How Strong Is Your Driver?

Machine Learning and Structural Features Predict Driver Mutation Strength and Clinical Outcomes

Noa Klugman, Dorit Shemesh, Jakob Landau, Ora Schueler-Furman*, Shai Rosenberg*

The Wohl Institute for Translational Medicine, Sharett Institute for Oncology, Departments of Microbiology and Molecular Genetics & Quantitative Molecular Medicine, Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

*These authors jointly supervised this work.

Contact: noa.klugman@mail.huji.ac.il

The Problem: Beyond Binary Classification

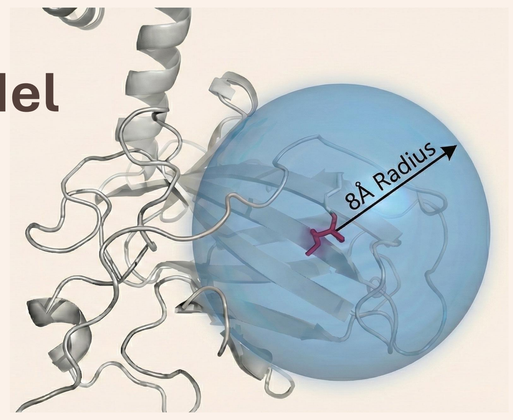
- Traditional classification dichotomously splits mutations into "Drivers" or "Passengers".
- This oversimplifies biology: driver mutations contribute unequally to tumor progression.
- There is a critical need to prioritize mutations based on their specific oncogenic effect size.

The Solution: STRIDE (STRuctural-Informed Driver Evaluation)

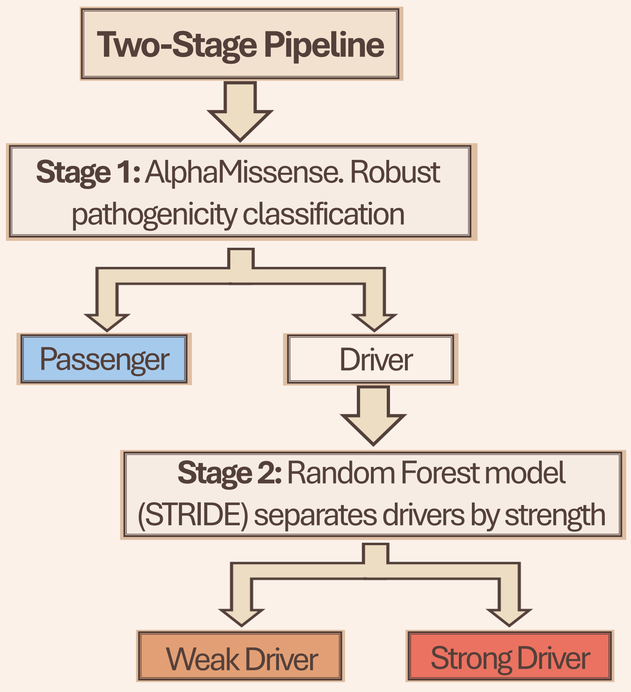
- Integrates machine learning with structural biology and evolutionary selective pressure signals.
- Classifies driver mutations into **Strong** and **Weak** categories.
- Approximates the Tumor Variant Amplitude (TVA), a selective pressure estimate with 3D residue- level context (also for positions with reduced sampled variance).

Methods

Key Features For Model Training

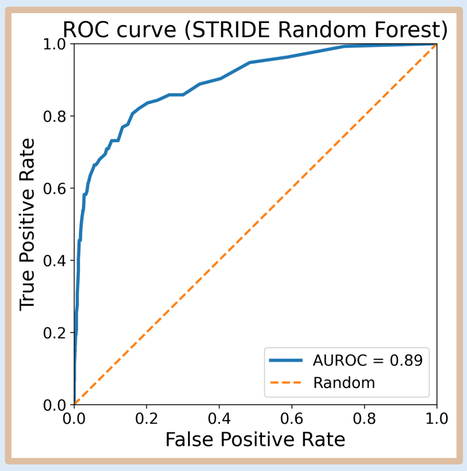


- Stability ΔΔG
- Conservation
- Functional Pockets
- Neighbourhood Features: analysing the mutational landscape within an 8Å radius to estimate local selective pressure



Model Performance

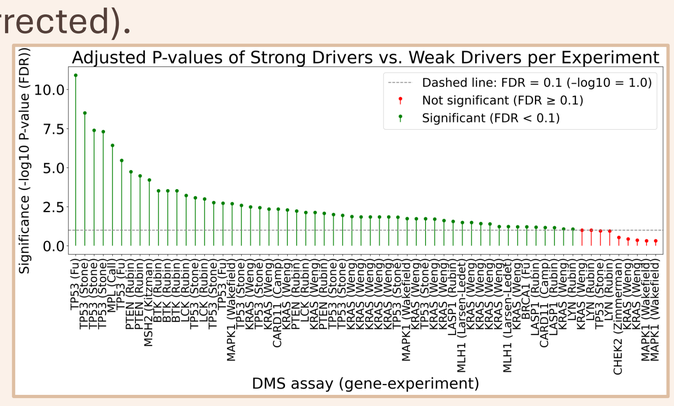
STRIDE achieves AUROC of 0.89 on held-out test data



External Validation

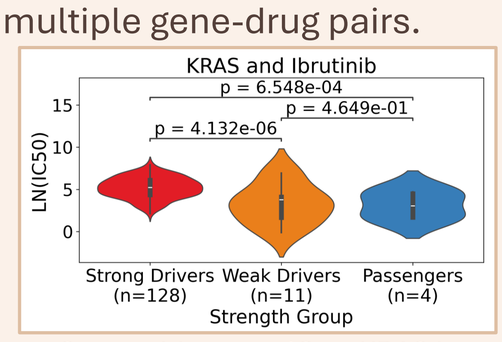
Functional

In 86% of deep mutational scanning assays, strong vs. weak drivers show significantly different functional scores (multiple testing-corrected).

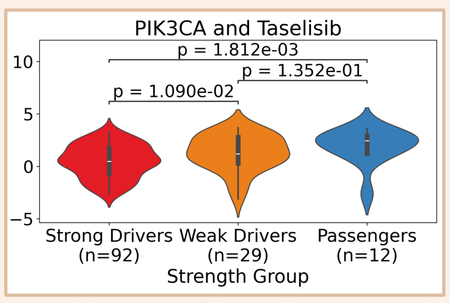


Drug Sensitivity

Driver strength predicts drug sensitivity in cancer cell lines for both direct inhibitors and upstream pathway inhibitors across multiple gene-drug pairs.



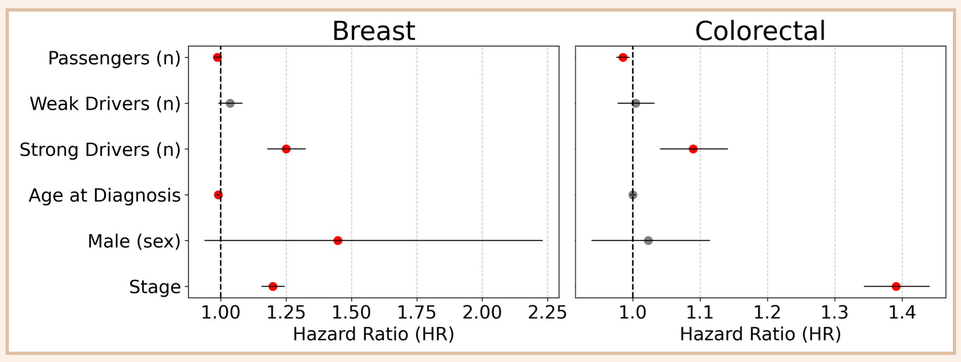
Strong drivers → ↑IC_{50} - KRAS is constitutively active and less BTK-dependent



Strong drivers → ↓IC_{50} - direct targeting, higher pathway dependence

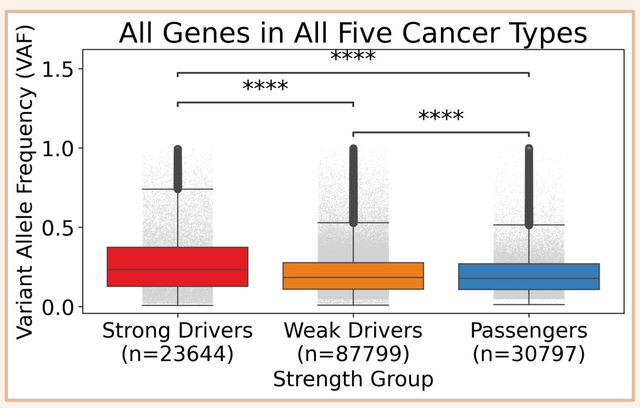
Clinical Outcomes

Strong-driver burden is associated with poorer Overall Survival in breast and colorectal cancer, based on Cox regression (also true for prostate and pancreatic).



Clonality

Strong drivers possess higher Variant Allele Frequencies (VAF), indicating early, critical events in tumor evolution.



Conclusions

- STRIDE enables a granular, multi-layered understanding of the mutational landscape beyond binary labels.
- By identifying high-impact drivers, it can improve mutation prioritization, advancing personalized medicine.

References

1. Landau, J., Tsaban, L., Yaacov, A., Ben Cohen, G. & Rosenberg, S. Shared Cancer Dataset Analysis Identifies and Predicts the Quantitative Effects of Pan-Cancer Somatic Driver Variants. Cancer Res 83, 74–88 (2023).
2. Cheng, J. et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science 381 (2023).

The Furman Lab @ HUJI

Ascola Alumni Network · הקרן הלאומית למדע · المؤسّسة الإسرائيلية للعلوم · Israel Science Foundation · THE HEBREW UNIVERSITY OF JERUSALEM · אוניברסיטה העברית בירושלים · ROSENBERG LAB