Network Model-Based Pathway Activity Quantification towards Personalized Drug-Response Prediction

Sungwon Jung, Ph.D.
Department of Genome Medicine and Science
Gachon University School of Medicine

The 15th Korea-Japan-China Bioinformatics Symposium 2017
Jun 21, 2017

가천유전체의과학연구소
Gachon Institute of Genome Medicine and Science
嘉泉基因组医科研究院

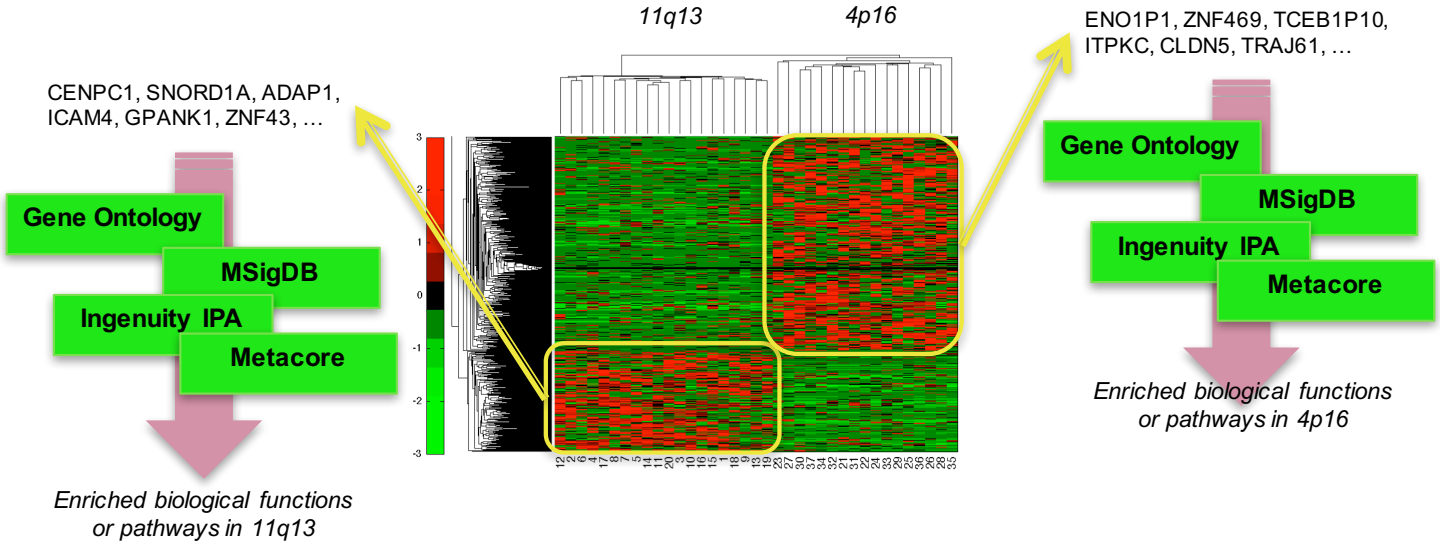
가천대학교
Gachon University

Talk Outline

- **Understanding biological contexts from pathway-centric perspectives**
- **Pathway activity quantification based on differential genetic associations**
  - Evaluation of Differential DependencY (EDDY)
  - Knowledge-based EDDY (KEDDY)
- **Extending pathway activity quantification to single sample cases**
- **Individualized pathway signaling quantification and its correlation with drug responses**

Gene-Centric Approach of Identifying Differential Contexts

- **Typical steps of gene-centric approaches (with an example of utilizing gene expression data)**
  - I. Prepare/preprocess gene expression data
  - II. Testing each gene for differential expression between conditions
    - Necessary adjustments (ex. P-value correction for multiple testing)
  - III. Identifying genes with differential expressions
  - IV. Based on annotations, investigating enriched annotations within the identified list of genes

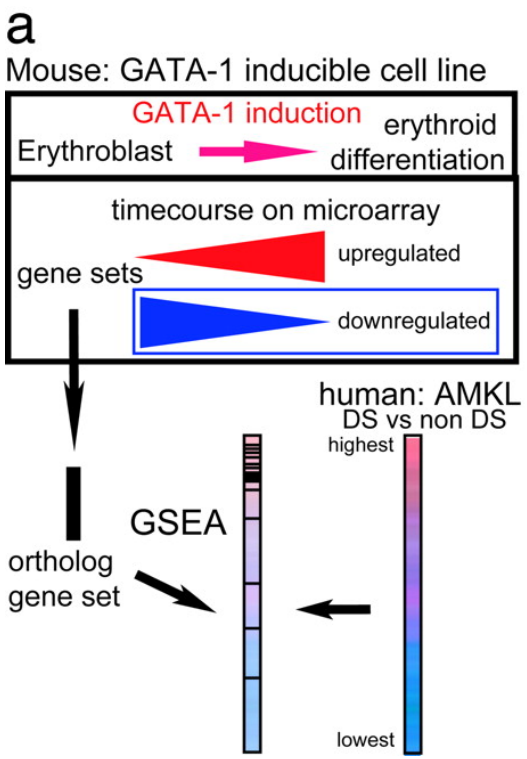


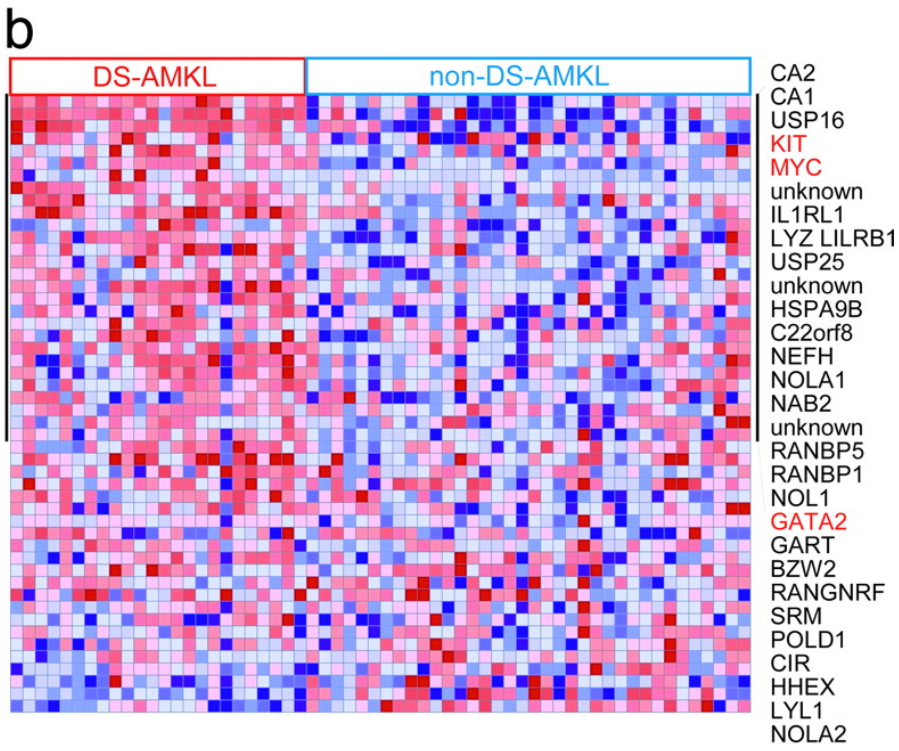
Limitations of Gene-Centric Approaches

- **No clear interpretation of genetic regulations**
  - "Okay, we identified a list of genes, which may be related to biological functions XX, YY, ZZ …"
- **Motivation: Can we provide more information on "differential biological functions"?**
  - More direct interpretation can be possible.

Pathway-Centric Approach of Identifying Differential Contexts

- **Identifying differential pathways**
  - GSEA (Subramanian et al., PNAS 2005 / Mootha, Nature Genetics, 2003)

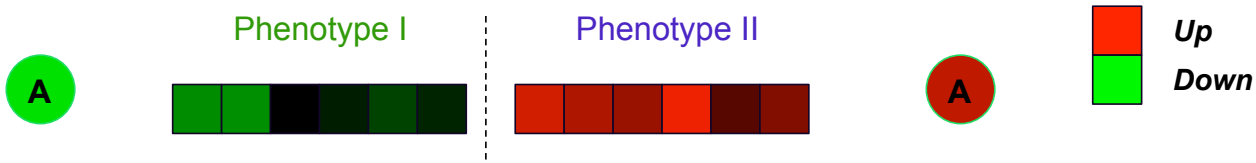




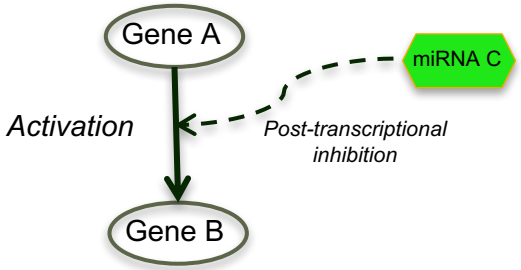
Based on (sort of) differential gene expressions

Limitation of Differential Expression Approaches

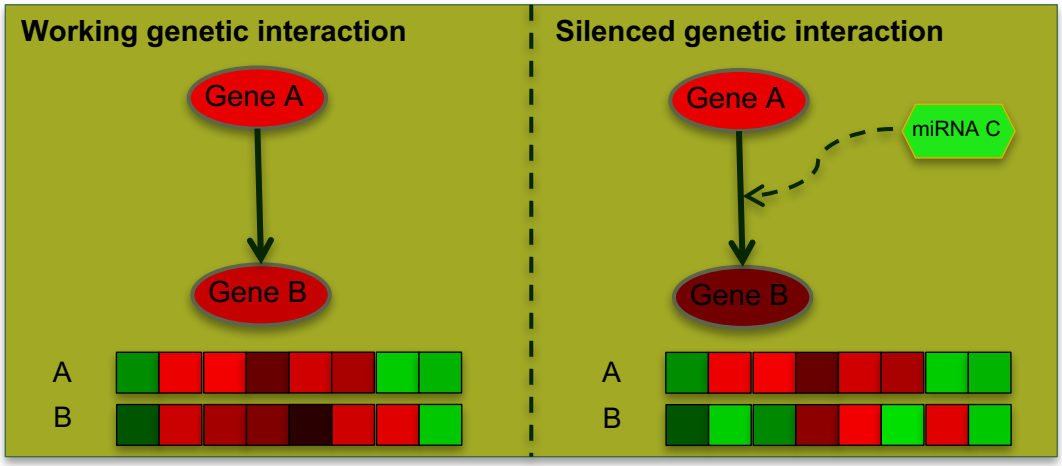
- **Differential expression-based approaches assume similar expression levels within a context**



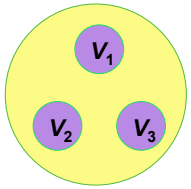
- **However, there can be biological contexts where genes do not show similar expressions within.**
  - Potentially due to other hidden factors



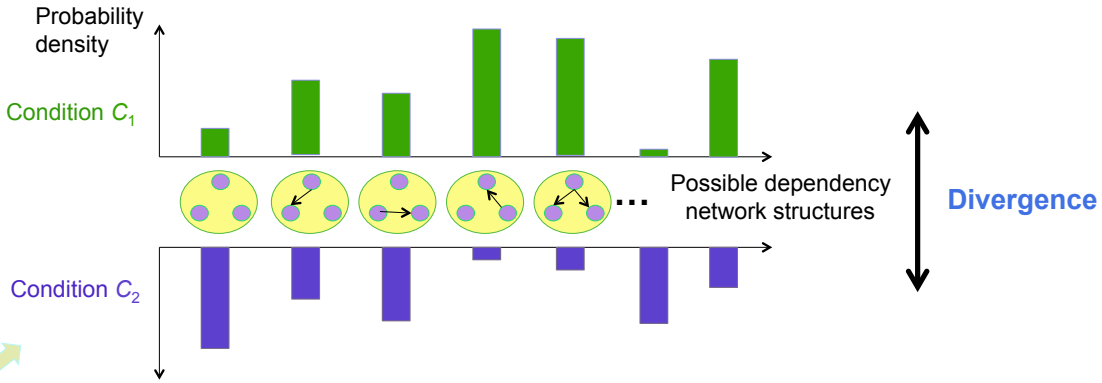
Biological contexts without similar gene expression



Evaluation of Differential DependencY (EDDY)

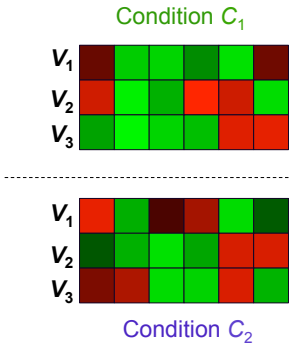


Target set of variables V

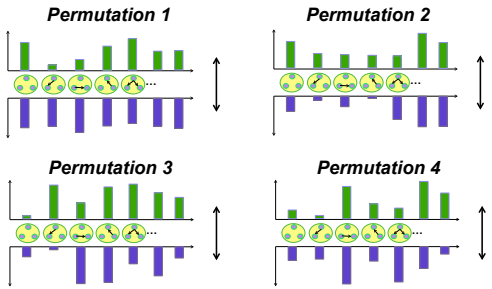


Computing divergence

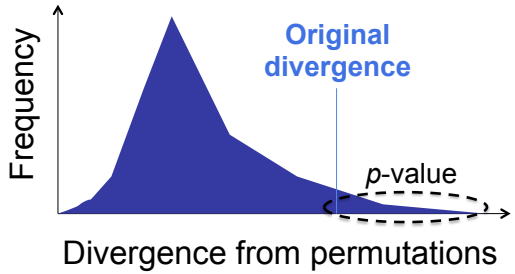
Data of two conditions



Evaluating statistical significance by randomly permuting condition labels



...



(Jung and Kim, Nucleic Acids Research 2014)

Application of EDDY to Glioblastoma Subtypes

- **TCGA GBM gene expression data**
  - Four subtypes (Classical, Mesenchymal, Neural, and Proneural)

- **Identifying subtype-specific pathways with differential genetic associations**
  - Testing of one subtype vs. the others
  - 2,101 pathway gene sets were tested (from MSigDB)

Table 1. The number of statistically significant gene sets for each subtype.

Method	Classical	Mesenchymal	Neural	Proneural
EDDY	13	10	22	22
GSEA	1 (0)	245 (1)	6 (0)	3 (0)
GSCA	1,590 (11)	1,432 (7)	1,681 (21)	1,563 (17)

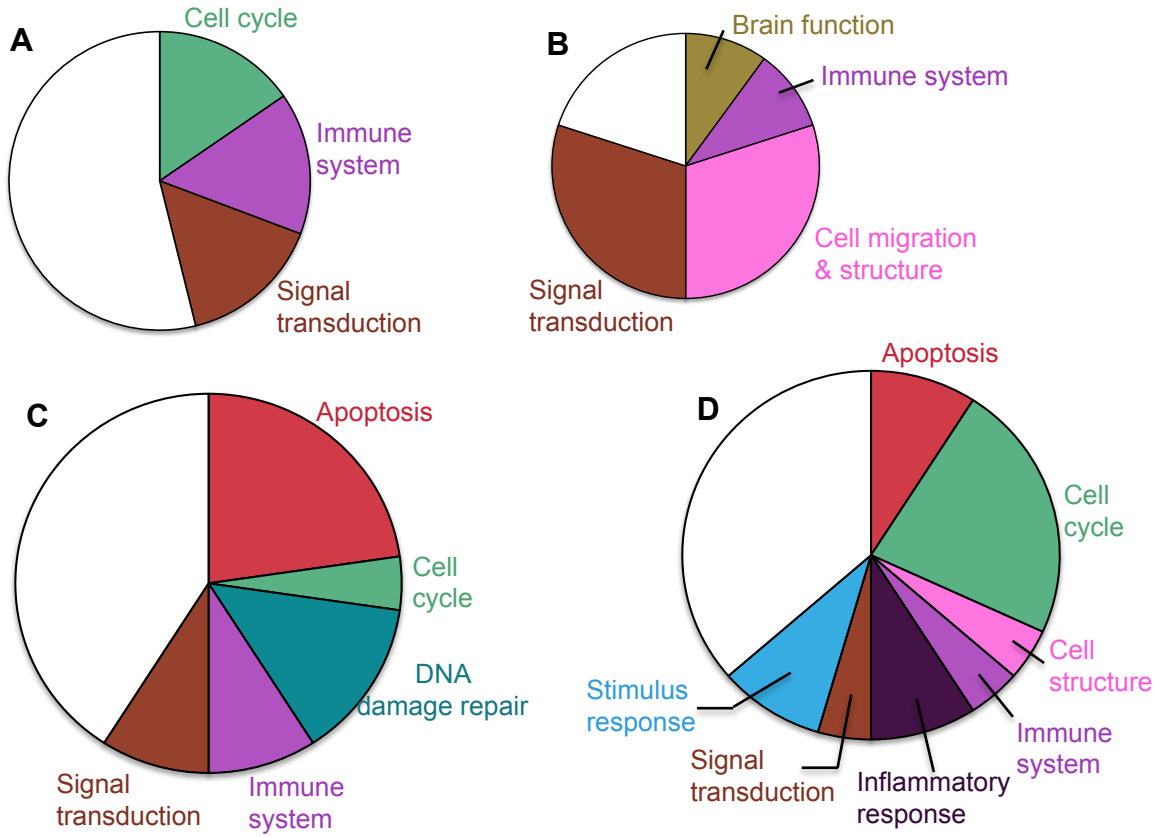
The number of common cases with EDDY is indicated in the parentheses.

Detects differentially expressed genes

Another method considering differential associations

(Jung and Kim, Nucleic Acids Research 2014)

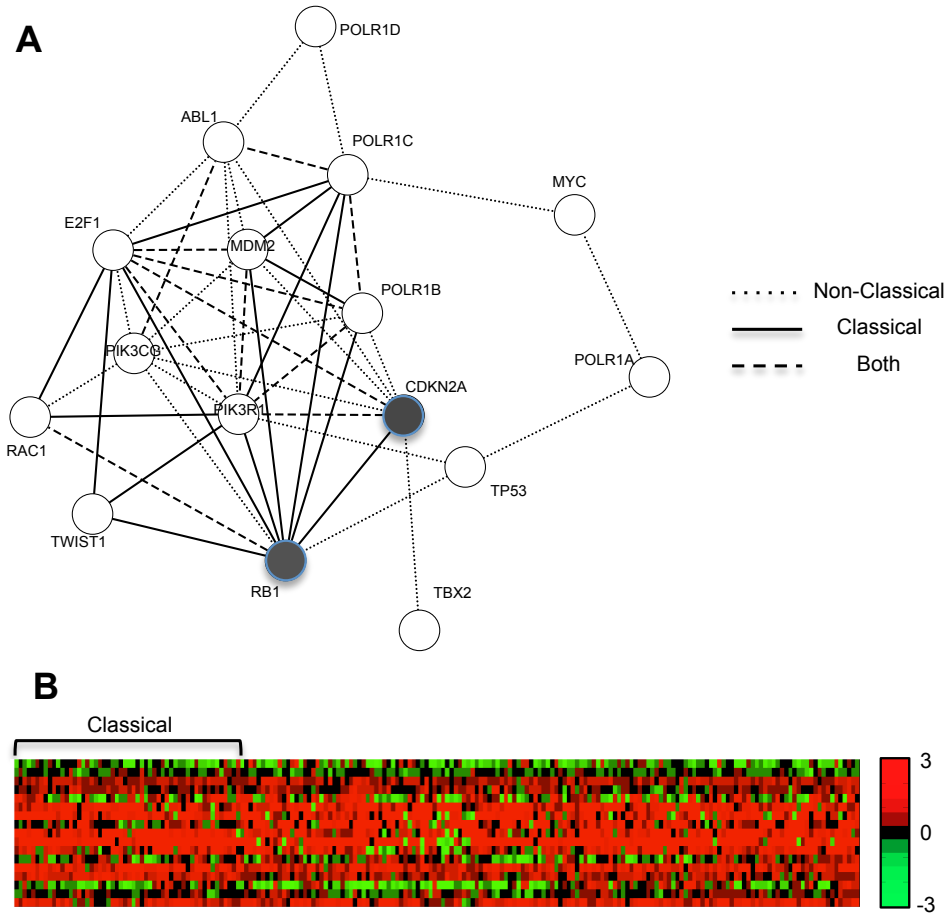
Application of EDDY to Glioblastoma Subtypes



(Jung and Kim, Nucleic Acids Research 2014)

Application of EDDY to Glioblastoma Subtypes

- One example: ARF pathway from the Classical subtype



(A) A lot of interaction differences between Classical and non-Classical samples

(B) No clear differential expression of individual genes between Classical and non-Classical

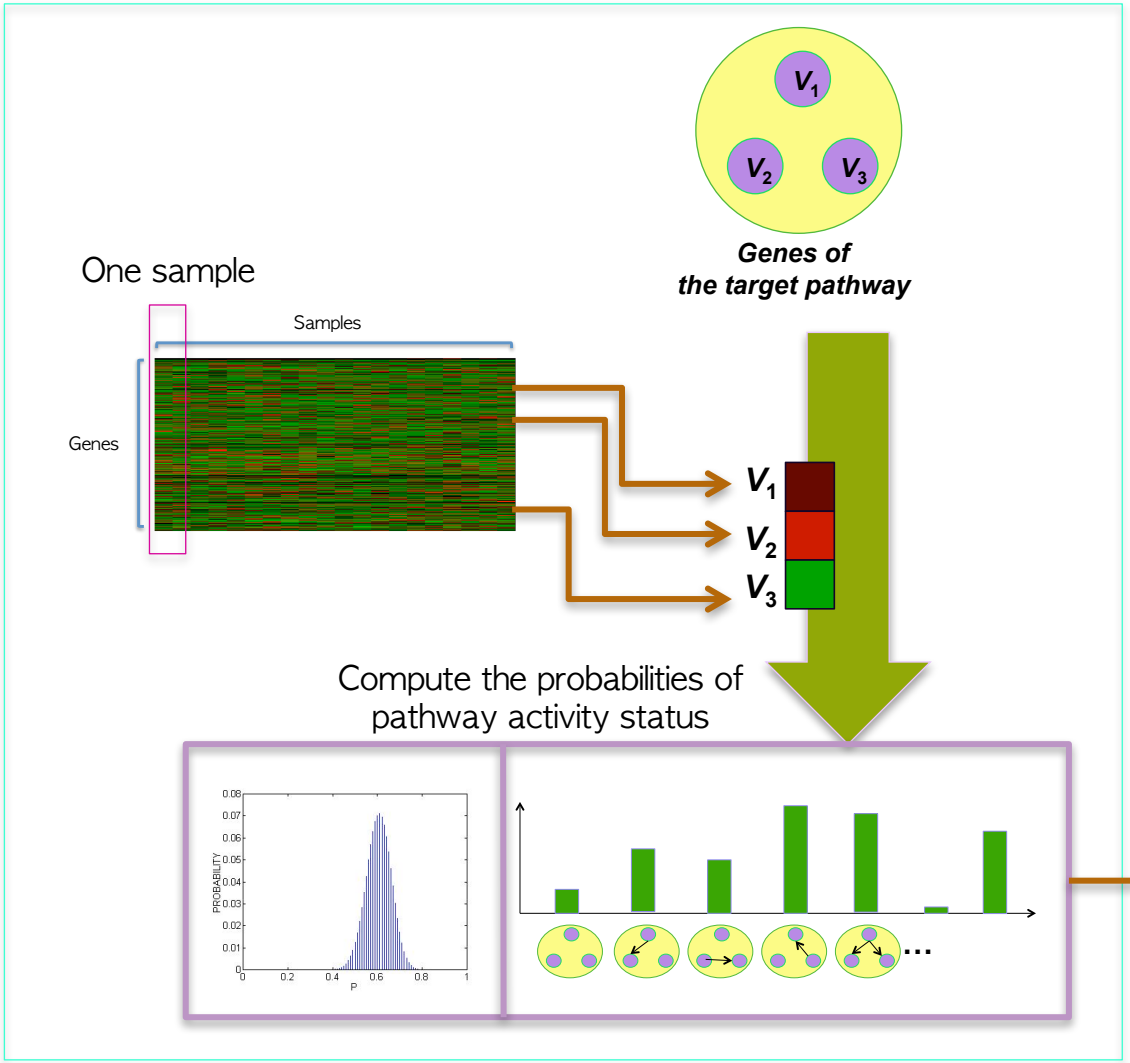
- From the Classical subtype, there is focal 9p21.3 homozygous deletion targeting CDKN2A. (Verhaak et al., Cancer Cell 2010)
- RB pathway is almost exclusively affected through CDKN2A deletion.
- (A) CDKN2A lost many interactions in Classical, while the only activated relationship is with RB1.

(Jung and Kim, Nucleic Acids Research 2014)

Limitation of EDDY

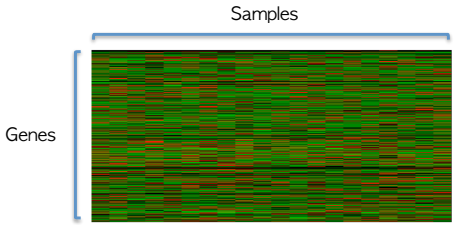
- **Requiring large computational resources**
  - For higher sensitivity, evaluation for more gene association networks is necessary.
  - For larger gene sets (pathways), evaluation of more gene association networks is necessary for comparable sensitivity.
  - => Needs heavy computation
  - Example)
    - Testing about 1,500 pathways for four classes with 202 samples: Took two weeks with about 1,000 HPC cores
  - 'Oh, I need something more handy.'

Extension of KEDDY Idea to Single Samples



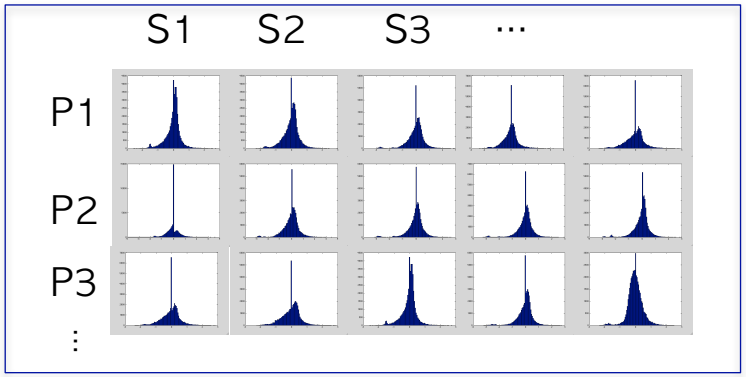
(Jung, BMC Medical Informatics and Decision Making 2016)

Extension of KEDDY Idea to Single Samples



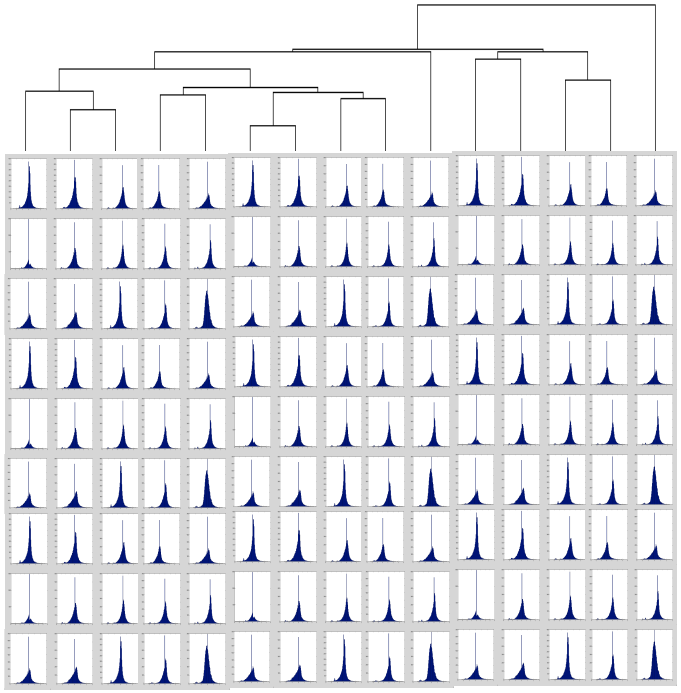
Gene x Sample matrix

Computing activity distributions of ALL PATHWAYS, for ALL SAMPLES



Pathway x Sample matrix

Clustering



Identifying group of samples based on sample pathway activity vectors

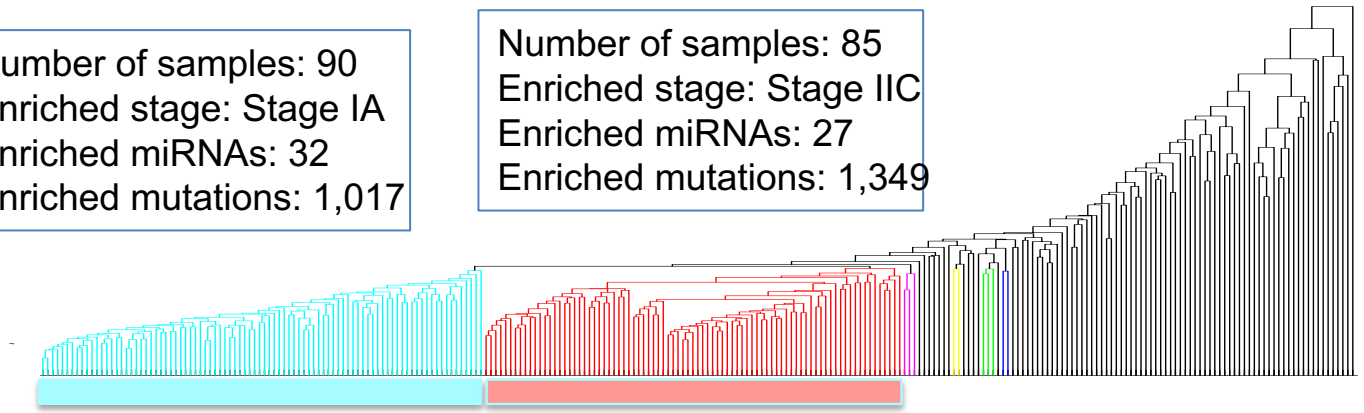
(Jung, BMC Medical Informatics and Decision Making 2016)

Extension of KEDDY Idea to Single Samples: Application

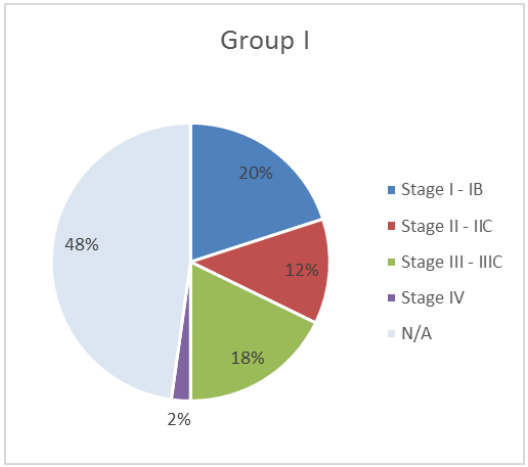
- **TCGA melanoma RNA-seq data**

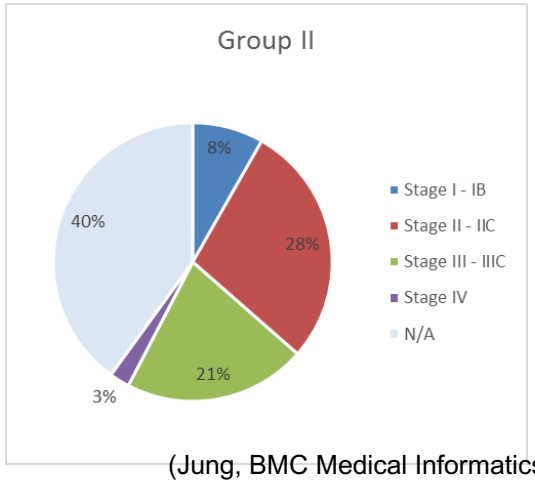
Number of samples: 90
Enriched stage: Stage IA
Enriched miRNAs: 32
Enriched mutations: 1,017

Number of samples: 85
Enriched stage: Stage IIC
Enriched miRNAs: 27
Enriched mutations: 1,349



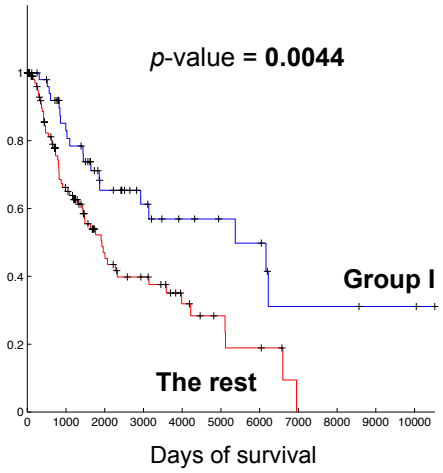
Group I **Group II**



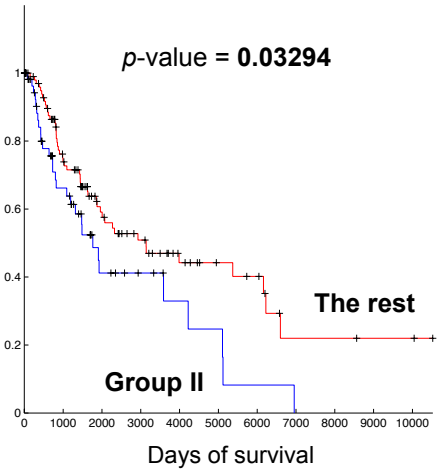


(Jung, BMC Medical Informatics and Decision Making 2016)

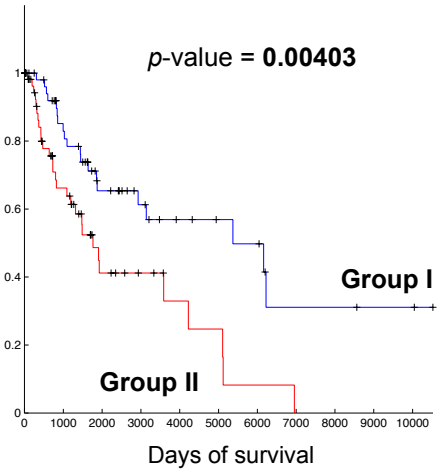
Extension of KEDDY Idea to Single Samples: Application



(A)



(B)



(C)

Group I median survival: 12.6 years
Group II median survival: 4.8 years

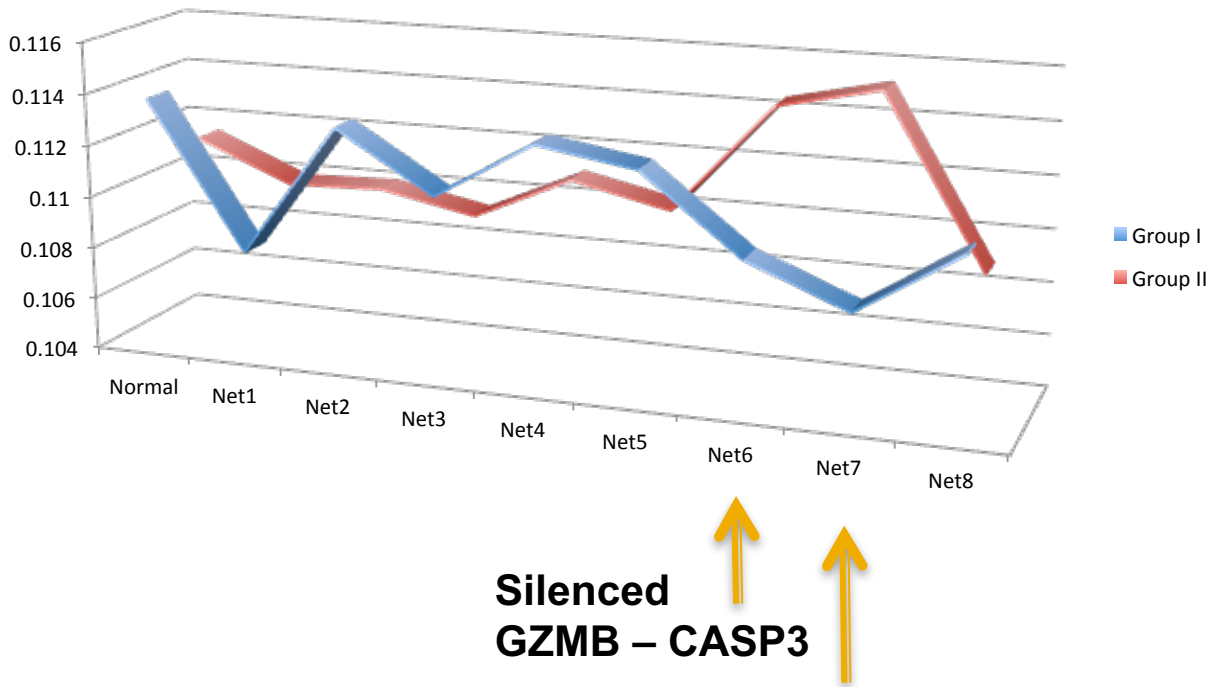
Hazard ratio of Group I respect to Group II: 0.328

(Jung, BMC Medical Informatics and Decision Making 2016)

Extension of KEDDY Idea to Single Samples: Application

- **One example: DNA fragmentation pathway**
  - One of the mechanisms used by immune cells to kill target cells

Average pathway activities of two groups



- **GZMB: Expressed by cytotoxic T lymphocytes and natural killer cells. Activates caspase cascade.**
- **CASP3, CASP7: Caspases. Activation plays a central role in cell apoptosis.**

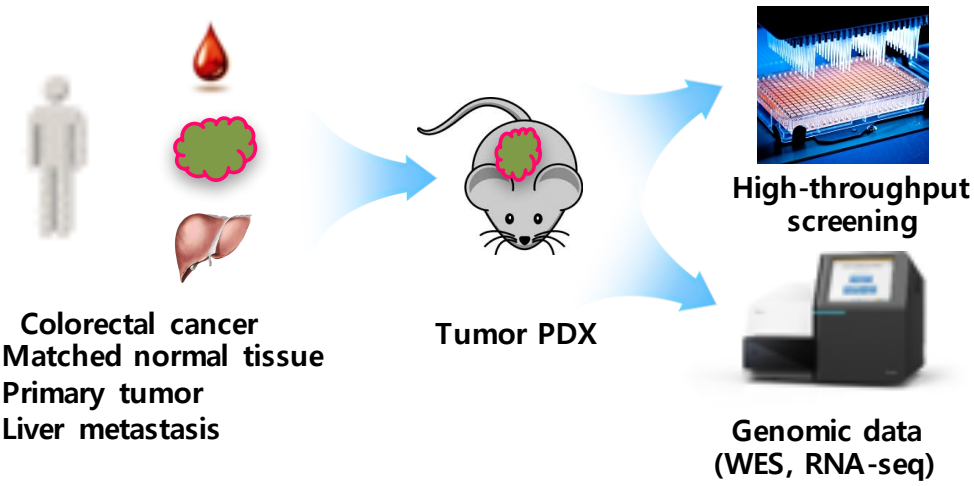
Suggests restricted immune response in Group II, where the caspase cascade activation by GZMB has been silenced.

(Jung, BMC Medical Informatics and Decision Making 2016)

Individual Signaling-Based Correlation of Genomic Profiles and Drug Responses

- **Genomic data of colorectal cancer**
  - 26 samples including primary tumor and liver metastasis
    - With matched normal colon tissues
  - Whole exome-seq, RNA-seq
  - DNA somatic mutations and gene expressions were evaluated.

- **High-throughput drug screening**
  - Screening of 65 FDA-approved drugs for every tumor sample
    - Response: IC50



Collaboration with Prof. Yongbeom Cho at Samsung Medical Center in Korea

Summary

- **The needs of pathway evaluation methods, with abilities of investigating individual genetic associations**
- **Wide range of applications for individualized pathway analysis**
  - Novel disease subtype identification
  - Individualized drug response prediction

Acknowledgements

- **Jung lab**
  - Sora Kim, M.S.
  - Hyojung Kim
  - Joigmin Lee
  - Opening for Post-Doc!
- **Samsung Medical Center**
  - Prof. Yongbeom Cho
  - Hyekyung Hong, Ph.D.
  - Yeosong Lee, Ph.D.
- **Gachon Institute of Genome Medicine and Science**

가천유전체의과학연구소
Gachon Institute of Genome Medicine and Science
嘉泉基因组医科研究院

KHIDI 한국보건산업진흥원

NRF 한국연구재단