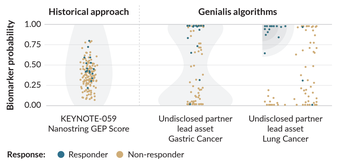
ResponderID™ KRAS:
Biology-Driven Machine Learning to Personalize KRAS Inhibitor Therapeutics

Joshua Wheeler, Anže Lovše, Klemen Žiberna, Janez Kokošar, Miha Štajdohar, Rafael Rosengarten, Aditya Pai, Daniel Pointing, Luka Ausec, Mark Uhlik Genialis, Inc., Boston MA, United States

ResponderID

Background

Traditional biomarker development methods rely on modeling drug responses directly within clinical cohorts that have outcome data. This approach presents several challenges, as clinical data is often limited, and there are rarely enough responders to support robust biomarker discovery. As a result, these biomarkers often suffer from overfitting, limited generalizability to new datasets, and poor ability to distinguish between responders and non-responders to treatment. ***Figure 1*** illustrates a comparison between the historical traditional approach and the Genialis approach for distinguishing non-responders from responders.

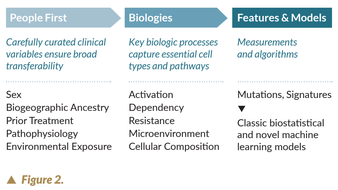


▲ ***Figure 1.***
Illustration of a traditional biomarker that directly models response using an 18-gene expression signature. Patients from the Keynote-059 study are color-coded based on their response to pembrolizumab, an immune checkpoint inhibitor. In this historical approach, it is challenging to distinguish responders from non-responders. However, with Genialis' approach, non-responders are readily separated from responders.

Biomarker Conceptualization

Genialis embraces an approach to biomarker development that identifies and abstracts the biology most likely to contribute to clinical response based upon therapeutic MoA, thereby enhancing the utility for clinicians and diagnostic organizations. By not fixating solely on outcome data for any given drug, our biomarkers can be applied broadly across a class of treatments rather than a single agent. This adaptability keeps Genialis biomarkers in sync with drug clinical development, ensuring lasting drug responses.

The implementation of ResponderID begins by defining crucial biological factors and variables (features) essential for predicting therapeutic responses. These features encompass both intrinsic properties of cancer cells (e.g., oncogene activation status, genetic dependency, resistance propensity) and extrinsic properties (e.g. tumor microenvironment, tumor composition). Additionally, a diverse range of patient clinical variables is incorporated as features to enhance model transferability and mitigate bias. These features are combined to represent potential phenotypic states of the disease, and are subsequently modeled using various machine learning techniques (***Figure 2***).



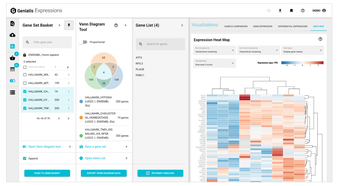
▲ ***Figure 2.***
The Genialis ResponderID approach to biomarker development prioritizes the patient, enabling the modeling of essential biology critical for predicting drug responses and prognoses. This is accomplished through the application of machine learning and proprietary algorithms.

Genialis Expressions

A cloud based software and repository for NGS data that standardizes data analysis, annotation, and visualization is essential for fostering cross functional collaboration.

Validated bioinformatic and quality control workflows ensure the uniform processing of data and assists in defining the analytical requirements for a downstream assay. FAIR data management facilitates annotation and curation, and supports reproducibility and reporting.

Genialis Expression is a technology layer that produces versioned machine learning- ready data at scale.

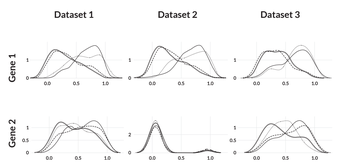


▲ ***Figure 3.***
Genialis Expressions. *A screenshot of the modules for interactive visualization of gene expression data.*

Genialis Precision Medicine (GPM)

This code base has been developed in accordance with the FDA guidance for Good Machine Learning Practice and extensively validated. GPM is used throughout the biomarker implementation lifecycle, commencing with the derivation of gene signatures that represent complex biological systems to the systematic mapping of independent datasets for real world validation. Some of the key methods that are readily deployed and dramatically improve the robustness of of a biomarker include:

- Evaluating the transferability of predictive signatures across molecular biomarker datasets **(U.S. Patent App. Ser. No: 17/154,683)** (***Figure 4***)
- System of preprocessors to harmonize disparate 'omics datasets by addressing bias and/ or batch effects **(PCT Application No. PCT/US2022/037860)**.
- Incorporating best practices for RNA-sequencing data normalization https://github.com/genialis/RNAnorm



▲ ***Figure 4.***
Feature transferability. *This method enables the identification of genes that are consistently and robustly expressed across different bias modalities including but not limited to tissue type and sequencing platform, thus broadening the applicability of and utility of the locked model. The consistency of expressions for two genes across multiple datasets is shown, Gene 1 is similarly expressed across all datasets and phenotypes, in contrast, gene 2 is not consistently expressed across all three datasets.*

Data, Signatures and Algorithms

Genialis has meticulously curated, annotated, and aligned nearly 100k unique publicly available samples, which complement a substantial collection of proprietary datasets. The Genialis data catalog is an asset that provides ready access to interoperable datasets to support the training, testing, and validation of biomarkers.

The ResponderID framework integrates both select public and proprietary biomarkers to map the biology of interest and model complex disease phenotypes. Modeling complex biology from diverse lines of evidence defines therapeutically relevant disease phenotypes, allowing customization and adaptability to biomarker development depending on the needs of the pharmaceutical and diagnostic developers.

Our long-term objective is to develop biomarkers that capture the essential cancer biology hallmarks, yielding a comprehensive model to support cancer care across clinical and therapeutic contexts. Recently, we co-developed a tumor microenvironment biomarker panel capable of delineating phenotypes defined by angiogenesis and immune biological axes, leading to improvements in therapeutic stratification (Uhlik et al., 2023). Collectively, our approach provides the necessary precision to guide real time clinical development and treatment decision making.

A Novel RNA-based Biomarker Framework for Guiding KRAS Therapeutic Development and Clinical Adoption

> ResponderID is a validated framework that enables the development and commercialization of multidimensional predictive biomarkers. This platform combines proprietary technology, data assets, and signatures with a comprehensive understanding of the underlying biology.
>
> The ResponderID framework takes a biology-first approach, yielding biomarkers that are applicable to a class of therapies rather than a single agent. Our approach increases the clinical utility for drug developers and diagnostic organizations alike to ultimately put patients and people first.
>
> Genialis is seeking like-minded partners interested in utilizing ResponderID KRAS as a novel KRAS biomarker to expedite the clinical development, approval and deployment of KRAS therapeutic modalities and eventually companion diagnostics.

Evaluation of KRAS Dependency Modeling from DepMap

Introduction

In an ever-evolving biological landscape, there is an imperative need for a biomarker that can effectively capture the inherent dynamism of complex disease states. Current biomarkers often adhere to a 'point-of-care' approach, which means that the information they provide is a static measure of a single variable (mutation status) and lacks information concerning additional aspects of contributing biologies.

To address these limitations, a biomarker was recently developed to assess KRAS dependency across diverse cancer cohorts, a pivotal metric in evaluating drug efficacy (Tsai et al., 2023). In this study, the authors introduce the K20 model, which incorporates 20 key features, including gene expression patterns and KRAS mutation status. The model achieved an AUC of 0.94 during validation, enabling accurate predictions of KRAS dependency in both mutant and wild-type cell lines. When applied to TCGA datasets, it effectively pinpointed subpopulations exhibiting heightened KRAS dependency, demonstrating its potential to elucidate real-world tumor biology.

Herein, we evaluated the performance of the K20 model. Our re-implementation demonstrates poor reproducibility of the K20 model, underscoring the continued need for a new generation of biomarkers capable of delivering on the promise of predicting patient response to KRAS therapies.

Methods

Data

- KRAS dependency scores from DEMETER2 RNAi screens (DepMap)
- RNA expression profiles and KRAS mutation status data from CCLE
- RNA expression profiles of LUAD, PAAD COAD, READ cohorts from TCGA

Models

1. **ElasticNet Regression** with optimization of alpha and L1 ratio hyper-parameters
2. **XGBoost Regression** with optimization of alpha, gamma and lambda hyper-parameters
3. **Linear Regression on one feature:** KRAS mutation
4. **Linear Regression on two features:** KRAS mutation and KRAS expression
5. **ElasticNet Regression** excluding both KRAS mutation and KRAS expression

ML Pipeline

1. Data integration in Genialis Expressions cloud platform
2. TPM normalization of gene expression profiles and log2 transformation
3. Feature filtering based on gene expression distribution in TCGA cohorts (std > 0.5 resulted in a subset of 9020 genes)
4. The KRAS mutation status was dichotomized: a "zero" was assigned for no activating mutations and a "one" for the presence of any activating mutation
5. Feature standardization by removing the mean and scaling to unit variance using a StandardScaler
6. Target variable was defined as dichotomized DEMETER2 KRAS dependency score
7. Training the regressors, tailored to the specific analytical context (re-implementation of ElasticNet alongside with XGBoost and LinearRegression models)
8. RepeatedKFold cross-validation with 10 splits and 100 repeats to evaluate model performance using the Receiver Operating Characteristic Area Under the Curve (ROC AUC) metric

Results

Two-feature model (KRAS mutation and KRAS expression) shows similar ROC AUC for predicting KRAS dependency as more complex ElasticNet and XGBoost models (***Table 1***).

Exclusion of KRAS features results in a lower ROC AUC of 0.86 but remains comparable to the model using only KRAS mutation status.

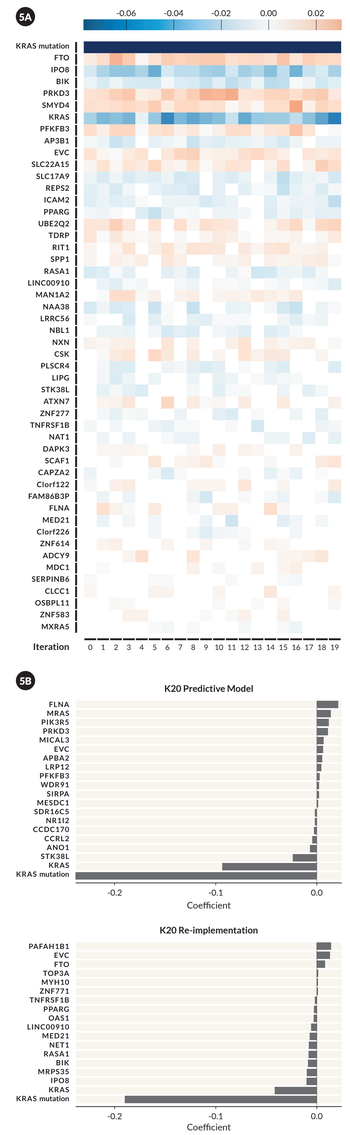
In the repeated cross-validation analysis (***Figure 5a***), the KRAS mutation feature was the sole constant selection. A handful of other features, such as FTO, IPO8, BIK, PRKD3, SMYD4, and KRAS, exhibit consistent selection, while the remaining features appear to be chosen somewhat arbitrarily (***Figure 5b***).

▼ ***Table 1.***
The ROC AUC of different models. *The ROC AUC was calculated using repeated K-fold cross-validation (n_splits = 10, n_repeats = 100).*

Model	ROC AUC
1. ElasticNet Regression with optimization of alpha and L1 ratio hyper-parameters	0.936 ± 0.047
2. XGBoost Regression with optimization of alpha, gamma and lambda hyper-parameters	0.926 ± 0.033
3. Linear Regression on 1 feature KRAS Mutation Only	0.921 ± 0.050
4. Linear Regression on 2 features KRAS Mutation & KRAS Expression	0.921 ± 0.050
5. ElasticNet Regression excluding both KRAS mutation & KRAS expression	0.858 ± 0.067

Summary

> In the absence of patient response data, current methodologies for modeling KRAS drug response heavily depend on incorporating a dependency score as a crucial variable to assess the probability of a favorable treatment response
>
> Re-implementation of dependency modeling approaches championed in Tsai et al., 2023 reveals KRAS mutation status and KRAS expression are reproducibly the two most important features in predicting KRAS dependency
>
> The two feature model (KRAS mutation and expression) have a similar ROC AUC for predicting the KRAS dependency group as more complex ElasticNet and XGboost models, suggesting additional features add little to the overall performance of the model (***Table 1***)
>
> The limited reproducibility and transferability of these models underscore the necessity for improved quality control, dataset alignment, and feature selection, with a focus on the fundamental KRAS biology — an approach being pioneered at Genialis



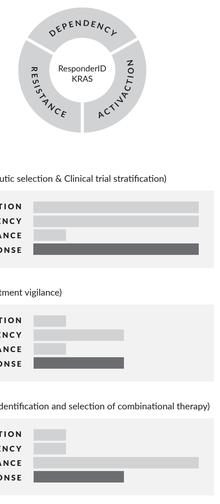
▲ ***Figure 5.***
K20 model re-implementation highlights KRAS mutation and expression status as the most predictive features. ***A.*** *The heatmap illustrates the significance of the top 50 features, selected with re-implementation of ElasticNet regression model, across multiple cross-validation iterations. Features are ranked by their frequency of selection. Different shades of color indicate varying degrees of feature importance in each iteration.* ***B.*** *Comparison of feature coefficients identified in the original study and re-implementation identifies only three shared features: KRAS mutation status, KRAS expression, EVC. Upper panel in Figure B adapted from Tsai et al., 2023*

ResponderID KRAS

Building the Next Generation of KRAS Biomarkers

Genialis' ResponderID KRAS is a RNA-based sequencing classifier rooted in KRAS biology, providing a comprehensive disease profile in one sequencing assay. The primary output predicts KRAS drug response by considering three essential biological axes: dependency, activation, and resistance. In combination, these four metrics (overall response, dependency, activation, resistance) generate a scorecard for real-time drug development and clinical decisions. Our approach is also capable of delivering insights into potential genetic targets for co-therapeutic approaches, identified from resistance gene monitoring with assigned dependency and activation scores, enhancing the potential for multi-drug combinations.

Genialis' ResponderID KRAS is poised to support partners in ongoing clinical development and across prospective validation during clinical trials.



▲ ***Figure 6.***
ResponderID deciphers essential KRAS biology, offering real-time insights into adaptive responses to guide KRAS therapeutic development and administration. ***A.*** *An ensemble of biologic axes converge to predict response.* ***B.*** *Multiple biological attributes create an adaptive biomarker panel that empowers both drug developers and clinicians to make real-time therapeutic decisions at every stage.*

References

- Tsai et al., 2023. An integrated model for predicting KRAS dependency. PLoS Computational Biology
- Uhlik et al., 2023, Xerna™ TME Panel s a machine learning-based transcriptomic biomarker designed to predict therapeutic response in multiple cancers. Frontiers in Oncology

www.genialis.com