Leveraging scale in precision oncology to measure pathway activation and detect genomic drivers in a large, real-world pan-cancer cohort

Joshua Drews[1*], Lee Langer[1*], Ezgi Karaesmen[1], Matthew MacKay[1], Andrew Sedgewick[1], Joshua SK Bell[1], Justin Guinney[1], Catherine Igartua[1]

[1]Tempus Labs, Chicago, IL

*These authors contributed equally to this work

INTRODUCTION

Tumorigenesis and cancer progression are driven by the dysregulation of oncogenic signaling pathways. The assessment of pathway disruption is clinically relevant given the increasing number of targeted therapies designed to modulate these signaling cascades, and the precise assessment of the degree of pathway dysregulation can inform variant classification and facilitate the discovery of novel biological drivers.

Here, we developed a machine learning platform integrating DNA alterations and RNA expression data from a large, pan-cancer cohort to develop RNA models of oncogenic signaling pathway dysregulation in near real time. This platform performs strongly across multiple cancer types, closely recapitulates published *in vitro* data, and identifies potential genetic alterations that may cause pathway disruption.

METHODS

- Real-world, de-identified patient records were selected from the Tempus Database (N=15,217; 22 cancer types). Gene expression was measured by whole-exome capture RNA sequencing applied to FFPE samples.
- Models were trained using gene expression to predict pathway dysregulation. Among samples within the positive and negative groups, 80% were used for training, 10% were used to determine the optimal cutoff, and the final 10% were used as a hold-out set to assess performance. The number of samples classified as positive or negative were balanced between each of these sets. Model performance was measured by area under the curve (AUC) and evaluated for each individual cancer type using a hold-out dataset.
- Pan-cancer models were built for each of the 10 TCGA pathways (**Figure 1**). The positive cohort consisted of samples with ≥ 1 short variant or copy number variation (CNV) in the pathway gene list (Sanchez-Vega, et al. 2018). The wild-type (WT) group consisted of samples with no somatic, pathogenic, or variant of unknown significance (VUS) alterations, no pathogenic CNVs, and no fusions in the pathway. All samples were high purity and had matched normal samples.
- A RAS model was built to assess the pathogenicity of uncommon *RAS* variants (**Figure 2**) in colorectal cancer (CRC). The positive cohort consisted of samples with a common *RAS* mutation, and the WT samples were WT for all *RAS*/*MEK*/*ERK* genes. Samples with uncommon *RAS* mutations were removed from the positive cohort in order to be tested in an unbiased manner.
- A HIPPO model was built in high-grade glioma to detect novel drivers of HIPPO activation (**Figure 3**). The cohort definitions were similar to those used in Figure 1.

RESULTS

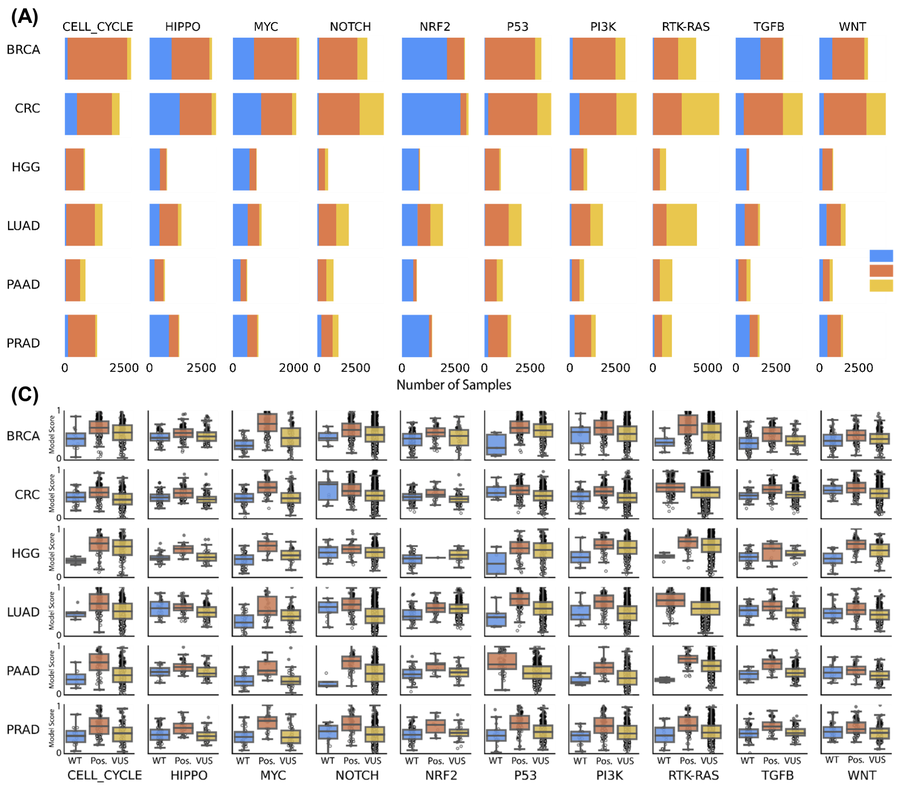

Figure 1. Characterization of Model Performance Across Common Cancers. Pan-cancer transcriptome models were trained across the 10 TCGA pathway gene sets. The results for 6 of the cancer cohorts are shown here. (A) The sizes of the pathway cohorts for the WT samples (blue), positive samples (orange), and those with a VUS in a pathway gene but no known pathogenic pathway mutations (gold). (B) The AUCs of the pan-cancer models in the holdout sets. (C) The distributions of model scores of the WT, positive, and VUS cohorts. Points shaded in black are above the optimal threshold calculated for the model.

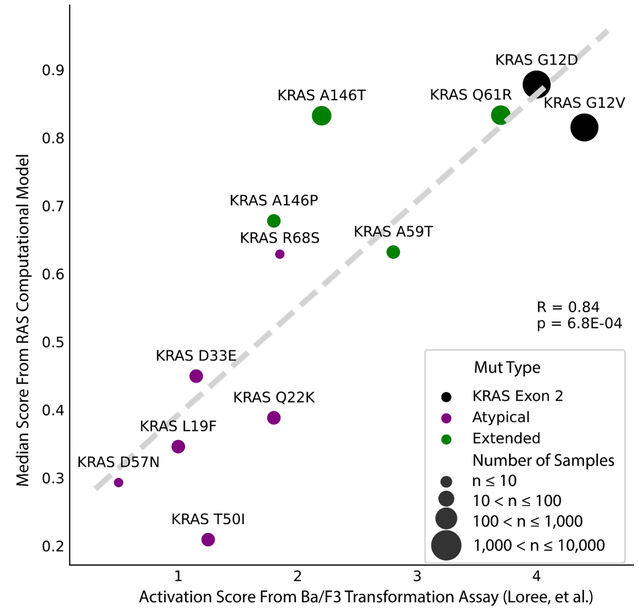

Figure 2. Prediction of *RAS* Variant Pathogenicity in Colorectal Cancer is Correlated with Functional Activation Scores. A model was trained in CRC where positive samples had detected pathogenic common *RAS* mutations, and the control was WT for all variants in *RAS/MEK/ERK* genes. Samples with uncommon *RAS* mutations that were previously assessed in vitro for pathogenicity using a Ba/F3 transformation assay (Loree, et al. 2021) were removed from the positive cohort to be tested in an unbiased manner. The median model output score for these variants was plotted against the reported pathogenicity as determined using the transformation assay.

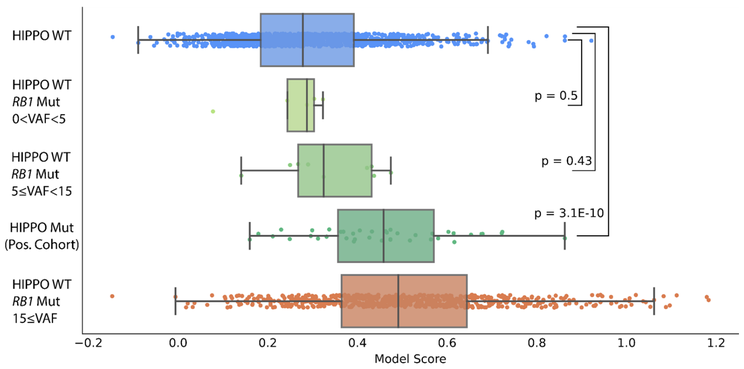

Figure 3. *RB1* is a Potential Driver of HIPPO Pathway Dysregulation in High-Grade Glioma. An integrated platform algorithm was implemented to search the space of detected genomic alterations for discovery of novel events associated with HIPPO pathway dysregulation in high-grade glioma. *RB1* pathogenic mutations (*RB1* mut) were significantly associated with high HIPPO dysregulation score, and the association increased with mutation variant fraction, suggesting that *RB1* could be a putative regulator of the HIPPO pathway.

SUMMARY

- We developed a platform for transcriptomic modeling of oncogenic pathway activation in near real time and applied it to a large real-world database.
- The platform differentiated between oncogenic drivers and passenger alterations with high accuracy, and measured the relative levels of pathway activation states between and within cancer subtypes.
- The platform was able to recommend novel regulators of the HIPPO pathway, suggesting that it could facilitate the discovery of biomarkers and the improvement of transcriptional models.

Acknowledgments: We thank Matthew Kase for poster preparation & review.

Correspondence: Katie.Igartua@Tempus.com