A SYSTEM-LEVEL APPROACH FOR PREDICTING PATIENT SURVIVAL USING MULTI-OMICS DATA

MOUTAIN WEST CTR-IN PROGRAM

AUBURN UNIVERSITY

Ha Nguyen[1], Phi Hung Bya[1], Quang-Huy Nguyen[1] & Tin Nguyen[1,2,*]

[1] Department of Computer Science and Software Engineering, Auburn University, AL
[2]Department of Computer Science and Engineering, University of Nevada, Reno
Contact: tinn@auburn.edu. Website: https://tinnguyen-lab.com

01/30/2024

TCGA BY THE NUMBERS

TCGA data describes **33** DIFFERENT TUMOR TYPES ...including **10** RARE CANCERS ...based on paired tumor and normal tissue sets collected from **11,000** PATIENTS

Racial Categories	Ethnic Categories									Total
	NOT HISPANIC OR LATINO			HISPANIC OR LATINO			UNKNOWN			
	Female	Male	Unknown	Female	Male	Unknown	Female	Male	Unknown	
AMERICAN INDIAN OR ALASKA NATIVE	11	6	0	7	1	0	1	2	0	**28**
ASIAN	284	358	0	1	0	0	23	19	0	**685**
NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER	9	1	0	1	0	0	2	0	0	**13**
BLACK OR AFRICAN AMERICAN	517	302	0	11	8	0	93	54	0	**985**
WHITE	3474	3412	0	178	135	0	692	566	0	**8457**
UNKNOWN	20	18	0	21	21	0	472	443	4	**999**
Total	**4315**	**4097**	**0**	**219**	**165**	**0**	**1283**	**1084**	**4**	**11167**

Figure 1. *Summary of TCGA data*

CHARACTERISTICS OF THE METHODS

- It is capable of incorporating KEGG pathway-level knowledge that holds key characteristics of cancer disparities.
- It standardizes pathway expression to enable cross-platform consistency and effective integration of multi-omics data.
- It highly accurately predicts patient survival without recourse to clinical information (e.g., age, gender, disease stage, etc.).

WORKFLOW OF THE METHOD

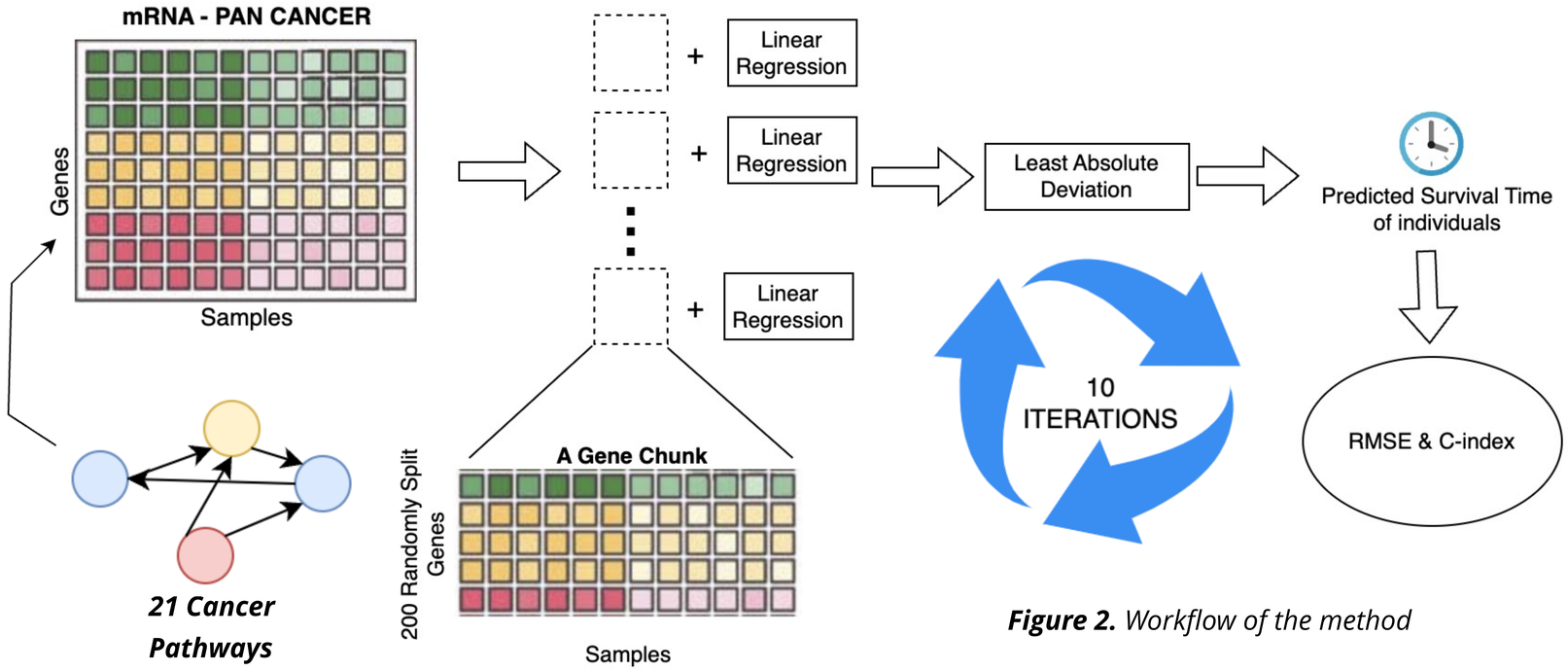

Figure 2. *Workflow of the method*

DATA ACQUISITION

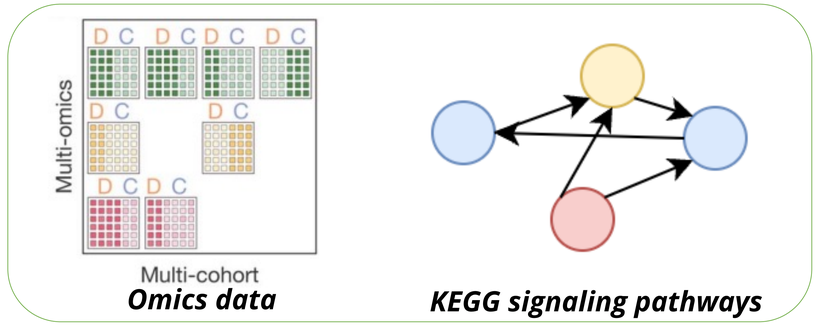

Figure 4. *The required inputs of the method*

- **Omics data:** The method takes as input two types of omics data, including combined mRNA expression and *combined methylation**, associated with 33 cancer types in TCGA and *8 cancer types in TARGET**.
- **Clinical data:** Clinical information of cancer patients, including survival time and survival status, is imported.
- **KEGG signaling pathways in cancer:** A list of KEGG pathways downloaded via API from the KEGG pathway database. Only 21 pathways related to cancer are retained for downstream analysis.

** Future works*

RESULTS

- The method is executed on PAN-cancer data, combining 33 cancer types, for 10 iterations. Each iteration is evaluated based on RMSE and C-index metrics on both the training and test sets.
- The final performance of the method is represented by the average result (bold text).
- The performance of the model is an **RMSE of 31.566 (months)** and a **C-index of 0.742.**

RMSE Train	RMSE Test	Cindex Train	Cindex Test
32,192	30,900	0,760	0,748
31,832	32,652	0,765	0,753
32,182	30,855	0,767	0,734
32,057	31,150	0,762	0,739
31,827	32,696	0,767	0,746
31,846	32,285	0,768	0,747
31,968	31,767	0,766	0,741
31,670	32,806	0,763	0,728
32,210	30,985	0,770	0,731
32,453	29,569	0,765	0,756
32,024	**31,566**	**0,765**	**0,742**

Figure 3. *Performance of the model*

ACKNOWLEDGEMENT

This work was partially supported by National Institute of General Medical Sciences (5U54GM104944), and National Cancer Institute (1U01CA274573-01A1). We would like to thank Juli Petereit, Director of Nevada Bioinformatics Center, for help and support throughout the project. We thank Marianne Berwick, Cheryl Jorcyk, and Tyler Bland for their insightful discussion.