Machine Learning Analysis of H&E Lung Adenocarcinoma Tumor Microenvironment Shows Association of Human Interpretable Histopathological Features with *KEAP1* Mutations

Abstract #449

BACKGROUND

Patients with advanced non-small cell lung carcinoma (NSCLC) have benefited from improved first-line treatments,[1] but ≥ 80% of patients who progress do not respond to subsequent lines of therapy, and the median progression free survival is only 2-4 months.[2] Treatment options for those patients are limited, and drugs are being developed against new targets, such as *KRAS*.

In a Phase 2 clinical trial, objective response to sotorasib (Amgen, Thousand Oaks, CA) was 37.1%, but co-occurring mutations in *KEAP1* reduced the response rate to 14% in an exploratory analysis.[2] Approximately 20% of LUAD tumors have *KEAP1* mutations.[3] KEAP1^{MUT} patients may therefore benefit from alternative treatments.

Here we present machine-learning (ML) algorithms that can predict *KEAP1*^{MUT} directly from routinely available hematoxylin and eosin (H&E) stained tissue samples. ML-based genomic assessment of the H&E samples that are routinely collected from each patient has potential to decrease the cost and turnaround time of genotyping because it requires no additional tissue collection or processing for genomic sequencing. This could increase the efficiency of clinical development program and may allow patients to receive the most effective therapies sooner.

ML MODEL TME PREDICTIONS

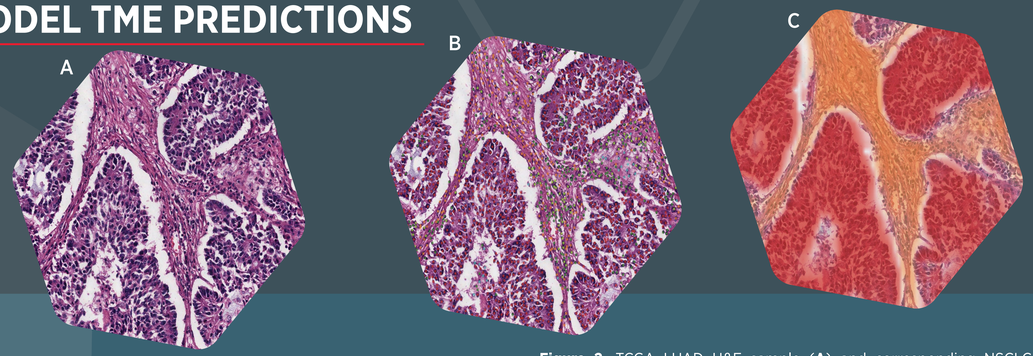


Figure 2. TCGA LUAD H&E sample (**A**) and corresponding NSCLC TME Model Overlays (**B and C**). ML-model predictions are classified into cell types (**B**) : lymphocyte (green), plasma cell (lime), fibroblast (orange), macrophage (aqua), and cancer cell (red), plus tissue types (**C**) of cancer tissue (red), cancer-associated stroma (orange), and necrosis (black).

METHODS

ML models, pre-trained to comprehensively characterize cells and tissue regions in NSCLC[4], were applied to 208 digitized whole slide images (WSI) of hematoxylin and eosin (H&E)-stained LUAD from The Cancer Genome Atlas (TCGA)[5] (**Figure 1**). Without further training, the models identified and quantified areas of cancer epithelium, cancer stroma, and necrosis, as well as counts of fibroblasts, cancer and immune cells (lymphocytes, macrophages, and plasma cells) (**Figure 2,** top right). Human Interpretable Features (HIFs), that capture the cell composition and tissue architecture of each WSI, were automatically extracted from the model generating a quantitative description of the tumor microenvironment (TME). TCGA-provided metadata indicated that 17% of the dataset (N=35) was *KEAP1*^{MUT}.

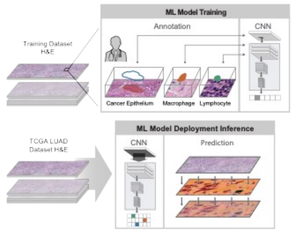


Figure 1. Machine Learning Model Training and Deployment

Associations between HIFs and *KEAP1*^{MUT} were determined by:

- Univariate analysis followed by false discovery rate (FDR) correction to investigate associations between single HIFs and *KEAP1*^{MUT}
- Hierarchical clustering using cross correlation and combining p-values for HIF groups determined by correlations between HIFs using Empirical Brown's Method determined correlations between groups of HIFs and *KEAP1*^{MUT}
- Constructing a separate logistic regression model for each HIF, with mutational status as a binary dependent variable, to determine the most statistically significant predictive features (by AUC, after Benjamini Hochberg FDR correction)

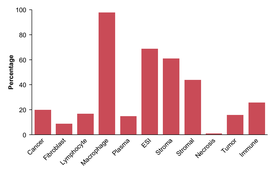
Confounding factor influence was accounted for after positive associations were identified. Independent validation of any associations was conducted using transcriptomic data (from TCGA) to correlate the expression of specific marker proteins representative of significant HIFs with *KEAP1*^{MUT}.

RESULTS

ML models generated a total of 4,443 HIFs from LUAD WSI, which were reduced to 2,684 HIFs after removal of outliers, HIFs with degenerate features, ones describing extensive properties of the WSIs (*e.g.*, absolute cell counts).

HIFs most significantly associated with *KEAP1*^{MUT} were identified by univariate analysis (193 HIFs [p<0.05 after FDR correction]), and hierarchical clustering that identified 4 groups of HIFs that were associated with *KEAP1*^{MUT} with 211-264 HIFs per group (p<0.05 after group wise FDR correction). **Table 1** shows the four cluster types and representative examples of significant HIFs from each cluster.

Figure 3. Frequency of Cell and Tissue Features in 100 Significant HIFs



Focusing on 100 HIFs most significantly-associated with *KEAP1*^{MUT}, as determined by logistic regression, the frequency of occurrence of cell and tissue features was assessed revealing that the term "macrophage" was a component of most HIFs (**Figure 3**). Thus, a relationship between macrophages and other cell or tissue features may be strongly predictive of *KEAP1*^{MUT}.

Table 1: HIF Clusters and Examples of Significant HIFs

HIF Cluster Type	Example Significant HIF from Cluster
Group 1: Density of immune cells in stroma relative to cancer tissue	Density of macrophages in the tissue
Group 2: Proportion of macrophages to immune cells in tumor	Within the stroma, the total number of immune cells divided by the total number of macrophages
Group 3: Proportion of immune cells to macrophages in cancer	Within the tumor area, the total number of immune cells (all types including macrophages) within a 40μm radius of a macrophage cell, over the total number immune cells
Group 4: Proportion of macrophages to cancer cells in tumor	Within an area defined as 60μm into the epithelial region from the interface of the epithelium and stroma regions, the total number of all cells (all types) divided by the total number of macrophages

RESULTS

To further explore the significance of macrophages as predictors of *KEAP1*^{MUT}, we compared the density of macrophages within the region of tumor tissue, and the expression of CD14 mRNA, a macrophage marker, in *KEAP1*^{MUT} and *KEAP1* wild type respectively (**Figure 4 and 5**). There were significantly fewer macrophages within the tumor region in *KEAP1*^{MUT} (p=0.001), and correspondingly expression of CD14 was reduced.

Figure 4. Density of Macrophages in LUAD Tumor Area

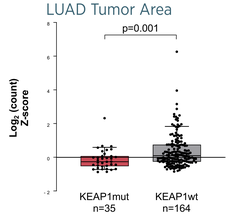
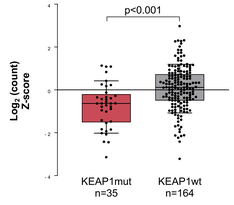


Figure 5. CD14 mRNA Expression



Finally, the predictive power of *KEAP1*^{MUT}-associated HIFs was determined by AUC. **Table 2** shows that two HIFs describing the proportion of macrophages to other TME features are highly predictive of *KEAP1*^{MUT}.

Table 2: Prediction of *KEAP1*^{MUT} with Example HIFs

HIF	AUC
Proportion of the Count of Macrophage Cells within 40μm of a Lymphocyte cell over the Count of Immune Cells within 40μm of a Lymphocyte Cell	0.89
Proportion of the Count of Macrophage Cells within 40μm of a Plasma cell over the Count of Stromal Cells within 40μm of a Plasma Cell within 60μm of the ESI	0.89

CONCLUSIONS

In TCGA LUAD WSI, ML-quantified TME histological features generated 193 HIFs that correlate with *KEAP1*^{MUT}, with HIFs describing macrophages in the TME being the most significant. These exploratory results highlight the potential for ML models to predict *KEAP1* status, and other mutations, from H&E-stained resections, to better support matching patients with the most effective therapies.

AUTHORS

Robert Egger,[1] Martin Ieong,[2] Murray Resnick,[1] Amaro Taylor-Weiner,[1] Victoria Mountain,[1] Ilan Wapinski,[1] Michael Montalto,[1] Andrew Beck,[1] Josie Hayes,[2*] Benjamin Glass[1*]

[1]PathAI, Boston, MA;
[2] Revolution Medicines, Redwood City, CA

*contact email: ben.glass@pathai.com and josie@revmed.com

REFERENCES

1. Howlader N, Forjaz G, Mooradian MJ, et al. The effect of advances in lung-cancer treatment on population mortality. N Engl J Med 2020;383:640-9.
2. Skoulidis F, Li BT, Dy GK, Price TJ, Falchook GS, Wolf J, Italiano A, Schuler M, Borghaei H, Barlesi F, Kato T, Curioni-Fontecedro A, Sacher A, Spira A, Ramalingam SS, Takahashi T, Besse B, Anderson A, Ang A, Tran Q, Mather O, Henary H, Ngarmchamnanrith G, Friberg G, Velcheti V, Govindan R. Sotorasib for Lung Cancers with KRAS p.G12C Mutation. N Engl J Med. 2021 Jun 24;384(25):2371-2381. Campbell, J., Alexandrov, A., Kim, J. et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet 48, 607–616 (2016).
3. Diao, J.A., Wang, J.K., Chui, W.F. et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. Nat Commun 12, 1613 (2021)
4. Setoguchi M, Nasu N, Yoshida S, Higuchi Y, Akizuki S, Yamamoto S (July 1989). "Mouse and human CD14 (myeloid cell-specific leucine-rich glycoprotein) primary structure deduced from cDNA clones". Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression. 1008 (2): 213–22. doi:10.1016/0167-4781(80)90012-3.
5. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550 (2014)

ACKNOWLEDGMENTS

The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.
We thank Biosciences Communications for their expert assistance with data visualization. https://www.bioscicom.net/
This poster template was developed by SciStories LLC. https://scistories.com/

PathAI