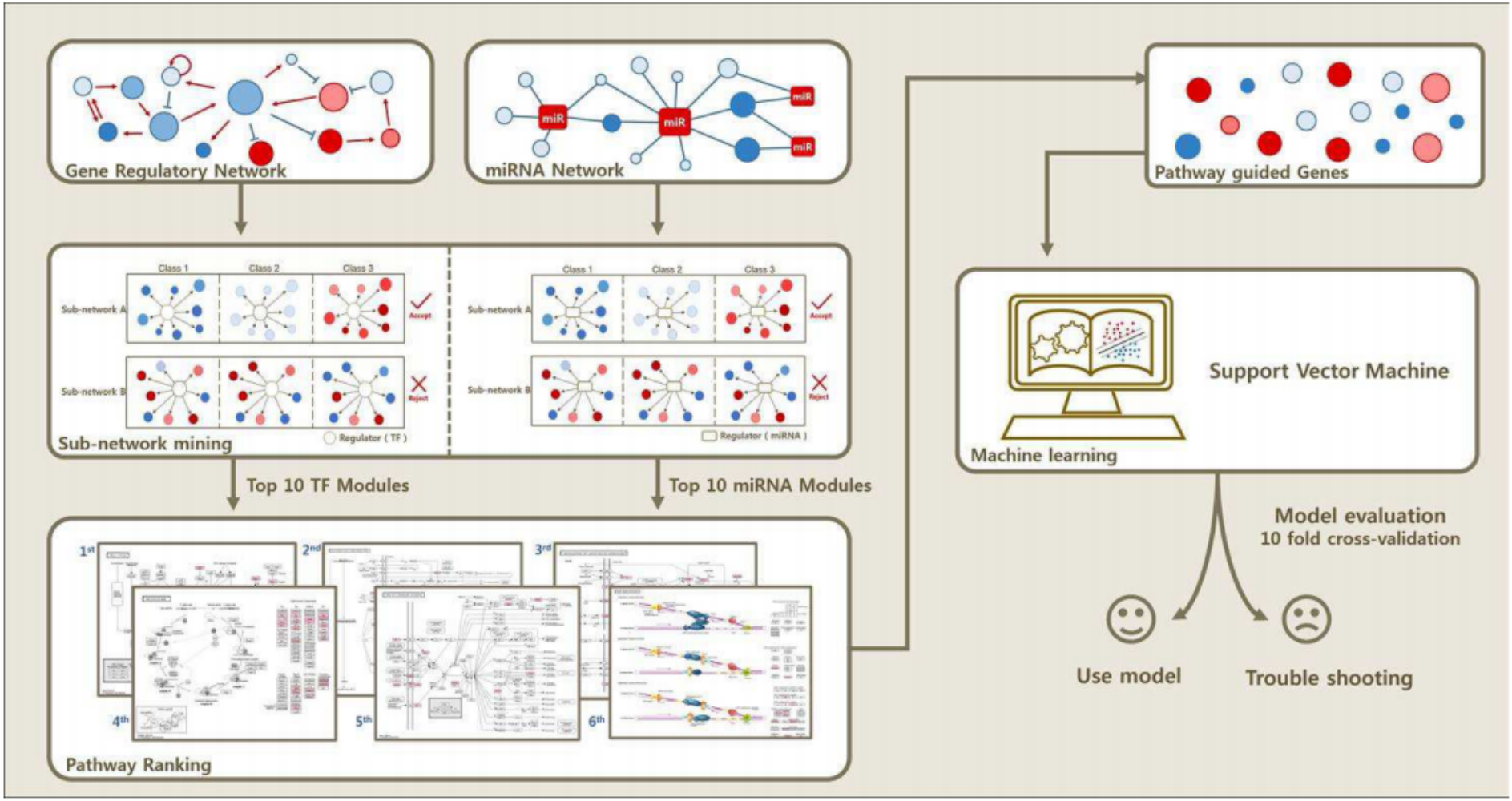
- An information theoretic sub-network mining algorithm

- To characterize **differences among breast cancer subtypes** in terms of cancer hallmarks, or **core mechanisms** (DNA replication, cell cycle, and p53 pathway) and identify **regulators** (TF and miRNA) that potentially control the core mechanism.

Contents

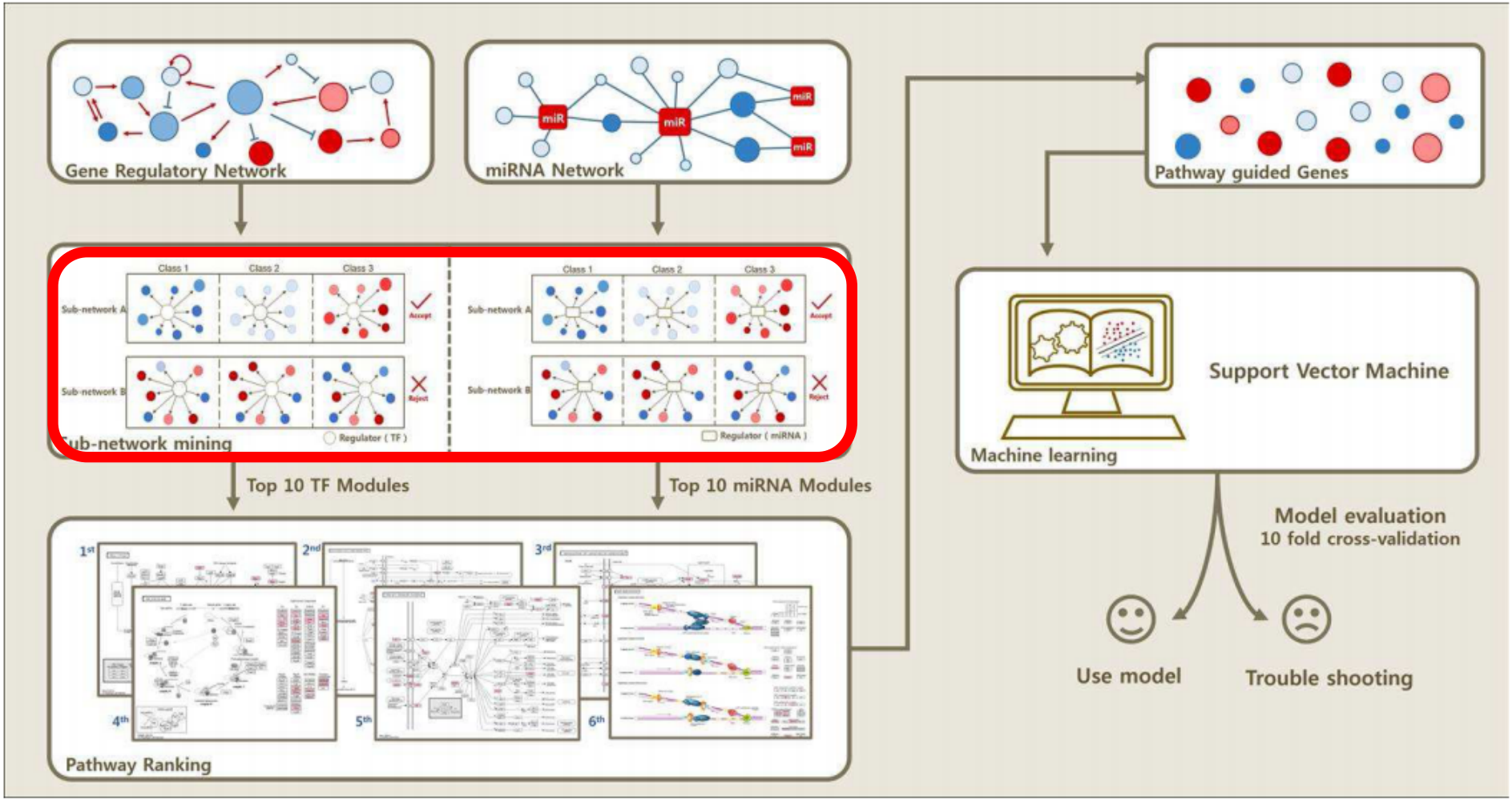
Methods

Overview



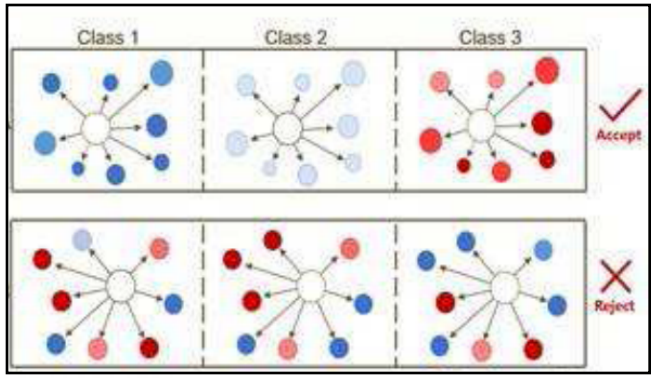
Methods

Overview



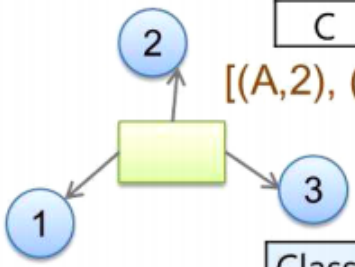
Methods

Entropy based score of a regulator-module



Class	Exprs
A	2
B	5
C	7

[(A,2), (B,5), (C,7)]



Class	Exprs
A	1
B	4
C	6

[(A,1), (B,4), (C,6)]

Class	Exprs
A	3
B	8
C	9

[(A,3), (B,8), (C,9)]

S_1 (A,1) (A,2) (A,3)

S_2 (B,4) (B,5) (C,6)

S_3 (C,7) (B,8) (C,9)

Sorted by expression

$$H(S_1) = -1\log(1) = 0$$

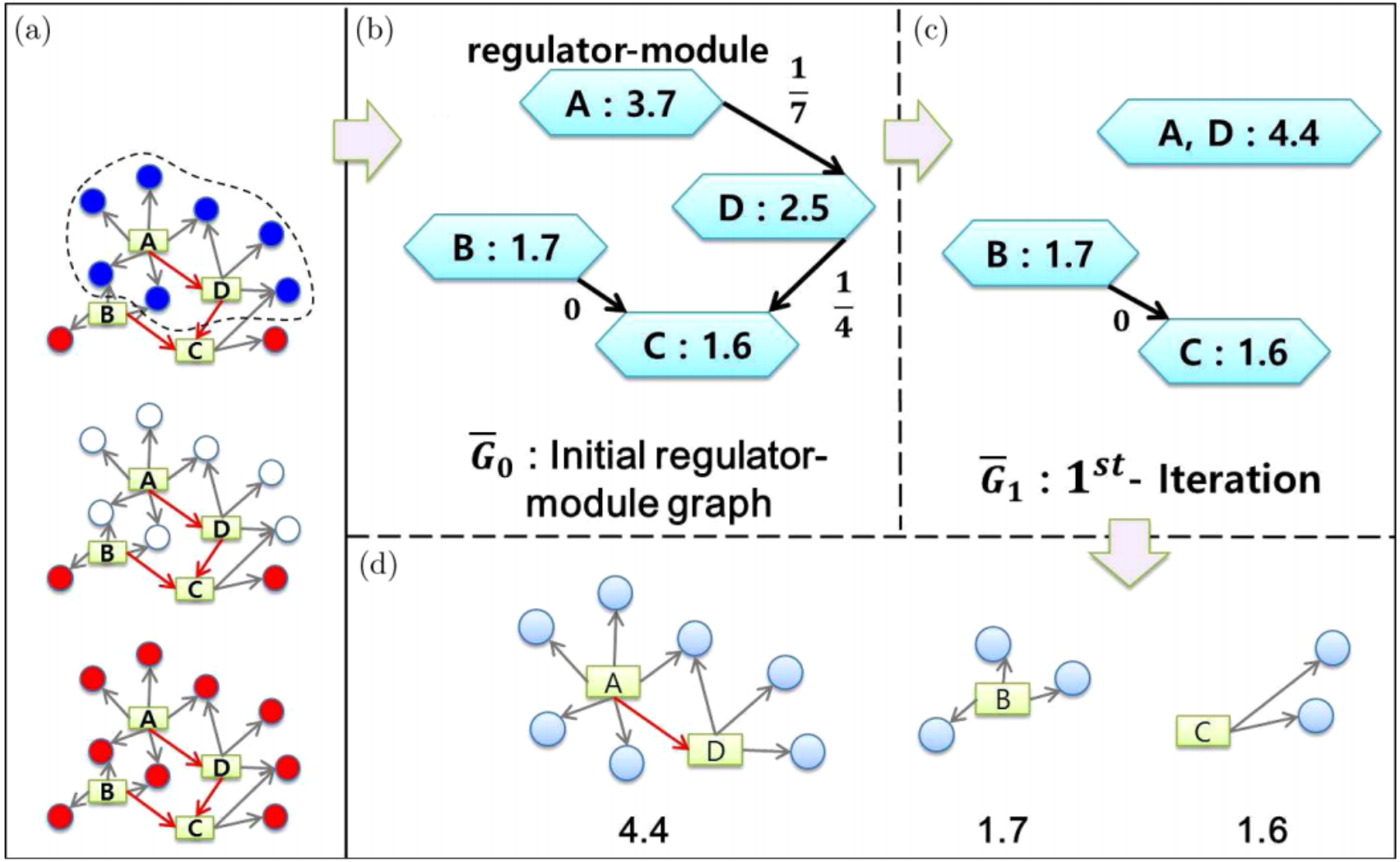
$$H(S_2) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.92$$

$$H(S_2) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0.92$$

$$H(L) = \left(\frac{3*0 + 3*0.92 + 3*0.92}{3+3+3}\right) = 0.61$$

Methods

Sub-network mining



Methods

Pathway Prioritization

- With DeSPA, 555 genes were selected
  - From top 10 TF-modules and top 10 miRNA-modules
- The genes are then mapped to KEGG pathway to rank the pathways
- Ranks of the pathways were calculated
  - Based on the mapping ratio of the mapped genes to the total genes in the pathway.

Contents

Result and Discussion

What we discovered

- (1) Top 50 genes from DeSPA achieved comparable classification performance to the PAM50 genes although only four genes are common
- (2) DeSPA was able to rank pathways that are highly related to the core mechanisms of breast cancer by mapping genes to KEGG pathway
- (3) As DeSPA considered regulatory network for its input, regulators (TFs and miRNAs) of the core gene set could be found

Result and Discussion

10-Fold Cross-validation

"Top 50 genes from DeSPA achieved comparable classification performance to the PAM50 genes although only four genes are common"

- Purpose
  - To evaluate the explanatory power of DeSPA
  - Compare top 50 genes from DeSPA results VS Pam50 genes
- Samples & Configuration
  - From TCGA-BRCA, we extracted the PAM50 subtype label and genome-wide expression levels
  - SVM model implemented with the SMO algorithm in the WEKA
- Results
  - DeSPA(73.58%) v.s. PAM50(73.88%)

Result and Discussion

Prioritized pathways

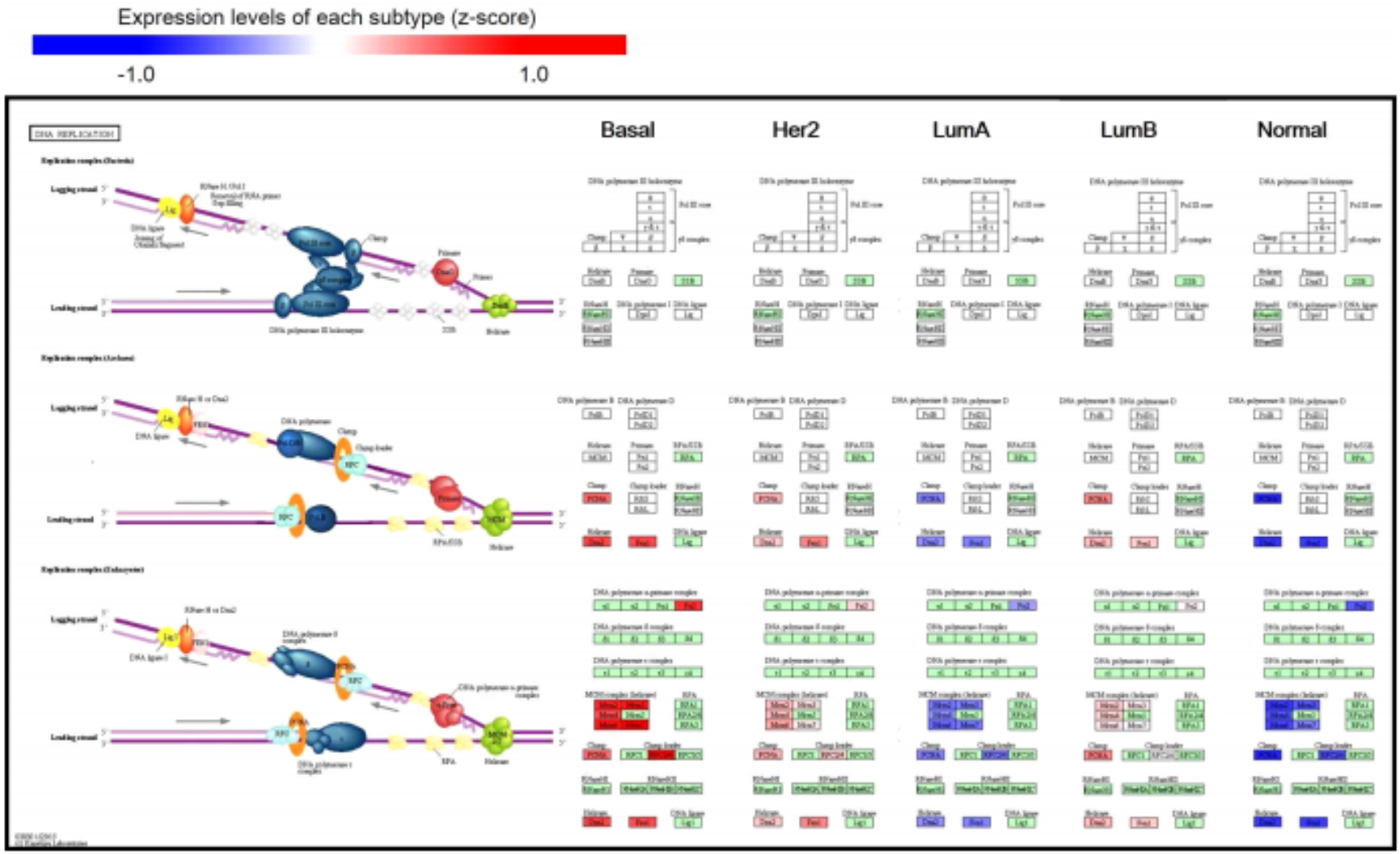
"DeSPA was able to rank pathways that are highly related to the core mechanisms of breast cancer by mapping genes to KEGG pathway"

Pathway rank	Pathway name	Mapping ratio
1	DNA replication	0.39 (14/36)
2	Cell cycle	0.27 (33/124)
3	Circadian rhythm	0.23 (7/30)
4	Oocyte meiosis	0.17 (19/110)
5	p53 signaling pathway	0.16 (11/68)
6	Progesterone-mediated oocyte maturation	0.15 (13/86)
7	Fanconi anemia pathway	0.15 (13/86)
8	Vasopressin-regulated water reabsorption	0.11 (5/45)
9	Small cell lung cancer	0.08 (7/86)
10	Gap junction	0.08 (7/89)

Result and Discussion

DNA replication pathway

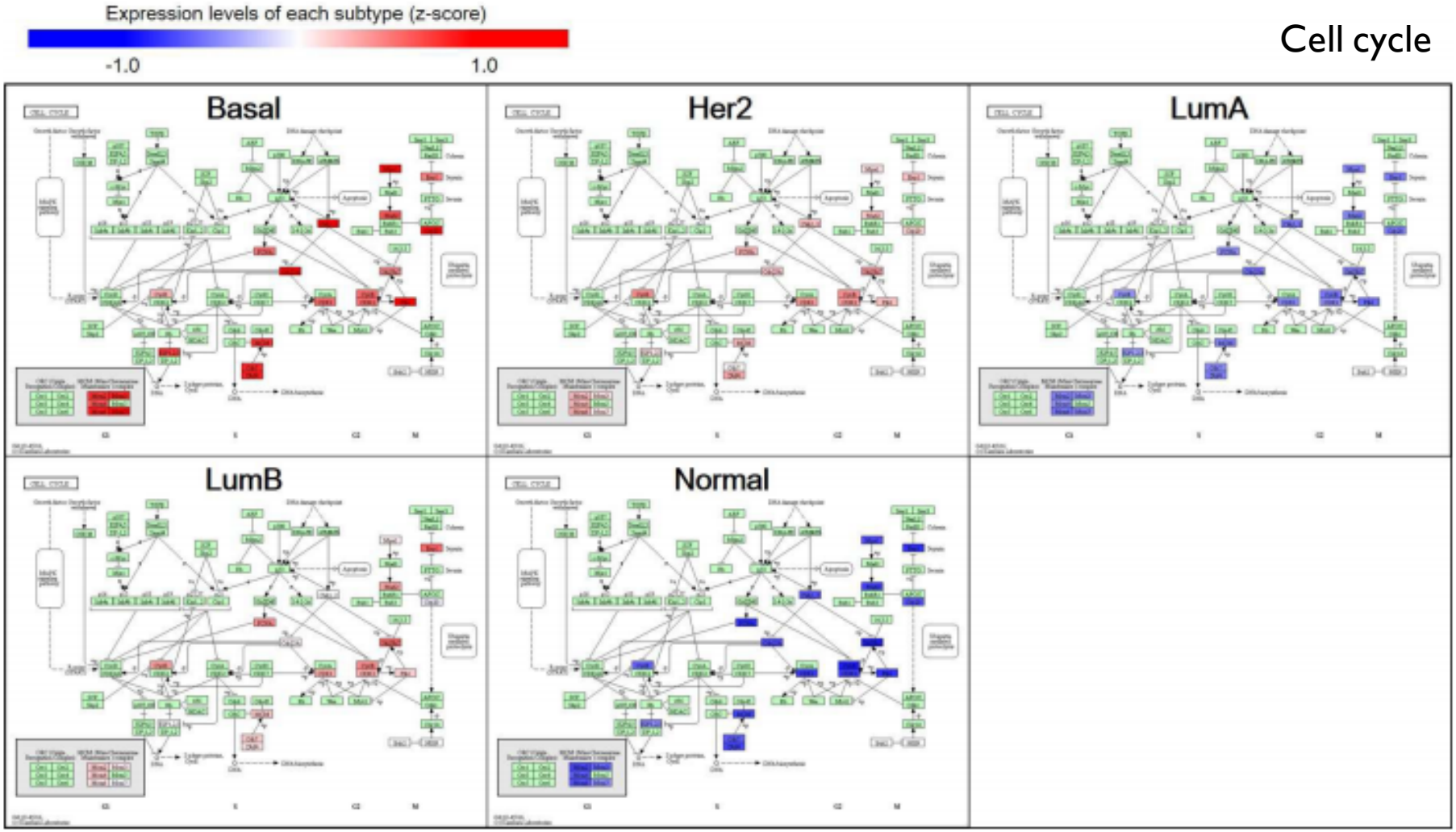
"Prioritized pathways also explain difference among the subtypes"



Result and Discussion

Cell cycle pathway

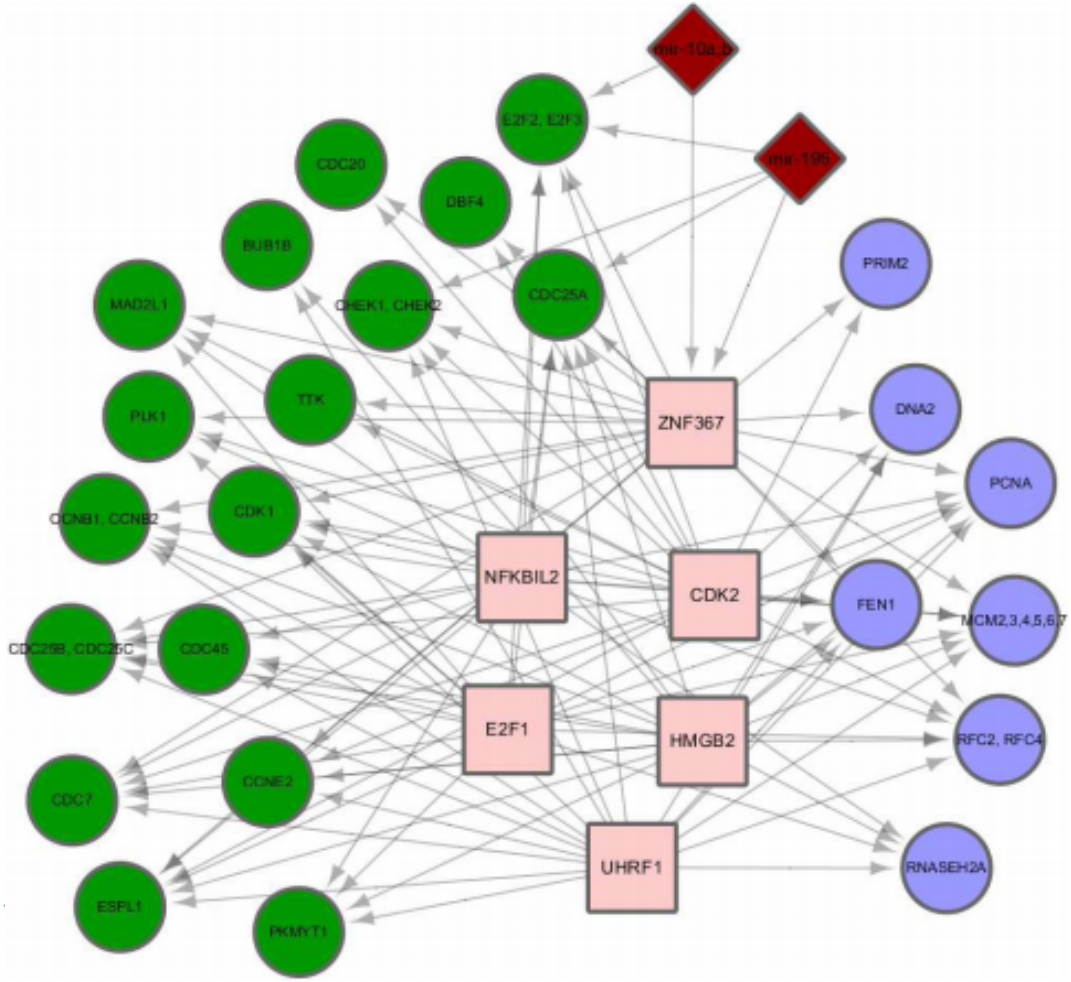
"Prioritized pathways also explain difference among the subtypes"



Result and Discussion

Regulators of the core gene set

"We investigated the major regulators (TF and miRNAs) in the top two pathways (DNA replication and cell cycle)"



Result and Discussion

Regulators of the DNA replication pathway

- We found genes that mapped to DNA replication pathway are regulated by TFs of UHRF1, NFKBIL2, E2F1, ZNF367 and CDK2

- TFs such as UHRF1, CDK2, and E2F1 are well known for a crucial role in breast cancer progression
  - CDK2 is extensively studied in its role in tumorigenesis of breast cancer as it controls the estrogen mediated growth signaling
  - UHRF1 is known as a key factor of DNA replication system
  - E2F1 has also been studied for its function to drive breast cancer and observed to have different expression patterns among several subtypes

Result and Discussion

Regulators of the cell cycle pathway

- ZNF367, CDK2, UHRF1, and HMGB2 are identified as TFs that regulates cell cycle pathway by DeSPA

- Most of these TFs (three out of four) are common with the TFs that regulates DNA replication
  - It is natural to have common regulators as DNA replication is involved in the cell cycle process
  - UHRF1, a key regulator for DNA replication, is also known to cause abnormal cell proliferation due to its expression status
  - CDK2 is also studied to cause immediate arrest at late G1 and S phase of the cell cycle when CDK2 is inhibited

Result and Discussion

miRNA regulating TF

- We also investigated miRNAs that regulates six core TFs targeting the both DNA replication and cell cycle pathways

- miRNAs that have a negative correlation with TFs in terms of expression levels are selected for further analysis
  - hsa-mir-195 and hsa-mir-10a,b showed strong negative correlation with ZNF367
  - hsa-mir-15 is known to regulate genes involved in cell division and angiogenesis
  - mir-195 is down-regulated and has an effect in various cancers

Result and Discussion

Genomic alteration and gene expression

"Genomic alteration did not affect the expression level of gene sets"

- To check on that the alteration of gene expressions from genomic alterations such as SNP, rather than by regulators such as TFs and miRNA
- The number of mutated samples of the three genes are very small as we can see the fourth and fifth column
  - → Indicating that genomic alteration did not affect the expression level

				# of mutated samples/total samples	
Gene	Subtype A	Subtype B	*p*-value	Subtype A	Subtype B
BLM	Her2	Luminal A	0.0343	4/151	1/343
BTRC	Her2	Luminal A	0.0296	3/152	0/344
BTRC	Her2	Luminal B	0.0466	3/152	0/274
FANCD2	Basal	Luminal A	0.0307	4/245	0/344
FANCD2	Luminal A	Luminal B	0.0382	0/344	4/270
FANCD2	Luminal A	Normal	0.0446	0/344	3/187

Contents

Summary

- We developed an information theoretic sub-network mining algorithm, DeSPA combined
  - (1) TF and miRNA regulatory networks
  - (2) information theoretic sub-network mining that is designed to handle multiple classes or subtypes.

- DeSPA was able to find genes that not only can **classify subtypes** but also **explain the important cancer hallmarks in breast cancer** (such as DNA replication and cell cycle)

- In summary, our study is significant in that we suggested a new set of gene sets that not only can classify subtypes but also explain the mechanism of breast cancer that can be very useful in clinical applications

Funding

- (1) This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and Future Planning (No. NRF-2012M3C4A7033341)

- (2) Collaborative Genome Program for Fostering New Post-Genome industry through the National Research Foundation of Korea(NRF) funded by the Ministry of Science ICT and Future Planning (NRF-2014M3C9A3063541)

- (2) BK21 Plus for Pioneers in Innovative Computing (Department of Computer Science and Engineering, SNU) funded by National Research Foundation of Korea (NRF) (21A20151113068).

Thank you for your attention