NETRAMARK

An AI aprroach to unravelling treatment response in pancreatic cancer: Insights from the COMPASS trial leveraging large language models (LLMs)

Joseph Geraci[1,2,3,4], Bessi Qorri[1], Mike Tsay[1], Christian Cumbaa[1], Paul Leonchyk[1], Luca Pani[5,6], Larry Alphs[1]

[1]NetraMark Corp, [2]Department of Pathology and Molecular Medicine, Queen's University, [3]Centre for Biotechnology and Genomic Medicine, Medical College of Georgia, Augusta University, [4]Arthur C. Clarke Center for Human Imagination, School of Physical Sciences, University of California San Diego , [5]Miller School of Medicine, University of Miami, [6]University of Modena and Reggio Emilia

INTRODUCTION

Pancreatic cancer is often deemed the “silent killer” due to its vague symptoms, late diagnosis, and rapid progression continuing to challenge oncologists due to poor response to therapy prognosis. Pancreatic cancer is the 10th most common cancer and 3rd leading cause of cancer-associated deaths.

Standard-of-care regimens, such as FOLFIRINOX (FFX) and Gemcitabine + *nab*-paclitaxel (GnP), show varied efficacy among patients. Understanding the factors that contribute to these varied responses is crucial for advancing personalized treatment strategies.

THE OPPORTUNITY

The COMPASS trial was the first prospective evidence of the potential predictive power of molecular profiling showing that there are unique advanced pancreatic ductal adenocarcinoma (PDAC) genomic and transcriptomic subtypes with molecular heterogeneity between individuals and differing responses to chemotherapy.

OBJECTIVES

Use a novel mathematically-augmented machine learning (MAML) approach to identify causal genomic and transcriptomic factors that capture heterogeneity between individuals and differing response to chemotherapy (GnP vs FFX) to pave the way toward precision medicine.

METHODS

MATHEMATICALLY-AUGMENTED MACHINE LEARNING APPROACH

NetraAI is a novel mathematically-augmented ML approach designed to deconstruct patient populations into explainable and unexplainable subpopulations. By focusing on the explainable subpopulations, it is possible to identify variables that characterize specific subpopulations within clinical trial data related to drug response measured as best tumor response while avoiding overfitting by not forcibly explaining everyone.

LARGE LANGUAGE MODEL-POWERED NETRAGPT

Our MAML approach produces insights by identifying patient subpopulations based on gene expression variables and their expression level intervals. These subpopulations exhibit differentiated responses in tumor shrinkage between GnP and FFX. The resulting data can then be used to train Large Language Models (LLMs), enabling them to serve as clinical trial companions (NetraGPT) and improve the design of follow-on pivotal trials.

DATASET

Comprehensive Molecular Characterization of Advanced Pancreatic Ductal Adenocarcinomas (PDAC) for Better Treatment Selection: A Prospective Study (COMPASS; NCT02750657)

Study Goal: Identify predictive mutational and transcriptional features in advanced PDAC for treatment selection between GnP or FFX.

n=208 PDAC patients prior to GnP or FFX first-line chemotherapy

Primary Outcome: Tumor response characterized as:

- Stable Disease: 20% growth to 100% loss of tumor
- Partial Response: 30% to 100% loss of tumor
- Complete Response: Disappearance of tumor

Data Types:

- Whole genome sequencing (WGS)
- RNA sequencing (RNASeq)

RESULTS

NetraAI-identified subpopulations characterized by variables that differentiate between FFX and GnP response in PDAC patients

Genetic drivers differentiating FFX and GnP response in stable disease PDAC patients

- Subpopulation 1: 9 GnP responders characterized by downregulated *LRRC8E* and *UGT1A5*
- Subpopulation 2: 13 FFX and 4 GnP responders characterized by downregulated *LRRC8E*, *PTPRH*, and *SORT1*

These genes are related to neuro death and regulation, neurotrophin signaling, and neuron apoptotic processes.

Drivers differentiating FFX and GnP response in stable disease, partial response and complete response PDAC patients.

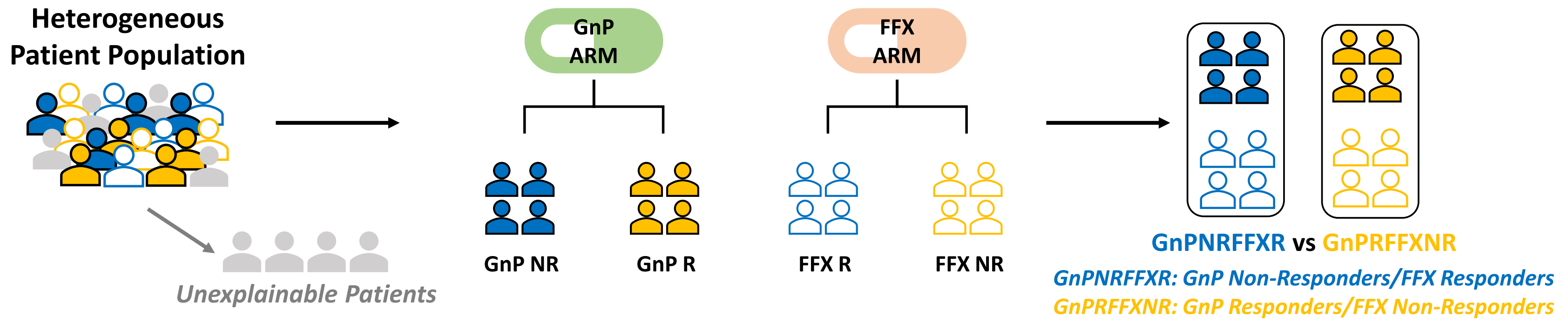
- Subpopulation 1: 19/24 patients responded to FFX characterized by higher expression of *CLEC19A* and *LRRC29*
- Subpopulation 2: 84 patients (44 FFX and 40 GnP responders) 78% of GnP responders belong to this subpopulation.

Gene interaction networks reveal *LRRC29* is related to *HOOK1* which negatively regulates EMT)by inhibiting *SHP2* activity.

Decreased *HOOK1* and *SHP2* levels are associated with a better GnP response to GnP in NSCLC suggesting that chemotherapy response may be cancer-agnostic and may depend on the chemotherapeutic agent and genetic factors involved in disease progression, with *LRRC29* expression serving as a biomarker for selecting GnP versus FFX treatment.

Identifying personas and variables that explain preferential response to GnP or FFX in PDAC patients using NetraAI

Evaluating GnP and FFX response as characterized as GnP Non-Responder (GnPNR) + FFX Responder (FFXR) vs GnP Responder (GnPR) + FFX Non-Responder (FFXNR)



FFX Preferential Response

- **26 subjects (16 FFXR/GnPNR, 10 GnPR/FFXNR)** characterized by 3 variables (**p=0.00799, Cohen’s D=1.086**): *C3AR1, MIEF1, HOXB6*
- **25 subjects (13 FFXR/GnPNR, 12 GnPR/FFXNR)** characterized by 3 variables (**p=0.00731, Cohen’s D=1.129**): *SUV39H1, MBNL1, PHACTR4*
- **25 subjects (15 FFXR/GnPNR, 10 GnPR/FFXNR)** characterized by 3 variables (**p=0.0079, Cohen’s D=1.097**): *DDTL, YBX3, FRMD4A*

GnP Preferential Response

- **28 subjects (17 FFXR/GnPNR, 11 GnPR/FFXNR)** characterized by 3 variables (**p=0.00294, Cohen’s D=1.2**): *TRIM25, RAB40B, NUTF2*
- **26 subjects (16 FFXR/GnPNR, 10 GnPR/FFXNR)** characterized by 2 variables (**p=0.00799, Cohen’s D=1.28**): *MVP, AC004980.7*
- **25 subjects (15 FFXR/GnPNR, 10 GnPR/FFXNR)** characterized by 3 variables (**p=0.000612, Cohen’s D=1.49**): *DENND3, PDE8A, MMP15*

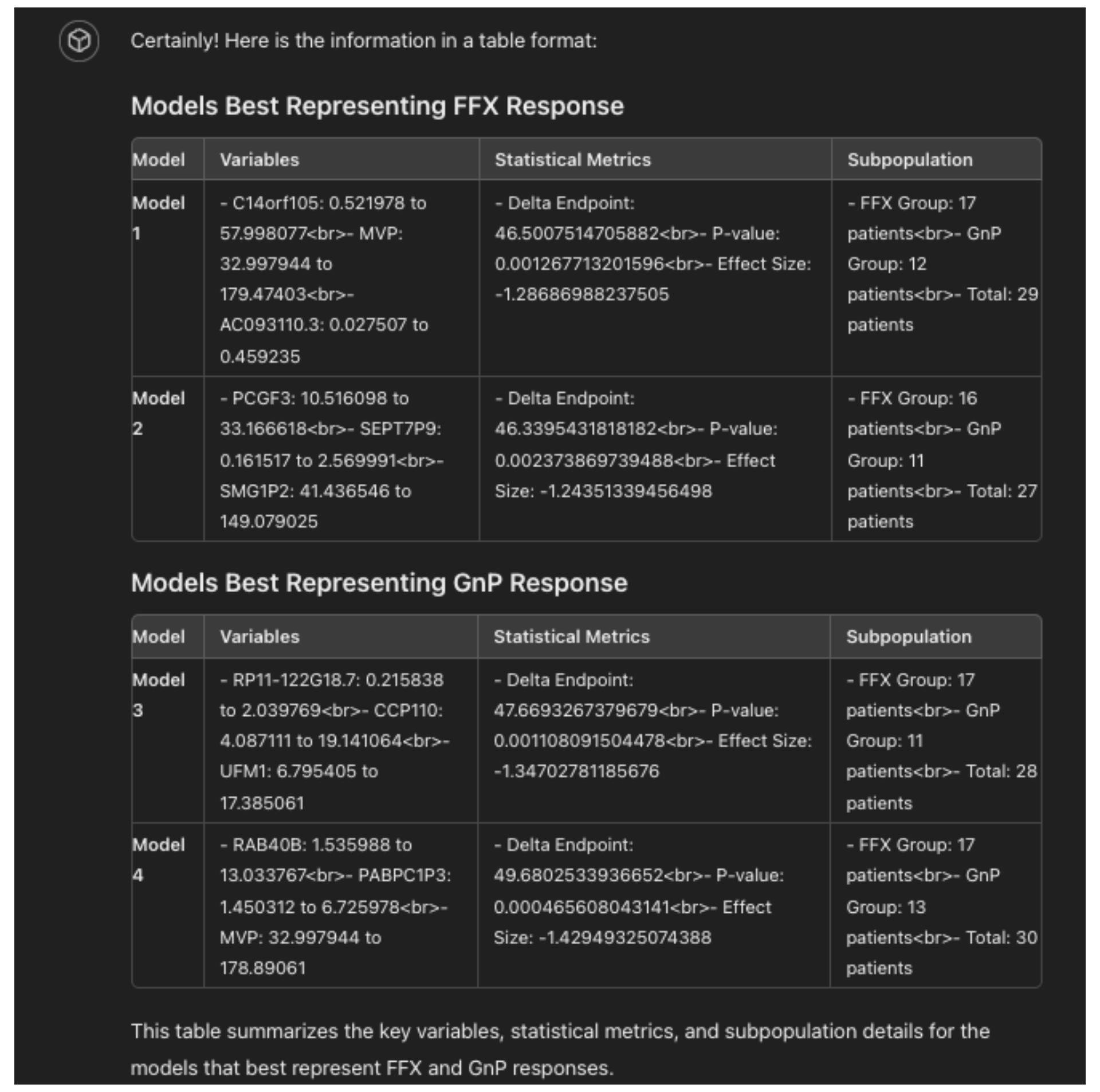
NetraGPT is a clinical trial companion built using a Large Language Model (LLM) trained through NetraAI-generated insights

NetraGPT is an LLM built on the NetraAI-derived insights, ***turning static clinical data into an interactive, dynamic tool.***

By leveraging a vast medical and scientific corpus, NetraGPT provides unbiased answers to clinical trial questions, avoiding the reinforcement of existing biases. This enables clinical trialists to explore new perspectives and improve decision-making without being confined to the limitations of traditional data analysis.

Key Features of NetraGPT:

- **MAML Technology:** Extracts precise insights from patient populations in oncology trials.
- **Value of Early Trials:** Harnessing critical data from earlier trials increases chances of follow-up pivotal trial success.
- **Handling Small Datasets:** MAML enables ML to learn from smaller datasets typical in oncology trials.
- **Managing Heterogeneity:** Capable of addressing the unknown heterogeneity inherent in oncology patient data.
- **Patient Differentiation:** Identifies variables that distinguish patients who will respond to a medication versus those who will not respond to a comparator drug.
- **Enrichment Strategies:** Enables the development of enrichment strategies to differentiate novel therapeutics.
- **Market Strategy Impact:** Aids in creating a powerful market strategy while addressing unmet patient needs.



Certainly! Here is the information in a table format:

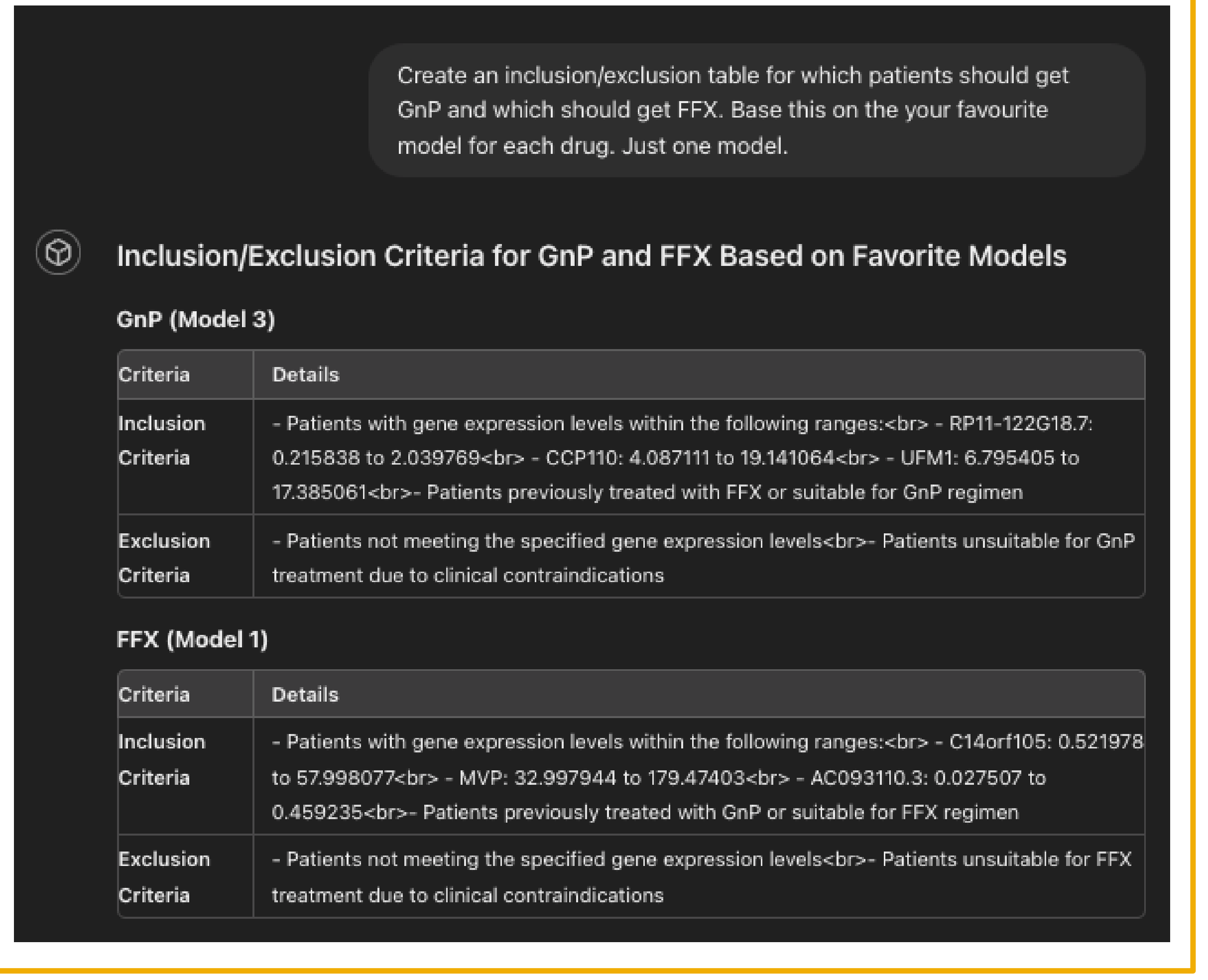
Models Best Representing FFX Response

Model	Variables	Statistical Metrics	Subpopulation
Model 1	- C14orf105: 0.521978 to 57.998077 - MVP: 32.997944 to 179.47403 - AC093110.3: 0.027507 to 0.459235	- Delta Endpoint: 46.5007514705882 - P-value: 0.001267713201596 - Effect Size: -1.28686988237505	- FFX Group: 17 patients - GnP Group: 12 patients - Total: 29 patients
Model 2	- PCGF3: 10.516098 to 33.166618 - SEPT7P9: 0.161517 to 2.569991 - SMG1P2: 41.436546 to 149.079025	- Delta Endpoint: 46.3395431818182 - P-value: 0.002373869739488 - Effect Size: -1.24351339456498	- FFX Group: 16 patients - GnP Group: 11 patients - Total: 27 patients

Models Best Representing GnP Response

Model	Variables	Statistical Metrics	Subpopulation
Model 3	- RP11-122G18.7: 0.215838 to 2.039769 - CCP110: 4.087111 to 19.141064 - UFM1: 6.795405 to 17.385061	- Delta Endpoint: 47.6693267379679 - P-value: 0.001108091504478 - Effect Size: -1.34702781185676	- FFX Group: 17 patients - GnP Group: 11 patients - Total: 28 patients
Model 4	- RAB40B: 1.535988 to 13.033767 - PABPC1P3: 1.450312 to 6.725978 - MVP: 32.997944 to 178.89061	- Delta Endpoint: 49.6802533936652 - P-value: 0.000465608043141 - Effect Size: -1.42949325074388	- FFX Group: 17 patients - GnP Group: 13 patients - Total: 30 patients

This table summarizes the key variables, statistical metrics, and subpopulation details for the models that best represent FFX and GnP responses.



Create an inclusion/exclusion table for which patients should get GnP and which should get FFX. Base this on the your favourite model for each drug. Just one model.

Inclusion/Exclusion Criteria for GnP and FFX Based on Favorite Models

GnP (Model 3)

Criteria	Details
Inclusion Criteria	- Patients with gene expression levels within the following ranges: - RP11-122G18.7: 0.215838 to 2.039769 - CCP110: 4.087111 to 19.141064 - UFM1: 6.795405 to 17.385061 - Patients previously treated with FFX or suitable for GnP regimen
Exclusion Criteria	- Patients not meeting the specified gene expression levels - Patients unsuitable for GnP treatment due to clinical contraindications

FFX (Model 1)

Criteria	Details
Inclusion Criteria	- Patients with gene expression levels within the following ranges: - C14orf105: 0.521978 to 57.998077 - MVP: 32.997944 to 179.47403 - AC093110.3: 0.027507 to 0.459235 - Patients previously treated with GnP or suitable for FFX regimen
Exclusion Criteria	- Patients not meeting the specified gene expression levels - Patients unsuitable for FFX treatment due to clinical contraindications

CONCLUSIONS AND SIGNIFICANCE

CONCLUSIONS

This study highlights the potential of mathematically-augmented ML in elucidating complex relationships within clinical trial data, offering valuable insights into the determinants of treatment response in pancreatic cancer. The identified biomarkers and patient characteristics provide a promising foundation for developing personalized treatment strategies, enhancing the precision of therapeutic decisions.

SIGNIFICANCE AND NEXT STEPS

Integrating MAML approaches into clinical practice could significantly improve patient outcomes by tailoring treatments to individual patient profiles. Our findings highlight the importance of further efforts in the field to achieve more effective and personalized cancer care.

REFERENCES

1. 1. Ushio, J. et al. Pancreatic Ductal Adenocarcinoma: Epidemiology and Risk Factors. Diagnostics 11, (2021).
2. Klein-Brill, A., Amar-Farkash, S., Lawrence, G., Collisson, E. A. & Aran, D. Comparison of FOLFIRINOX vs Gemcitabine Plus Nab-Paclitaxel as First-Line Chemotherapy for Metastatic Pancreatic Ductal Adenocarcinoma. JAMA Netw Open 5, e2216199–e2216199 (2022).
3. Aung, K. L. et al. Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. Clin Cancer Res 24, 1344–1354 (2018).
4. Pan, Z. et al. Analysis of dynamic molecular networks for pancreatic ductal adenocarcinoma progression. Cancer Cell Int 18, 214 (2018).
5. Yang, H. et al. The clinicopathological and prognostic implications of tyrosine phosphatase SHP2 and ankyrin Hook1 gene expression in non-small cell lung cancer patients treated with gemcitabine plus platinum as first-line chemotherapy. Ann Palliat Med 9, 2943952–2942952 (2020).