OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Network-based Survival Analysis Reveals Subnetwork Signatures for Predicting Outcomes of Ovarian Cancer Treatment

Wei Zhang[1], Takayo Ota[2], Viji Shridhar[2], Jeremy Chien[2], Baolin Wu[3], Rui Kuang[1]*

1 Department of Computer Science and Engineering, University of Minnesota Twin Cities, Minneapolis, Minnesota, United States of America, **2** Department of Laboratory Medicine and Experimental Pathology, Mayo Clinic College of Medicine, Rochester, Minnesota, United States of America, **3** Division of Biostatistics, School of Public Health, University of Minnesota Twin Cities, Minneapolis, Minnesota, United States of America

Presenter: Seo-Jin Bang

This paper aims to..

Identify signature genes that reliably predict the outcomes of ovarian carcinoma.

PLOS | COMPUTATIONAL BIOLOGY

They introduce themselves as…

> By making connections through the application of **computational methods** among disparate **areas of biology**, *PLOS Computational Biology* provides **substantial new insight** into living systems at all scales, from the nano to the macro, and across multiple disciplines, from molecular science, neuroscience and physiology to ecology and population biology.

PLOS | COMPUTATIONAL BIOLOGY

Network-Based Cox Regression (Net-Cox)

They introduce themselves as…

By making connections through the application of **computational methods** among disparate **areas of biology**, *PLOS Computational Biology* provides **substantial new insight** into living systems at all scales, from the nano to the macro, and across multiple disciplines, from molecular science, neuroscience and physiology to ecology and population biology.

PLOS | COMPUTATIONAL BIOLOGY

Network-Based Cox Regression (Net-Cox)

They introduce themselves as…

By making connections through the application of **computational methods** among disparate **areas of biology**, *PLOS Computational Biology* provides **substantial new insight** into living systems at all scales, from the nano to the macro, and across multiple disciplines, from molecular science, neuroscience and physiology to ecology and population biology.

Survival Analysis in Cancer Genomics

PLOS | COMPUTATIONAL BIOLOGY

Network-Based Cox Regression (Net-Cox)

They introduce themselves as…

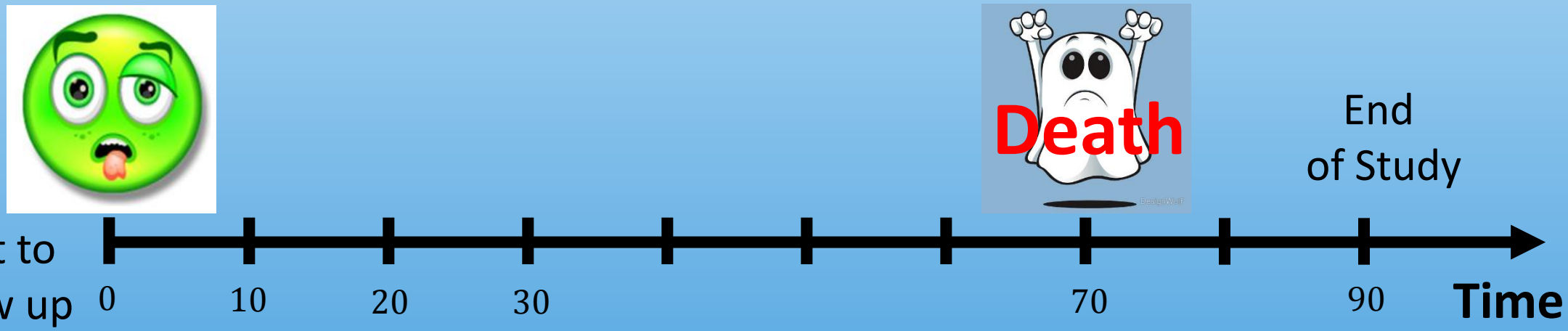
By making connections through the application of **computational methods** among disparate **areas of biology**, *PLOS Computational Biology* provides **substantial new insight** into living systems at all scales, from the nano to the macro, and across multiple disciplines, from molecular science, neuroscience and physiology to ecology

Prior network information improves..

Selected genes are enriched in ..

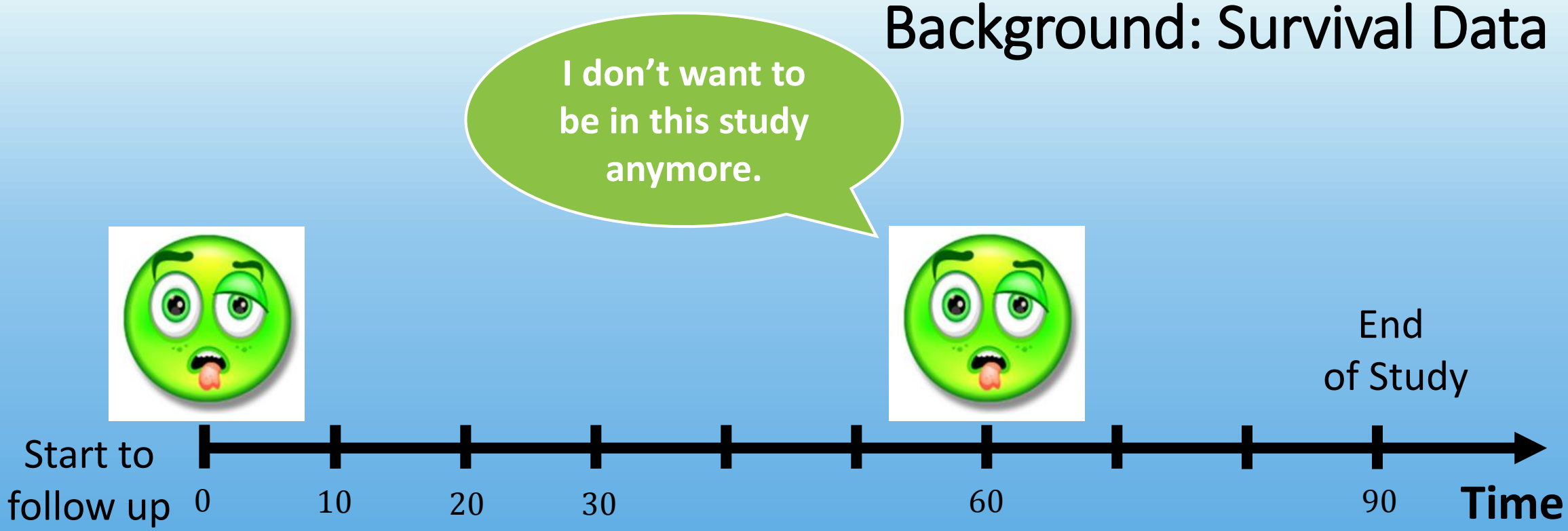
Survival Analysis in (Ovarian) Cancer Genomics

Background: Survival Data



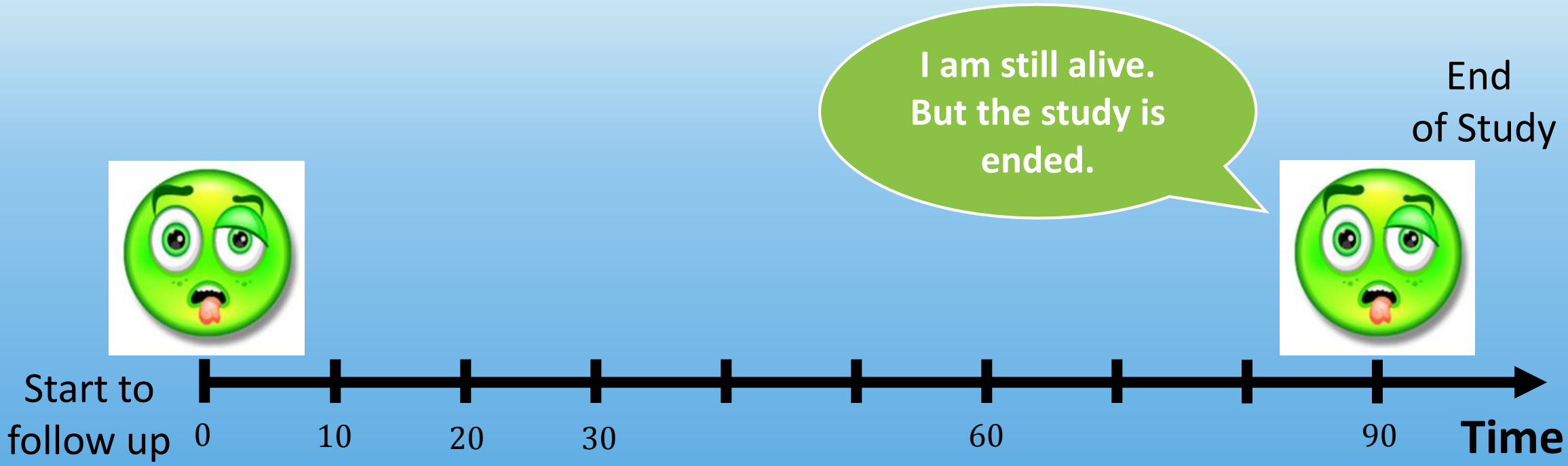
The death is observed ($\delta_i = 1$) at time 70 ($t_i = 70$)

Background: Survival Data



The sample is censored ($\delta_i = 0$) at time 60 ($t_i = 60$)

Background: Survival Data



The sample is censored ($\delta_i = 0$) at the study end ($t_i = 90$)

Background: Hazard Function $h(t)$

Instantaneous rate of event (ex. death) at time t on no event before t.

$$h(t) = \lim_{dt \to 0} \frac{P(t < T < t + dt)}{dt\ P(T > t)}$$

Background: Why we use Hazard Function?

- Intuitive interpretation.
- Easy to derive a survival function

$$P(T > t) = exp\left[\int_0^t h(u)du\right]$$

- Easily modeled with Cox Proportional Hazard Model

Background: Cox Proportional Hazard Model

$$h(t|\boldsymbol{X}_i) = h_0(t)\exp(x_{1i}\beta_1)\cdots\exp(x_{pi}\beta_p)$$
$$= h_0(t)\exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta}\right)$$

The underlying hazard function, $h_0(t)$, is a hazard function at time t at baseline levels of covariates.

The Cox model assumes that covariates, $\boldsymbol{X}_i^T = (x_{1i}, \cdots, x_{pi})$, are multiplicatively related to the hazard function.

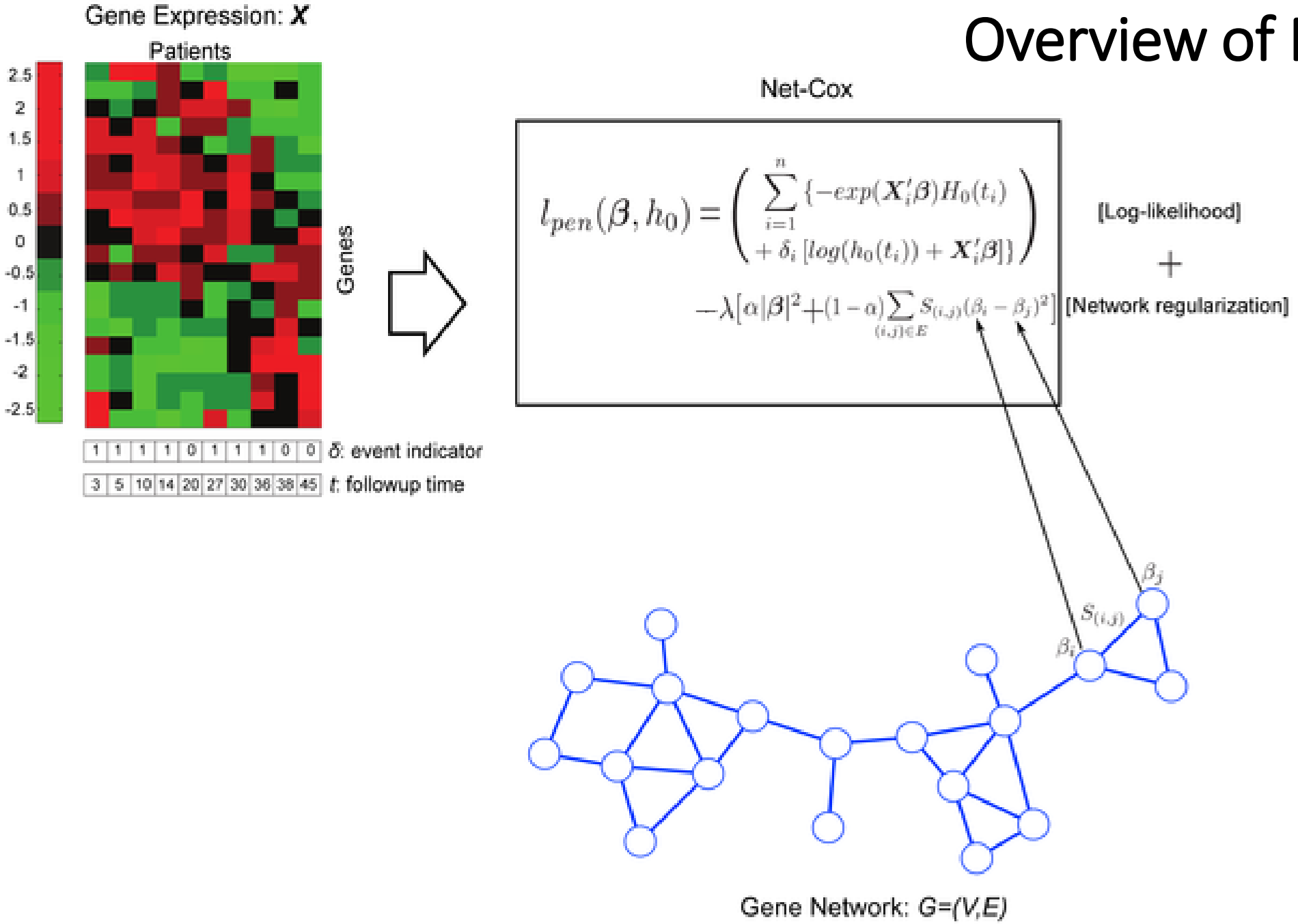
Background: Cox Proportional Hazard Model

Maximum Likelihood Estimate (MLE): Find estimates of $h_0(t)$ and $\boldsymbol{\beta}$ that are most likely to be in the model with the observed data.

$$\underset{\boldsymbol{\beta},\, h_0}{\operatorname{argmax}}\{l(\boldsymbol{\beta}, h_0)\} = \sum_{i=1}^{n}\left[-\exp\left(\boldsymbol{X}_i^T\boldsymbol{\beta}\right)\ H_0(t_i) + \delta_i\left\{\log\left(h_0(t_i)\right) + \boldsymbol{X}_i^T\boldsymbol{\beta}\right\}\right]$$

Where $H_0(t_i)=\sum_{t_k \le t_i} h_0(t_k)$

Overview of Net-Cox



Key Idea of Net-Cox

$$l_{pen}(\boldsymbol{\beta}, h_0) = l(\boldsymbol{\beta}, h_0) - \frac{1}{2}\lambda\left[\alpha|\boldsymbol{\beta}|^2 + (1-\alpha)\sum S_{(i,j)}(\beta_i - \beta_j)^2\right]$$

Original log-likelihood

L_2-term:
Coefficients β_j will be consistently estimated across different data sets. (said robust estimates)

Network Information term:

- $S_{(i,j)}$ should represents intensity of relationship between two genes i and j.
- As genes have strong relationship, they are assigned similar coefficient values.

Note1: By using singular value decomposition, the problem dimension is reduced from p (large) to n (small)
Note2: λ and α are obtained by maximizing the cross-validation log-partial likelihood

Details about Net-Cox

"$\boldsymbol{S_{(i,j)}}$ should represents intensity of relationship between two genes $\boldsymbol{i}$ and $\boldsymbol{j}$"

1. Gene co-expression network
   : $\boldsymbol{S_{(i,j)}}$ is defined as a function of the Pearson's correlation coefficients between gene i and j.

2. Gene functional linkage network
   : $\boldsymbol{S_{(i,j)}}$ is defined as a quantity of the functional relation between two genes.

Details about Net-Cox

- Genes are ranked by size of $|\widehat{\beta}_j|$
- If one with high hazard ratio is dead (or already dead) at time t or one from low hazard ratio is alive or unknown at time t, we can say **"the survival prediction for the individual at time t is successful."**
- Predict survival status of all patients at time t for every possible threshold for the hazard ratio. Then we could get a AUC value at each time t

Same method, similar interpretation

Data

"Death" is the event. ($\delta_i = 1$ if a dead is observed.)

- **Response (survival data)**

	Dataset (GEO ID)	TCGA (N/A)	Tothill (GSE9899)	Bonome (GSE26712)
Death	**# of Censored**	227	160	24
	# of Uncensored	277	111	129
Recurrence	**# of Censored**	241	86	N/A
	# of Uncensored	263	185	N/A

The number of patients categorized by censoring and uncensoring for the death and recurrent events is reported in each dataset. Note that the Bonome dataset does not provide information on recurrence.
doi:10.1371/journal.pcbi.1002975.t001

"Recurrence" is the event. ($\delta_i = 1$ if a recurrence is observed.)

- **2647 genes that are previously known to be related to cancer are used. (Sloan-Kettering cancer genes)**

Result

Top 15 signature genes identified by different methods.

Death

Net-Cox (Co-exp)	Net-Cox (FL)	$L_2-\mathrm{Cox}$
FBN1	COL11A1	COL11A1
COL5A2	MFAP4	FABP4
VCAN	TIMP3	MFAP4
SPARC	MFAP5	COMP
AEBP1	COL5A2	BCHE
AOC3	THBS2	FAP
COL3A1	FAP	COL5A2
THBS2	CXCL12	MFAP5
PLN	AEBP1	TIMP3
ADIPOQ	RYR3	THBS2
COL5A1	LOX	HOXA5
CNN1	COL5A1	NUAK1
COL6A2	EDNRA	COL5A1
COL1A2	NUAK1	SLIT2
DCN	LPL	CXCL12

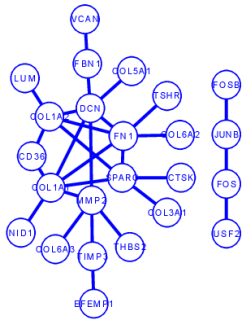
Explain columns

- **More ovarian cancer related genes are detected in Net-Cox**
- **Several ovarian cancer related genes are identified only in *Net-Cox***

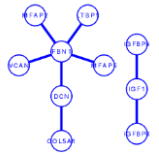
- Detected in both *Net-Cox* and L_2-*Cox*
- Already known to be relevant to ovarian cancer

- Detected only in *Net-Cox*
- Already known to be relevant to ovarian cancer

Result



Net-Cox



L_2-Cox

- Select top 100 signature genes.
- Construct the human protein-protein interaction (PPI) networks using the 100 genes.
- The PPI networks identified by Net-Cox are larger and denser

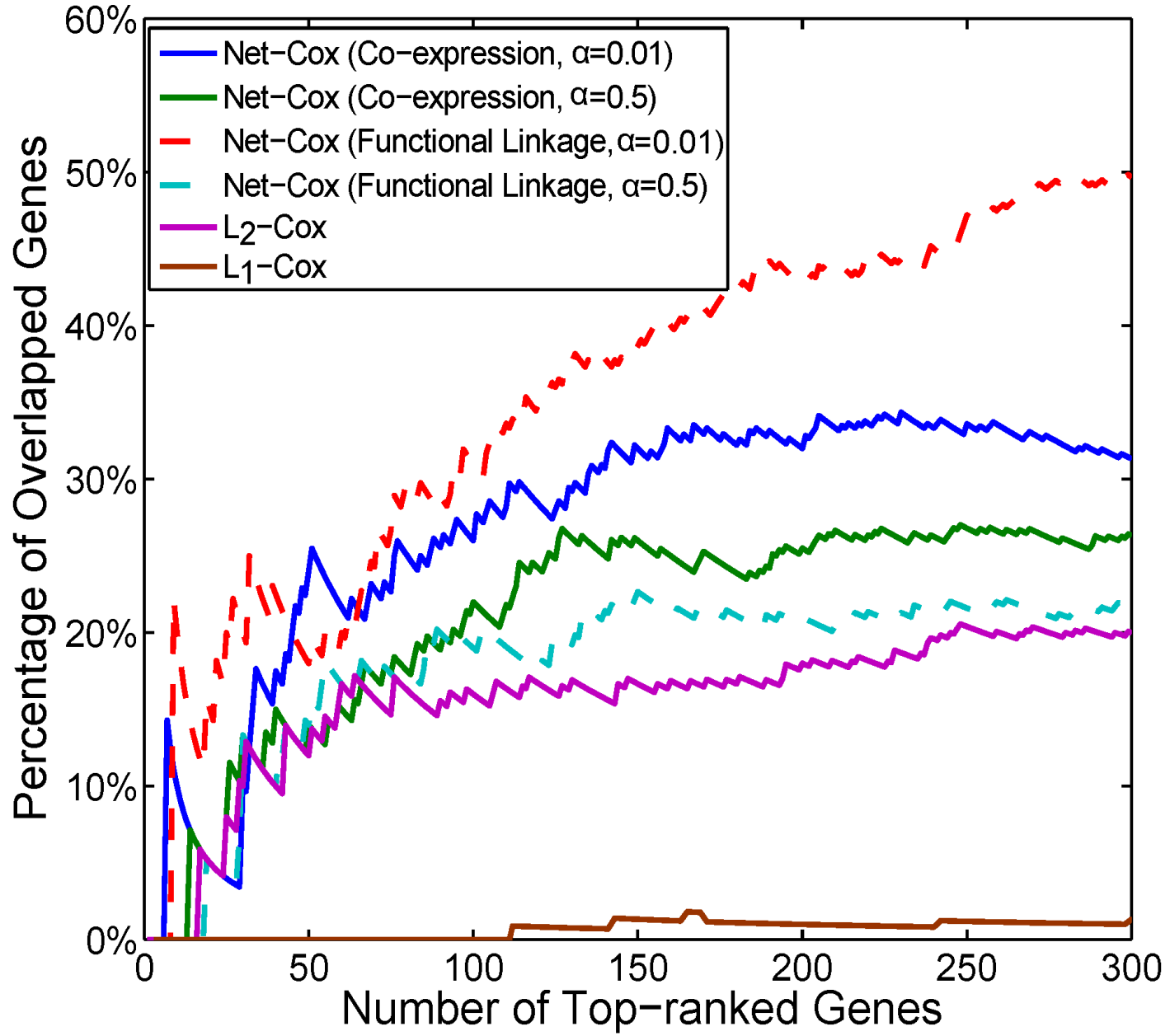
Result

- Most identified genes are stromal or either components or modulators of extracellular matrix (ECM)
- In KEGG pathway and GO enrichment analysis, extracellular matrix, region, and structure organization are also consistently the most significantly enriched
- It was shown ECM acts as a model substratum for the preferential attachment of human ovarian tumor cells in vitro

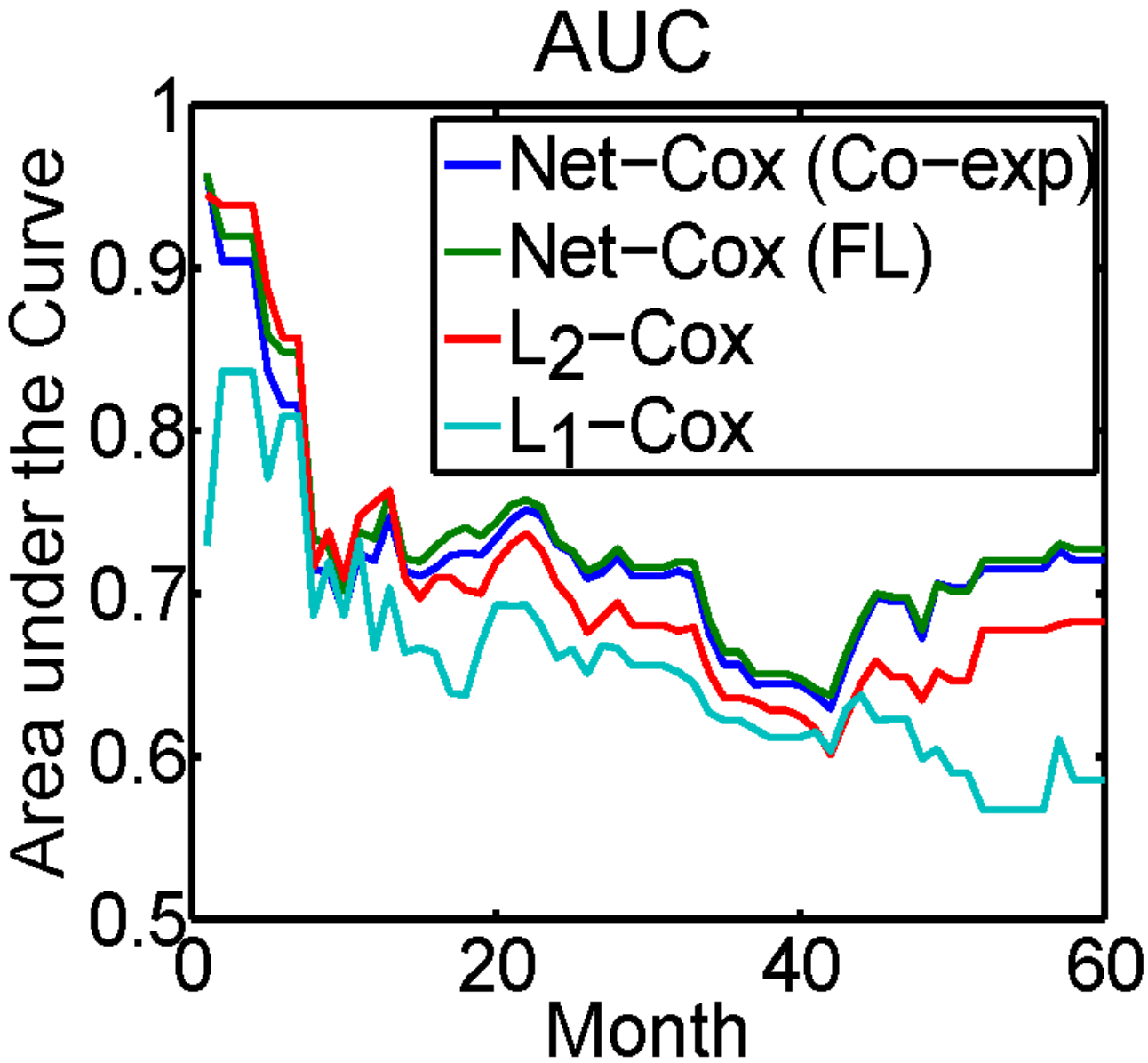
Quantitative Result

- How does the model **consistently select** signature genes across independent data sets
- How well the model **predict the survival rate**
- The role of **network information** ($S_{(i,j)}$)

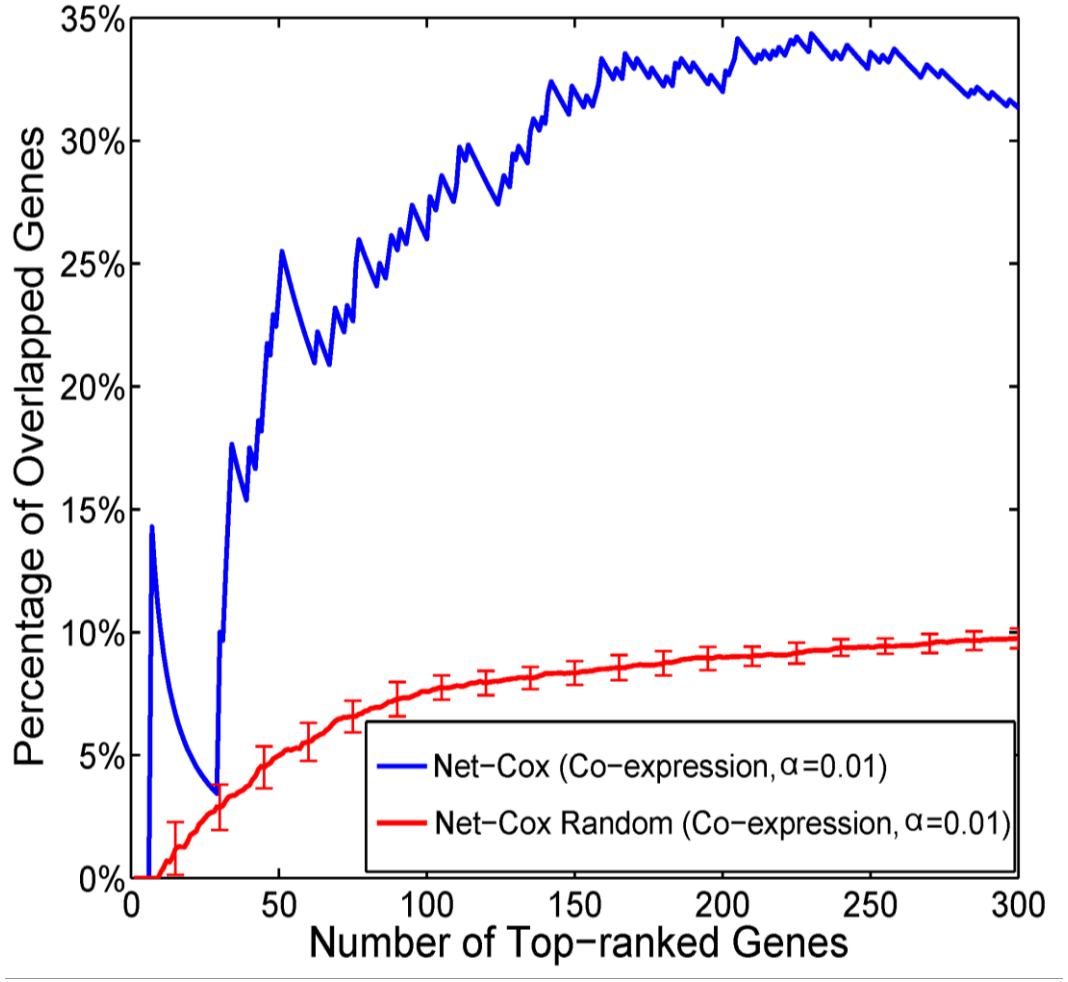
Result: Consistency of gene selection

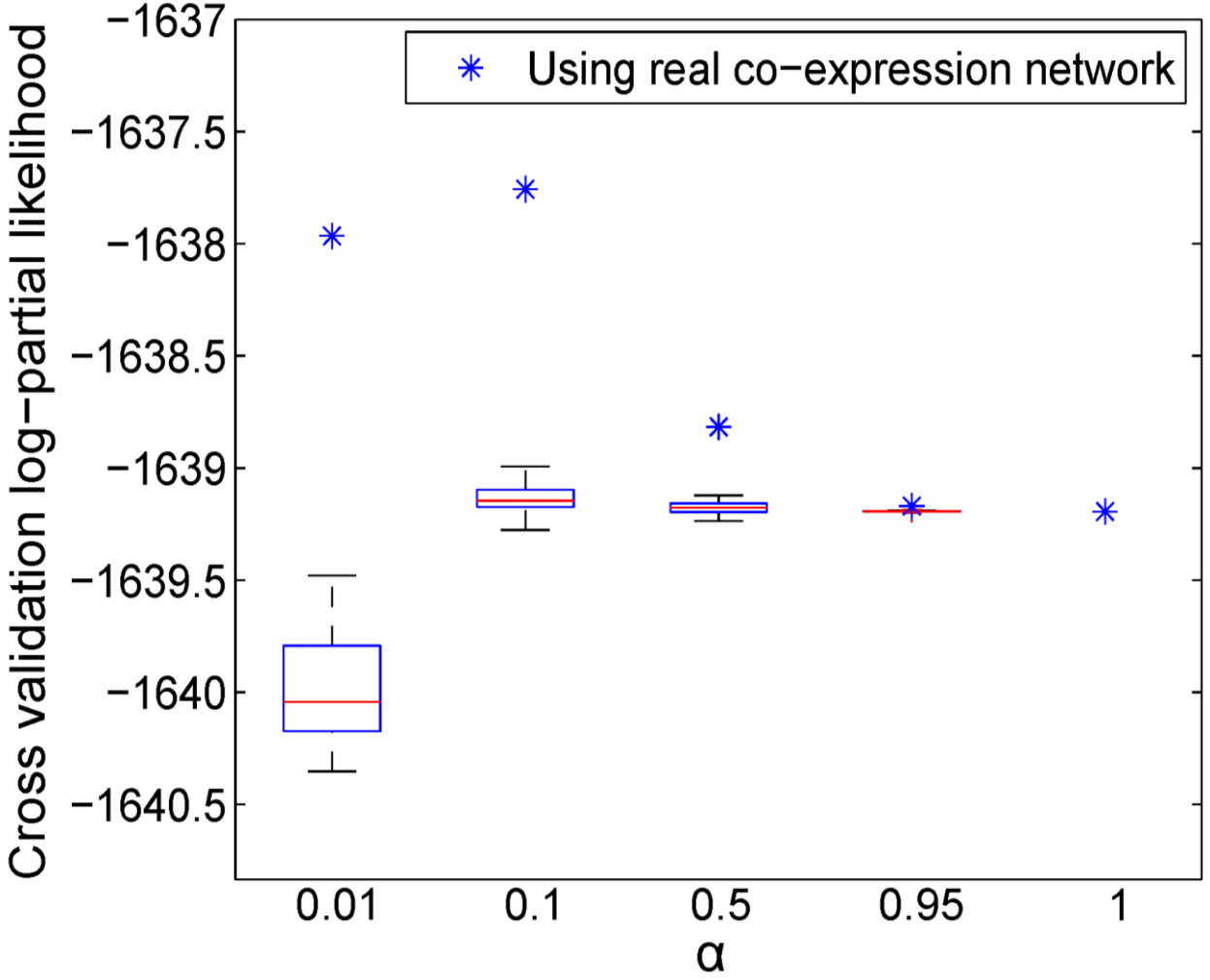


Result: Accurate Prediction Achievement



Result: Network Information improves the model





Summary

- This paper propose Net-Cox, a network-based survival model.
- The dual form of Net-Cox incorporates **prior information of a network** (from $S_{(i,j)}$), and **robust regression coefficient** (from L_2 term) in survival analysis.
- Net-Cox **consistently selects** signature genes and improves **prediction** achievement compared to L_1-Cox and L_2-Cox models.
- The literature research, enrichment analysis, and laboratory experiment of the signature genes also support Net-Cox model.