Using spatial omics and multiplexed imaging to discover new biomarkers of response or resistance to Immune Checkpoint Inhibitors (ICI) in Advanced Non-Small Cell Lung Cancer (NSCLC) using Lasso logistic regression

Sara Hussin[1] , Myrto Moutafi [2] , Sandra Martinez-Morilla [3] , Rolando Garcia Milian [2] , Thazin New Aung [2] , Ioannis Vathiotis [2] , Niki Gavrielatou [2] , Vasiliki Xirou [2] , David Rimm[2] , Leonidas Salichos[1,4]

[1]Department of Biological and Chemical Sciences, New York Institute of Technology, [2]Department of Pathology, Yale School of Medicine, New Haven, CT, USA , [3]Medical Library Bioinformatics Support Program, and Environmental Health Sciences Department, Yale School of Public Health, New Haven, CT, USA., [4]Department of Molecular Biophysics and Biochemistry, Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA

Abstract

Introduction

This study describes the use of spatial transcriptomics using the GeoMx Digital Spatial Profiler (DSP) as a discovery platform to find biomarkers for ICI that're associated with response or resistance to immunotherapy.

Methodology

Using spatial transcriptomics, we collected 224 pre-treated lung cancer tissue samples from 56 patients and run a panel of 18000 mRNAs, assessed by oligonucleotide-tagged in situ hybridization. The human whole transcriptome was sequenced on the NovaSeq platform to quantify the mRNAs present in each region of interest. Three tissue compartments, defined by fluorescence co-localization (tumor [panCK+], leukocytes [CD45+/CD68-], macrophages [CD68+] were generated to assess mRNA and were sampled 4 times (4 blocks, per compartment, per patient). Then, we used R to perform a Lasso logistic regression to generate a set of mRNA biomarkers associated with response or resistance to immunotherapy. Our models were trained using blocks 1,2 and 4 for each compartment and were validated using cross validation (CV) and block 3 (b3, which showed the highest heterogeneity).

Results

Our first model, using information from panCK+ (tumor cells) had a CV AUC of 0.905 and a block 3 AUC of 0.825, sensitivity of 0.885, and specificity of 0.615. This model included 65 mRNAs with *CDK18*, *GOLGA2*, *ERV3-1*, and *BLVRA* having the highest coefficients. Our second -mixed- model, derived from all 3 compartments (panCK, CD45 and CD68), had a CV AUC of 0.952, a b3 AUC of 0.767, sensitivity of 0.667, and specificity of 1. This model had 17 mRNAs from CK, 7 mRNAs from CD45, and 11 mRNAs from CD68 with *PDHB* (CK), *BLVRA* (CK), *ANPEP* (CK), *PPP1R10* (CD45), *WDR13* (CD68), and *COPS9* (CD68) having the highest coefficients.

Methodology

Tissue Microarray

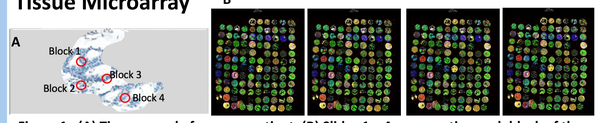

Figure 1. (A) Tissue sample from one patient. (B) Slides 1 – 4 representing each block of the tissue sample for 56 patients.

Figure 2. Workflow (A-B) 4 FFPE TMA were imaged and stained for RNA. (C) Expression levels of mRNA were counted on the NovaSeq platform.

Statistical Analysis

L1 Regularization Penalizes the sum of absolute values of the coefficients, performing feature (i.e., gene) selection by reducing those coefficients to 0.

L2 Regularization Penalizes the sum of squares of the coefficients, thus, doesn't perform feature selection as it only reduces the coefficients to values *near* 0.

Elastic Net Combines L1 and L2 penalties to reduce the number of genes being used in the model using the alpha (α) parameter, where α=1 is a lasso model while α=0 is a ridge model.

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(y_i-\beta_0-x_i^T\beta\right)^2+\lambda P_\alpha(\beta)\right),$$

where

$$P_\alpha(\beta)=\frac{(1-\alpha)}{2}\|\beta\|_2^2+\alpha\|\beta\|_1=\sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2+\alpha|\beta_j|\right)$$

Model Training Blocks 1, 2, and 4 were used for model training.

Model Validation Cross-validation (CV) and block 3 (b3) were used to validate the models. CV finds the best lambda (λ) while preventing overfitting. Showing the highest heterogeneity from the other blocks, b3 was used as a separate validation dataset.

Results

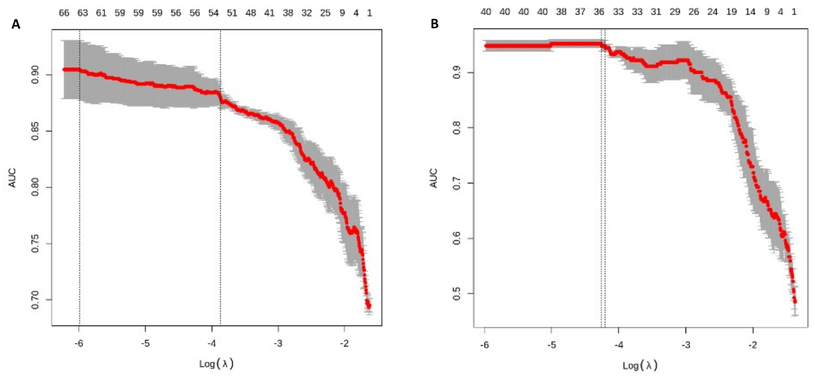

Figure 3. Performance of AUC as different λ values are being tested using CV. AUC measures how well predictions are ranked and the quality of our model's predictions on a scale of 0 (0% accurate predictions) to 1 (100% accurate predictions). AUC is significantly high for both the CK model (A) and the mixed compartments model (B). We tested 1,000 λ values using CV. The vertical dotted lines represent the λ value that gives the minimum mean cross-validated error and the λ value that gives the most regularized model such that the CV error is within one standard error of the minimum. The best λ for our models were 0.0025 (A) and 0.0141 (B).

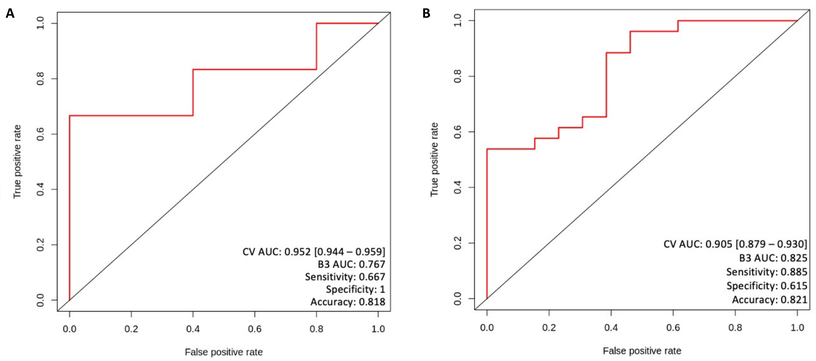

Figure 4. ROC curves for the CK model (A) and the mixed compartments model (B) measuring the true positive rate (TPR) against the false positive rate (FPR). The closer the curve is to the top left corner, the more accurate the model. The diagonal line represents a completely random classifier. AUCs (in addition to their confidence intervals), sensitivity, specificity, and overall accuracy are high across both models (based on a scale of 0 – 1). Sensitivity is the TPR while specificity is the true negative rate (TNR). High sensitivity and specificity indicate that our models can accurately classify responders vs. non-responders.

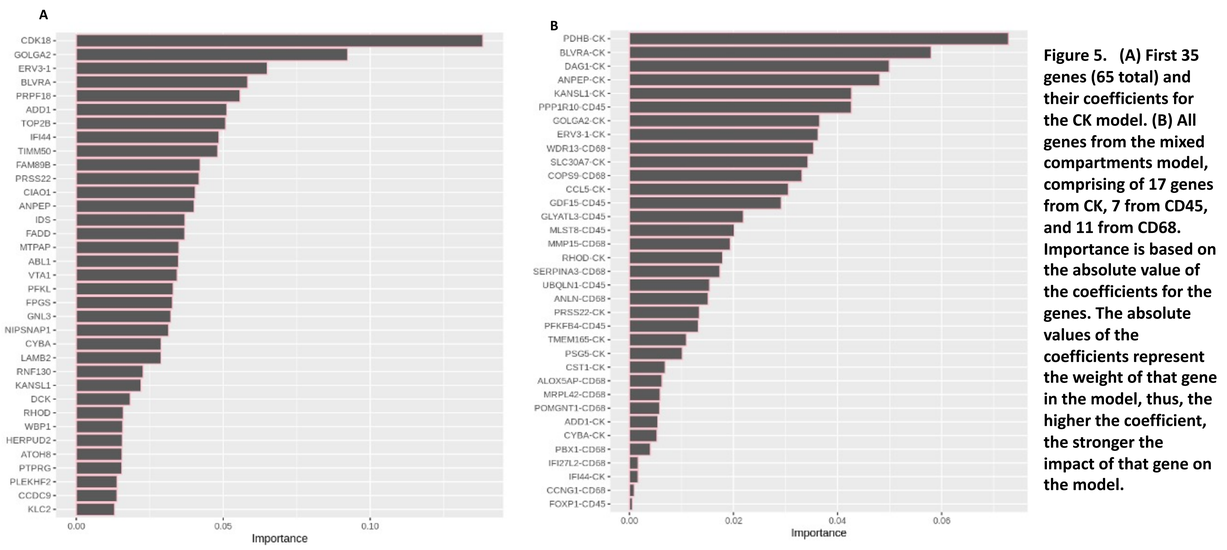

Figure 5. (A) First 35 genes (65 total) and their coefficients for the CK model. (B) All genes from the mixed compartments model, comprising of 17 genes from CK, 7 from CD45, and 11 from CD68. Importance is based on the absolute value of the coefficients for the genes. The absolute values of the coefficients represent the weight of that gene in the model, thus, the higher the coefficient, the stronger the impact of that gene on the model.

Genes

Gene	Compartment	Biological Role	Cancer Relationship
PDHB	CK	Encodes for E1 beta	Contributes to drug resistance in lung cancer cells[1]
BLVRA	CK	Key component in the conversion of biliverdin to bilirubin	Promotes colorectal cancer cell progression[2]
CDK18	CK	Plays a role in ATP binding and during the G1/S phases of mitosis	Its protein levels predict breast cancer disease progression and response to therapy[3]
DAG1	CK	Encodes for dystroglycan	Had predictive value in a previous model for lung cancer immunotherapy[4]
PPP1R10	CD45	Encodes for a phosphatase 1 binding protein	Observed to be elevated in expression in breast tumors[5]
GOLGA2	CK	Maintains the structure of the Golgi apparatus	Its downregulation may be a potential therapeutic option for lung cancer[6]
WDR13	CD68	Regulates AP1 target genes in the colon	Its absence resulted in reduced tumors in mice[7]
ADD1	CK	Encodes for the cytoskeletal protein, α-adducin	Involved in the pathogenesis of coronary artery disease and hypertension[8]
TOP2B	CK	Encodes as DNA topoisomerase	Its expression was higher in lung cancer tumors[9]
ERV3-1	CK	Mediate processes during infection	Overexpression observed in many cancers and diseases[10]
IFI44	CK	Negatively regulates host antiviral response and autoimmunity	Observed to be abnormally expressed in HNSC[11]
TIMM50	CK	Involved with mitochondrial target signaling	Promotes tumor progression in NSCLC[12]
COPS9	CD68	Involved in phosphorylation and cell proliferation	Its knockdown suppressed tumor cell growth and promoted apoptosis in NSCLC cells[13]

Table 1. Many of the most significant genes found in our models play important biological roles and have been observed in numerous cancers, including lung, breast, and colorectal cancers.

Conclusion

- Using information from different blocks within a tissue sample, we can investigate tumor heterogeneity and discover biomarkers that have unique, molecularly-defined compartments for tumor cells, lymphocytes, and macrophages.
- In the CK model, *CDK18* and *BLVRA* had very high coefficients along with *PDHB* and *DAG1* in the mixed compartments model.
- Results from this study may lead to a novel, spatially-defined, transcriptomic approach for developing new biomarkers for immunotherapy.

Contact Information

Sara Hussin
Dr. Leonidas Salichos
Department of Biological and Chemical Sciences
New York Institute of Technology
shussin@nyit.edu
lsalicho@nyit.edu

References

1. Cevatemre, Buse et al. "Pyruvate Dehydrogenase Contributes To Drug Resistance Of Lung Cancer Cells Through Epithelial Mesenchymal Transition". *Frontiers In Cell And Developmental Biology*, vol 9, 2022. *Frontiers Media SA*, https://doi.org/10.3389/fcell.2021.738916.
2. Mao, Haiyan et al. "<P≫Biliverdin Reductase A (BLVRA) Promotes Colorectal Cancer Cell Progression By Activating The Wnt/B-Catenin Signaling Pathway</P≫". *Cancer Management And Research*, Volume 12, 2020, pp. 2697-2709. *Informa UK Limited*, https://doi.org/10.2147/cmar.s242531.
3. Barone, Giancarlo et al. "The Relationship Of CDK18 Expression In Breast Cancer To Clinicopathological Parameters And Therapeutic Response". *Oncotarget*, vol 9, no. 50, 2018, pp. 29508-29524. *Impact Journals, LLC*, https://doi.org/10.18632/oncotarget.25686.
4. Guo, Nancy Lan et al. "A Predictive 7-Gene Assay And Prognostic Protein Biomarkers For Non-Small Cell Lung Cancer". *Ebiomedicine*, vol 32, 2018, pp. 102-110. *Elsevier BV*, https://doi.org/10.1016/j.ebiom.2018.05.025.
5. Grelet, Simon et al. "A Regulated PNUTS Mrna To Lncrna Splice Switch Mediates EMT And Tumour Progression". *Nature Cell Biology*, vol 19, no. 9, 2017, pp. 1105-1115. *Springer Science And Business Media LLC*, https://doi.org/10.1038/ncb3595.
6. Chang, Seung-Hee et al. "GOLGA2/GM130, Cis-Golgi Matrix Protein, Is A Novel Target Of Anticancer Gene Therapy". *Molecular Therapy*, vol 20, no. 11, 2012, pp. 2052-2063. *Elsevier BV*, https://doi.org/10.1038/mt.2012.125.
7. Singh, Vijay Pratap et al. "WD-Repeat Protein WDR13 Is A Novel Transcriptional Regulator Of C-Jun And Modulates Intestinal Homeostasis In Mice". *BMC Cancer*, vol 17, no. 1, 2017. *Springer Science And Business Media LLC*, https://doi.org/10.1186/s12885-017-3118-7.
8. Gerhard, Tobias et al. "A-Adducin Polymorphism Associated With Increased Risk Of Adverse Cardiovascular Outcomes: Results From Genetic Substudy Of The International Verapamil SR-Trandolapril Study (INVEST-GENES)". *American Heart Journal*, vol 156, no. 2, 2008, pp. 397-404. *Elsevier BV*, https://doi.org/10.1016/j.ahj.2008.03.007.
9. Syahruddin, Elisna et al. "Differential Expression Of DNA Topoisomerase Iiα And Iiβ Genes Between Small Cell And Non-Small Cell Lung Cancer". *Japanese Journal Of Cancer Research*, vol 89, no. 8, 1998, pp. 855-861. *Wiley*, https://doi.org/10.1111/j.1349-7006.1998.tb00640.x.
10. Kewitz, Stefanie, and Martin Sebastian Staege. "Expression And Regulation Of The Endogenous Retrovirus 3 In Hodgkin'S Lymphoma Cells". *Frontiers In Oncology*, vol 3, 2013. *Frontiers Media SA*, https://doi.org/10.3389/fonc.2013.00179.
11. Pan, Hua et al. "Interferon-Induced Protein 44 Correlated With Immune Infiltrzhang, Xiupeng et al. "TIMM50 Promotes Tumor Progression Via ERK Signaling And Predicts Poor Prognosis Of Non-Small Cell Lung Cancer Patients". *Molecular Carcinogenesis*, vol 58, no. 5, 2019, pp. 767-776. *Wiley*, https://doi.org/10.1002/mc.22969.
12. ation Serves As A Potential Prognostic Indicator In Head And Neck Squamous Cell Carcinoma". *Frontiers In Oncology*, vol 10, 2020. *Frontiers Media SA*, https://doi.org/10.3389/fonc.2020.557157.
13. Li, Jie et al. "CSN5/Jab1 Facilitates Non-Small Cell Lung Cancer Cell Growth Through Stabilizing Survivin". *Biochemical And Biophysical Research Communications*, vol 500, no. 2, 2018, pp. 132-138. *Elsevier BV*, https://doi.org/10.1016/j.bbrc.2018.03.183.

Acknowledgements

Many thanks to Dr. Salichos for his extensive mentorship and advice throughout this project as well as in my academic and professional career. I would also like to extend my thanks to Dr. Moutafi for the opportunity to participate in this project and for her continued support and mentorship.