# Utilizing Electronic Health Records (EHR) and Tumor Panel Sequencing to Demystify Prognosis of Cancer of Unknown Primary (CUP) patients


Intae Moon[1,3], Jaclyn LoPiccolo[2], Sylvan C. Baca[2,5], Kenneth L. Kehl[3], and Alexander Gusev[3,4,5]

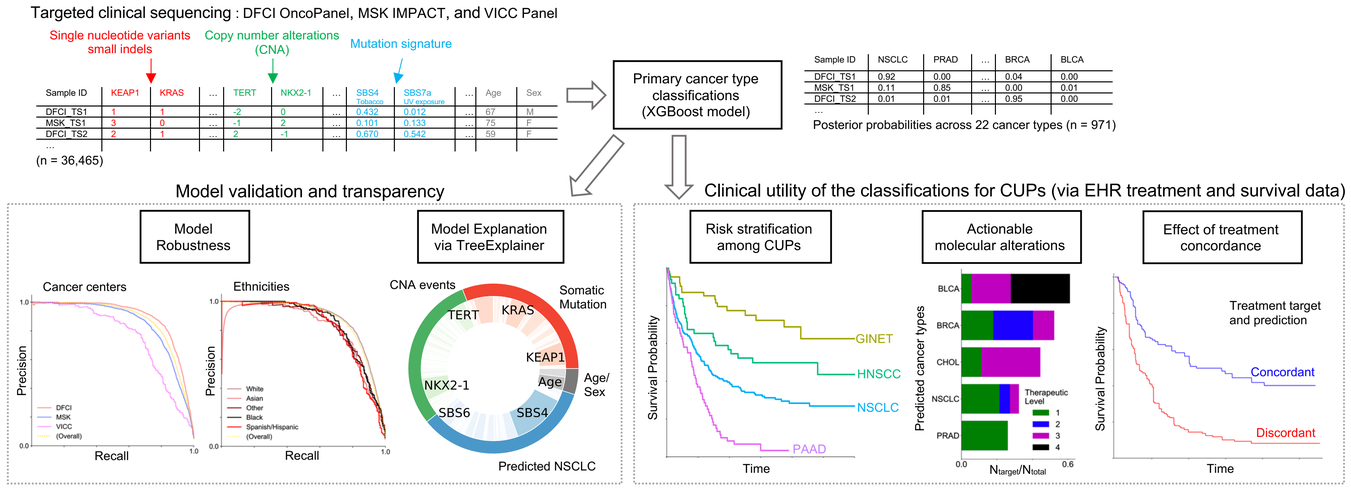
[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA
[3]Division of Population Sciences, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA
[4]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[5]The Broad Institute of MIT and Harvard, Cambridge, MA, USA


## Motivation

- CUP represents about 3-5% of all cancers worldwide and has poor prognosis (survival 6 to 16 months) & limited treatment options
- Accuracy of the pathology review suffers in difficult-to-diagnose tumors

➢ Need for an accurate method to identify the primary site of CUP to empower informed clinical decision making

## Method overview




## Key findings

- Obtained interpretable classifier with robust performance on held-out test data consisting of 7,289 primary and metastatic tumor samples from 22 known cancer types (weighted F1 : 0.789) (Fig. 1)
- Applied the classifier to 971 patients with CUP at the Dana-Farber Cancer Institute (DFCI) to predict their primary cancer types (Fig. 2)
- Identified primary based subtypes among the CUP cohort with significant prognostic differences (Fig. 4)
- Identified patients with CUP who have actionable alterations based on their predicted cancer type (Fig. 5)
- Demonstrated that *the first palliative therapies consistent with the molecular classification led to longer survival* than those discordant with the molecular classification (Fig. 6 and 7)

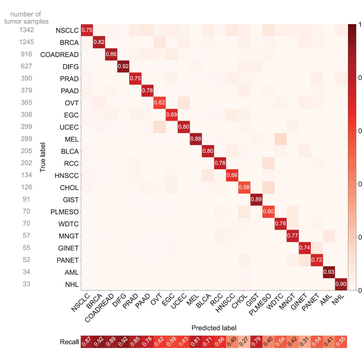



Fig. 1 model performance on the test set

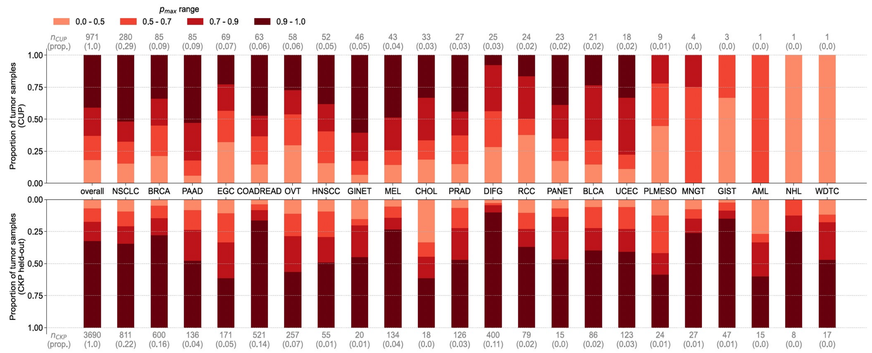



Fig. 2 proportion of tumor samples wrt. $p_{max}$, posterior probability for each prediction in CUP and Cancer with Known Primary (CKP)

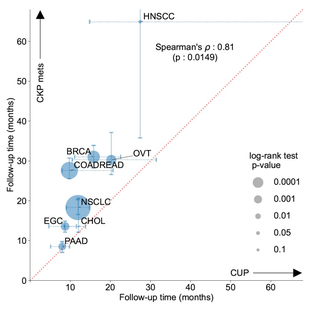



Fig. 3 median survival time: CUP vs. CKP

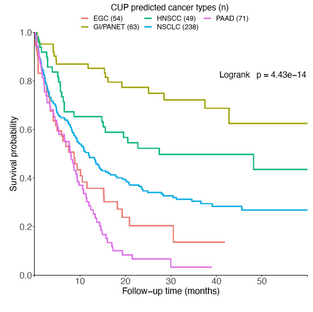



Fig. 4 : Risk stratification of CUPs based on their predicted cancer types

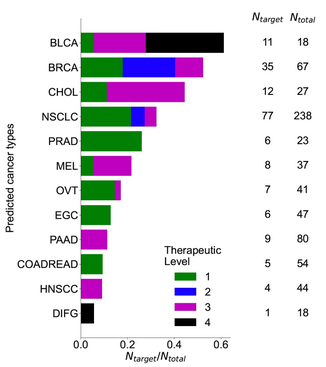



Fig. 5 : Proportions of patients with actionable targets

## Discussion

Our findings suggest that many CUPs can be classified into meaningful subtypes with somatic and prognostic differences, and this classification has the potential to aid clinical decision making.

*Future work*

- Validation of the analysis on a larger CUP cohort including multiple institutions
- Incorporate the model as an assistive tool for cancer diagnosis workflow

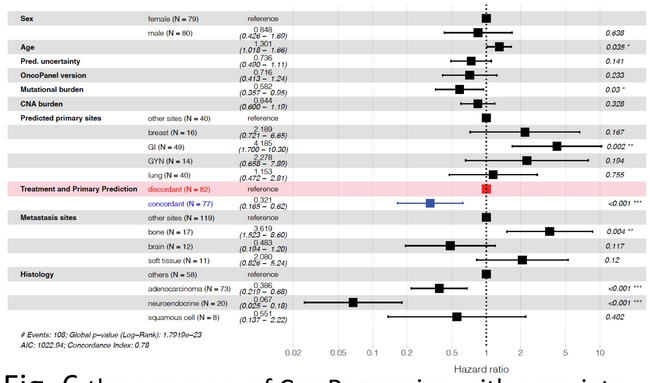



Fig. 6 the summary of Cox Regression with covariate adjustment on patients in the CUP cohort

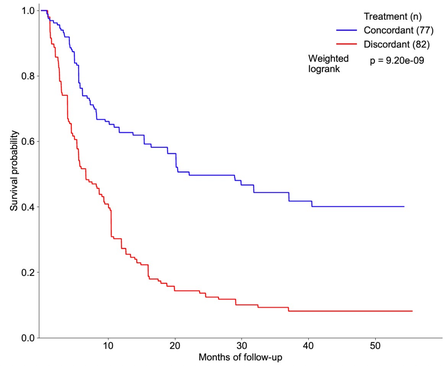



Fig. 7 Kaplan-Meier estimator with inverse probability of treatment weighting (IPTW)