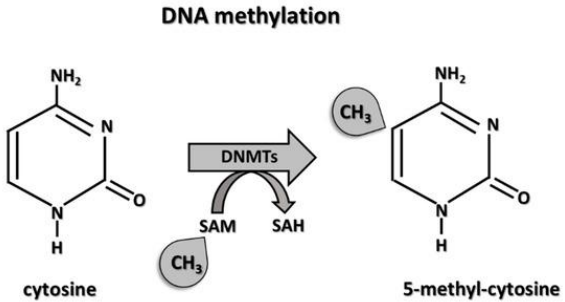
DNA Methylation as a Prognostic Indicator of Liver Hepatocellular Carcinoma (LIHC)

Harinarayana Mellacheruvu and Tony Zhang

Background

- **Clinicopathological information**: age, gender, cancer stage…
    - Usually used for prognosis, but does not provide enough details about the cancer
- **DNA Methylation**
    - A biologically-significant and reliable biomarker which is promising for prognosis



(https://www.mdpi.com/cells/cells-08-00953/article_deploy/html/images/cells-08-00953-g001-550.jpg)

Hu, R., et al. J. Comput. Biol. (2022)

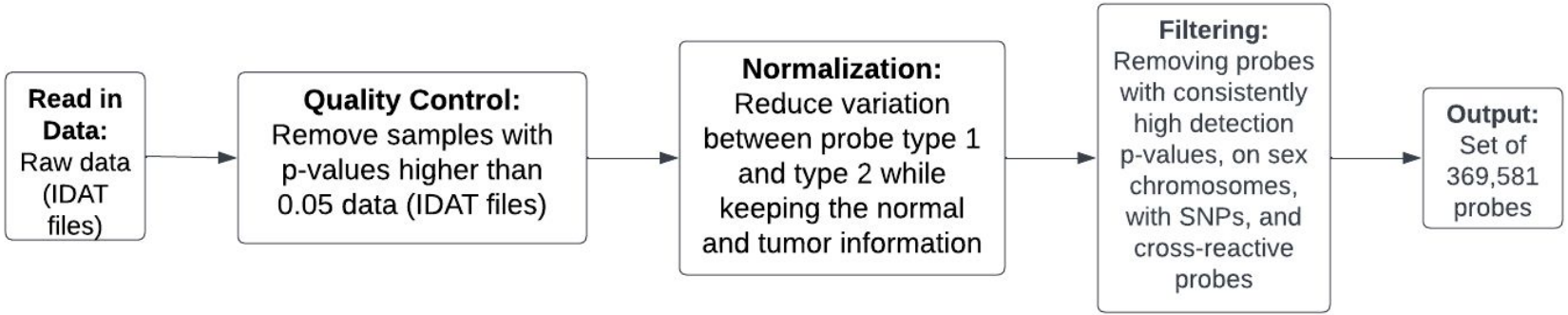
Challenges

- **Survival analysis** estimates the survival time of a patient

- Challenges with survival analysis on **high-dimensional data**
  - Censorship of survival data
    - Censoring rate of TCGA-LIHC data: 35% (132 alive, 245 dead)
  - High-dimensional data but **small sample size**
    - Dimension of data: 377 samples by 369,561 probes
    - Models likely to suffer from overfitting

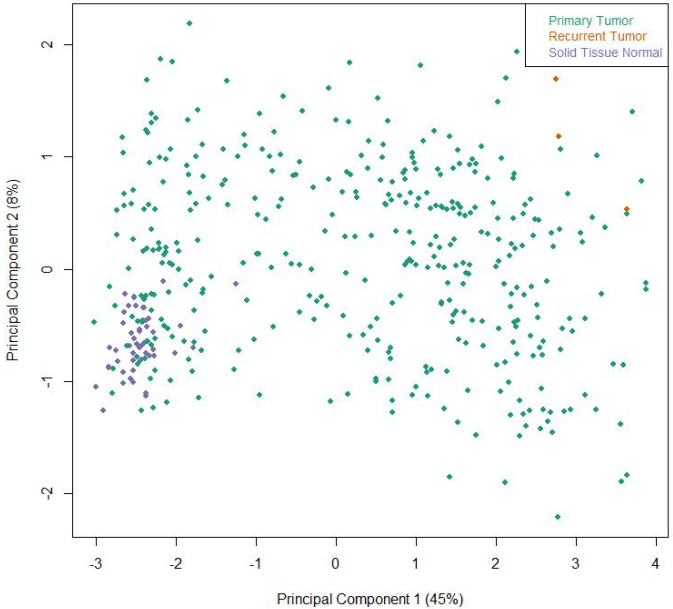
Hu, R., et al. J. Comput. Biol. (2022)

Methods: Data Preprocessing



Methods: Identifying Differentially Methylated Positions (DMPs)

1) Comparing m-values of 50 pairs of tumor and normal samples
2) Using the dmpFinder function to find DMPs
3) Keeping probes with FDR corrected p-values<0.05
4) Results: 213,392 DMPs

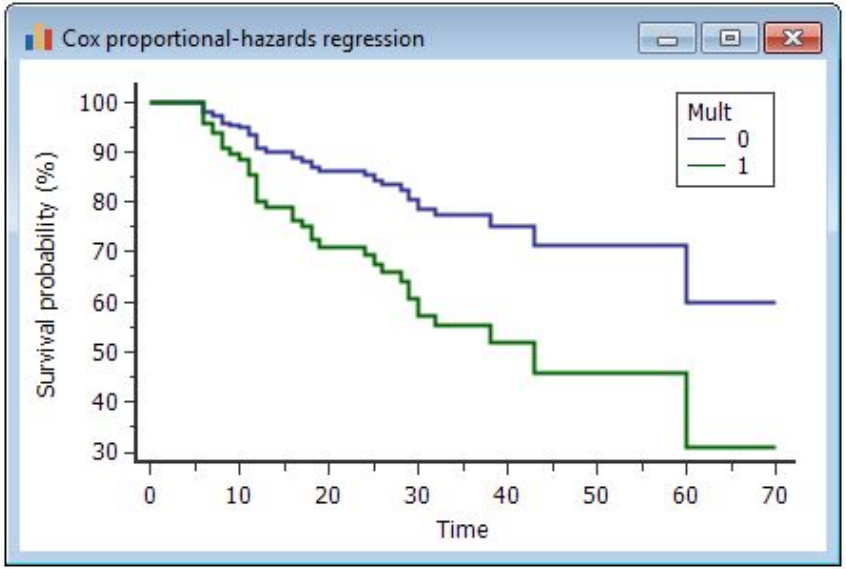


PCA plot of 377 tumor (green) and 50 normal (purple) samples

Methods: Marker Selection

- Step 1: Build a univariate **cox model** to select a smaller subset of probes to do further testing with
- Step 2: Build a multivariate **cox model** to select a significant combination of specific methylated CpG sites
- Elastic Penalty because many covariates (CpG sites)

- Data: **M Values** at specific CpG sites for patients
  - M values = log(**Methylated Signal / Unmethylated signal**)



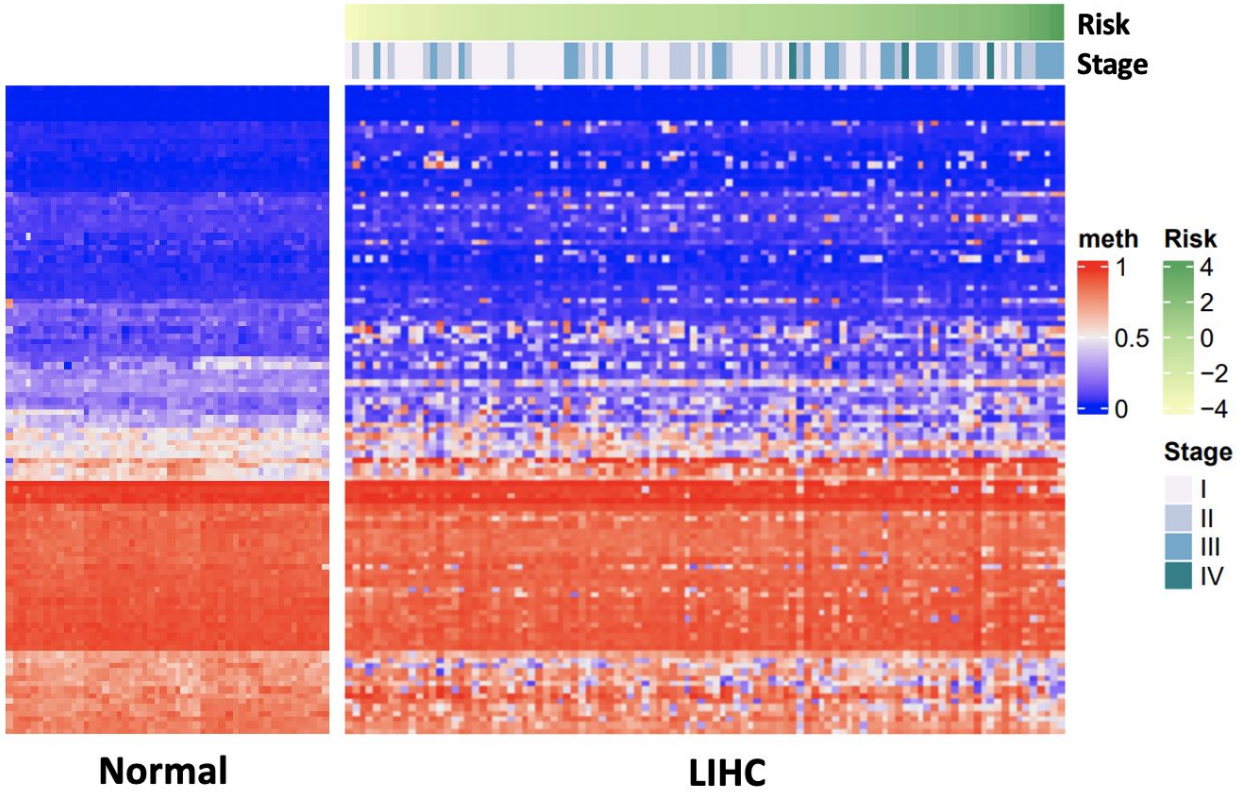
https://www.medcalc.org/manual/cox-regression.php

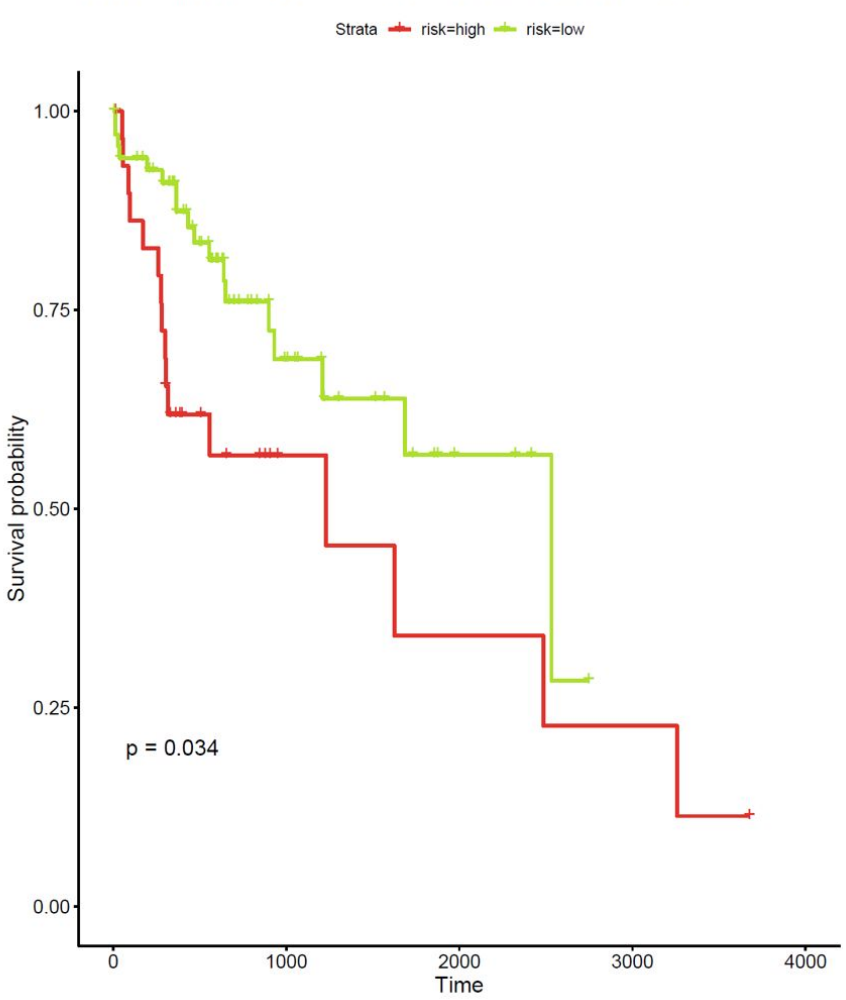
Methods: Model Training and Validation

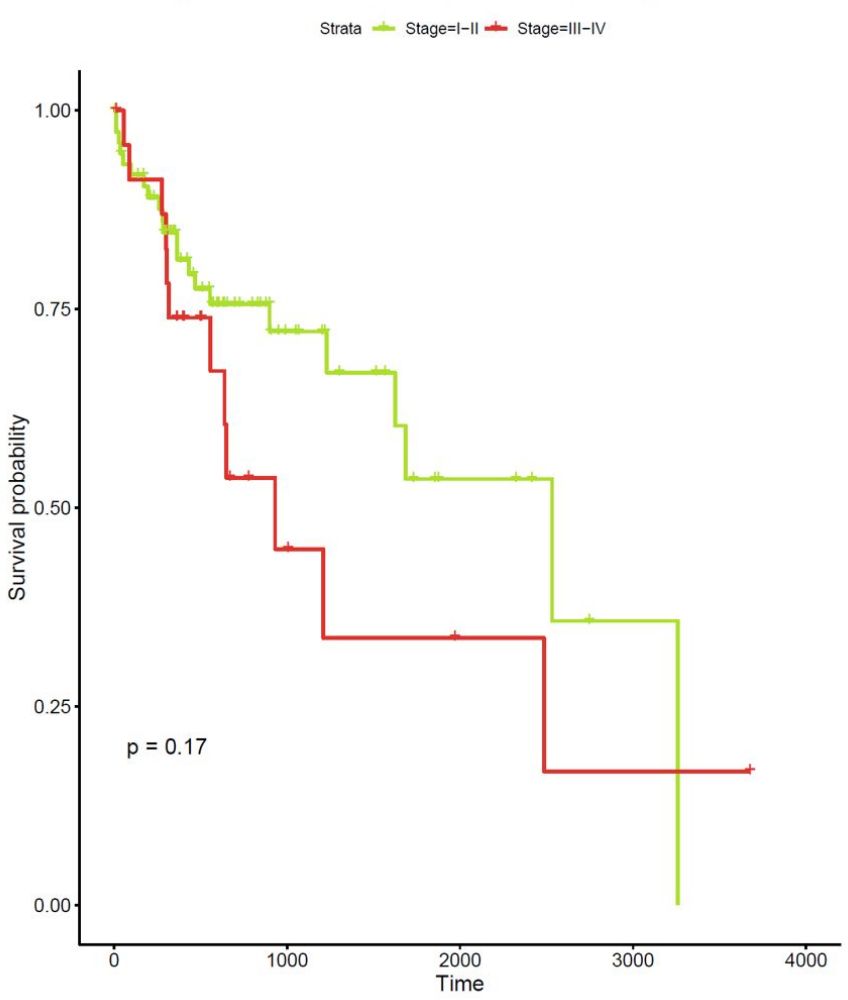
- Performed ⅔-⅓ training test split
- 5-fold cross-validation done to select best model
- Ultimately, 4288 probes were selected from multivariate cox hazard model

Results: DMPs in normal and tumor tissues

- Methylation measured using beta values
- Risk scores assigned to samples based on Cox hazard model
- Stratify LIHC patients into high and low risk based on methylation signal



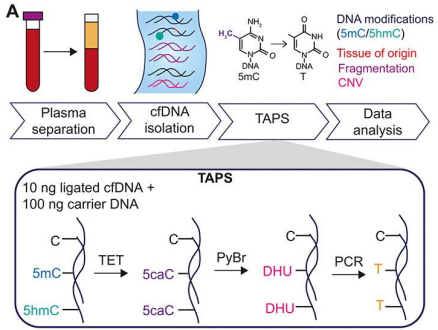

Strata
risk=high
risk=low
1.00
0.75
0.50
0.25
0.00
Survival probability
p = 0.034
0
1000
2000
3000
4000
Time


Strata
Stage=I−II
Stage=III−IV
1.00
0.75
0.50
0.25
0.00
Survival probability
p = 0.17
0
1000
2000
3000
4000
Time

Discussion

- **Takeaways**:
- Looked at differentially methylated CpG sites as a prognostic indicator in liver cancer
- Use methylation signal, specifically M values, we can stratify patients into high and low risk categories
- Methylation vs Cancer Stage

- **Next Steps**:
- Analyze methylation signals as prognostic indicator using cfDNA
- Validation of model on independent data set



https://www.science.org/doi/10.1126/sciadv.abh0534

Acknowledgements

PI: Jasmine Zhou

Direct Mentor: Ran Hu

The Zhou Lab

UCLA David Geffen School of Medicine