

Precision medicine in oncology through sub-insight learning from small data sets

Bessi Qorri¹, Mike J Tsay¹, Paul Leonchik¹, Larry Alphs¹, Luca Pani^{2,3}, Joseph Geraci^{1,4,5,6}

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

INTRODUCTION

PRECISION MEDICINE IN ONCOLOGY:
The cancer treatment era is characterized by the use of biomarkers to guide decisions in the prevention, diagnosis, and treatment of disease. This movement toward precision medicine aims to ensure that patients receive treatments specifically tailored to their unique disease pathology, with the goal of improving patient outcomes. Advancements in artificial intelligence (AI) and machine learning (ML) stand at the front line of this revolution, offering unprecedented capabilities in patient stratification and subtype discrimination in clinical trials.

CHALLENGES WITH PRECISION MEDICINE IN ONCOLOGY:
The promise of precision medicine requires sophisticated tools capable of deciphering complex biological data for complex diseases. The major challenge revolves around the datasets on which AI/ML models rely on.

- Requirement of Large Datasets:** Large datasets are often difficult to compile due to the rarity of some diseases.
- Limitations of Data:** Even large datasets, while rich in information, are not always reflective of the patient population, particularly in heterogeneous diseases like oncology. Many AI/ML methods only reinforce what is already known.

The demand for personalized treatment strategies in oncology necessitates novel AI/ML technologies that can leverage smaller or limited datasets to uncover actionable insights through refining patient stratification and subtype discrimination, overcoming the challenges faced with conventional methods that depend on large datasets.

OBJECTIVES:
Using NetraAI, a novel ML approach, we aim to analyze a non-small cell lung cancer (NSCLC) patient dataset to identify variables driving the etiology defining specific subpopulations of patients. This approach can pave the way for a more flexible framework in oncology, accommodating data variability and promoting the discovery of novel biomarkers and therapeutic targets. This ultimately can advance precision medicine with the goal of ensuring that more patients benefit from personalized treatment strategies.

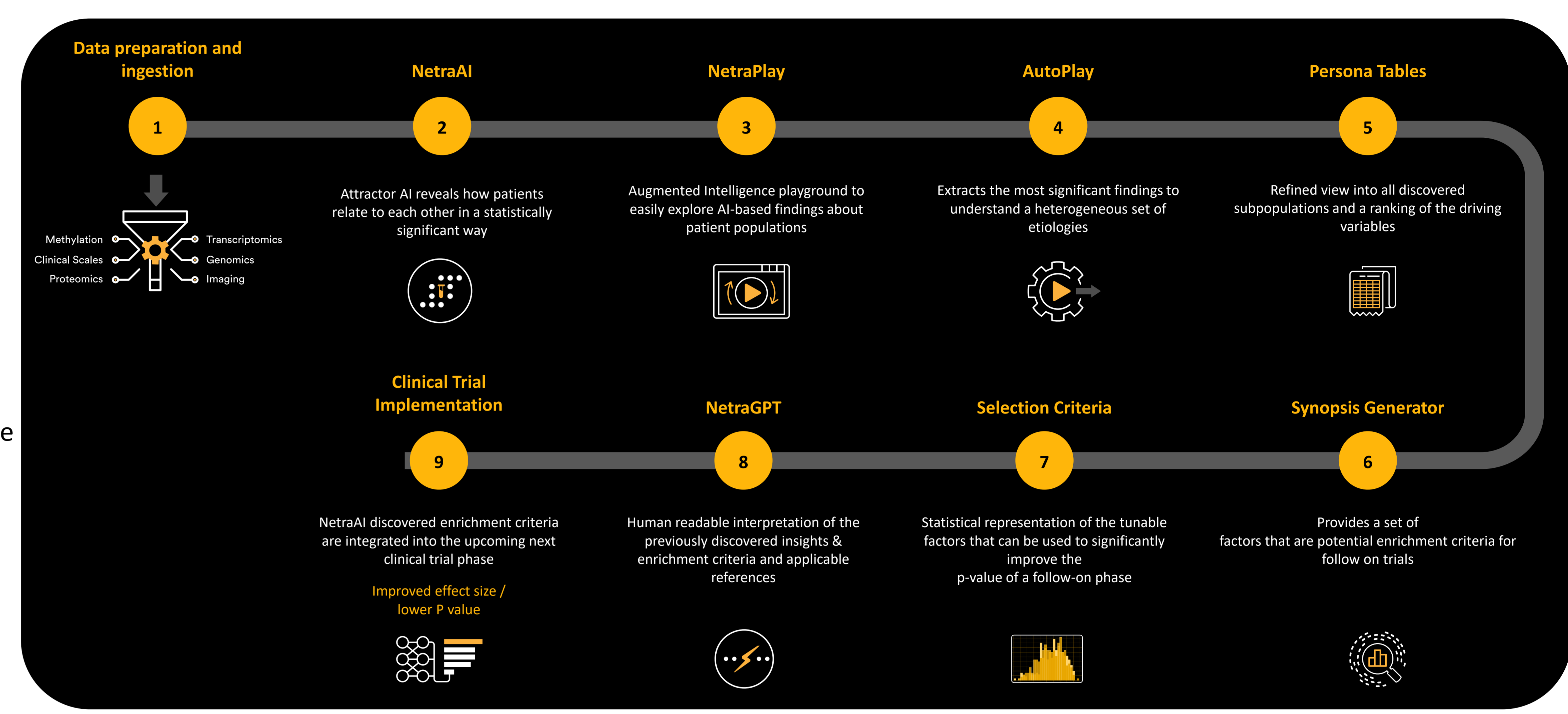
METHODOLOGY

NETRAAI – A NOVEL MACHINE LEARNING APPROACH:
NetraAI, powered by its unique Attractor AI algorithms, is designed to deliver actionable insights from limited data despite not reflecting the totality of the disease they represent.

SUB-INSIGHT LEARNING:
Using advanced mathematical methods, NetraAI segments data into explainable and unexplainable subpopulations (NetraPerspectives) according to a set of variables. This enables:

- The Identification of intricate relationships within the data.
- Reductions in the risk of overfitting by not adhering to the labels provided.
- A holistic and novel view of the patient population through the combination of different NetraPerspectives.

DATASET:
Using two NSCLC datasets (GSE18842 and GSE10245), we compiled a small dataset consisting of 104 gene expression samples of adenocarcinoma (ADC) and squamous cell carcinoma (SCC).

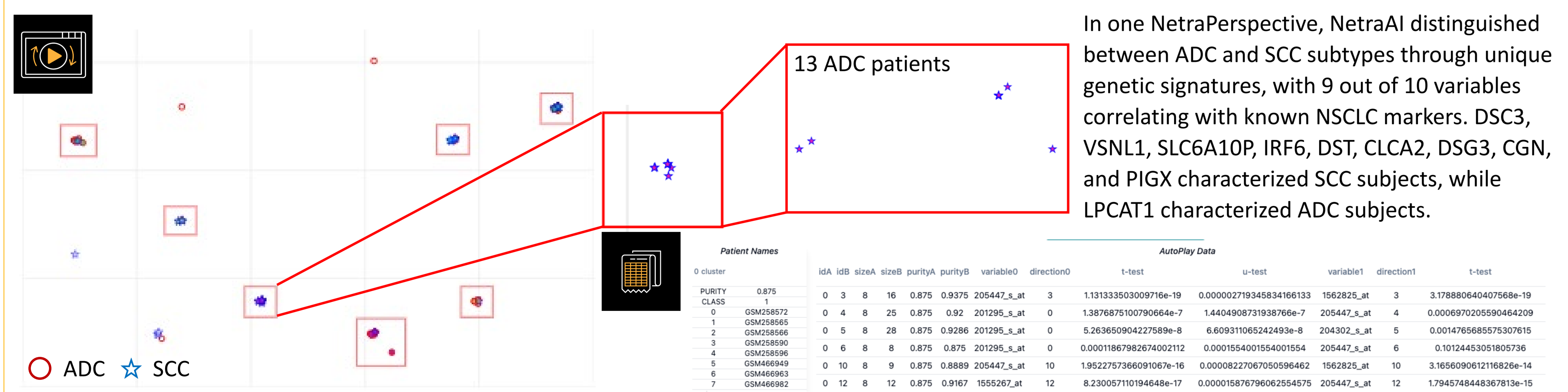


RESULTS

NetraAI-Identified ADC and SCC NSCLC Subpopulations

One NetraPerspective showed multiple explainable subpopulations of NSCLC, primarily stratified as ADC or SCC subpopulations. Interestingly, there are multiple subpopulations of each subtype suggesting that there are different combinations of variables driving specific ADC and SCC subtypes. Using the unique zoom capabilities of NetraAI, we are able to examine each subpopulation to identify the specific patients and characterizing variables.

From a single NetraPerspective, Persona Tables are generated, listing variables and their significance values characterizing each subpopulation. The characterizing variables and their significance values are reported based on the comparison of two subpopulations of opposite classes or labeling. The purity of the subpopulation is also reported to reveal the confidence with which the variables characterize the subpopulation.

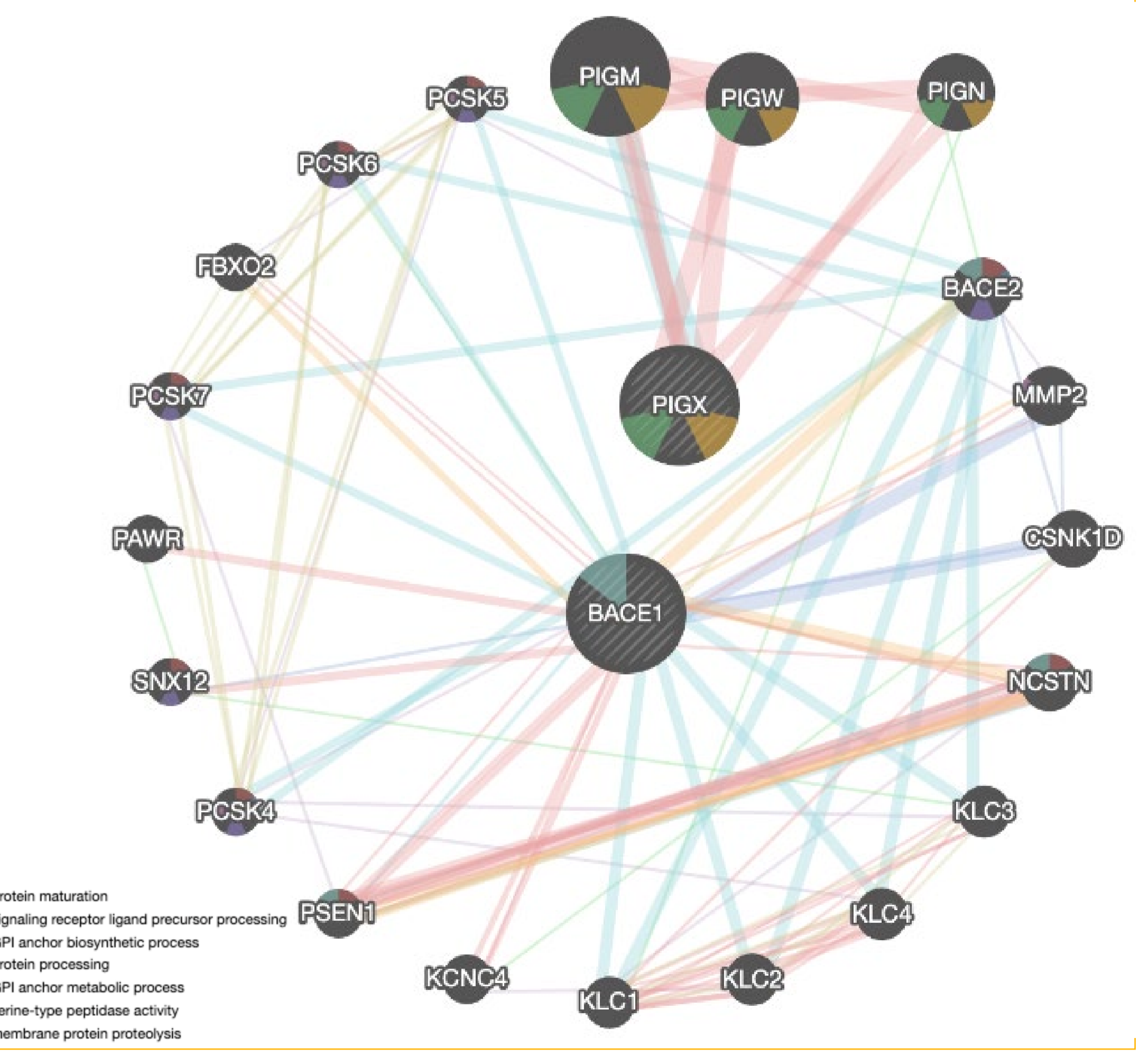


PIGX as a novel marker of NSCLC and protein-protein interaction (PPI) networks

With 9 out of 10 variables identified discriminating between ADC and SCC being previously reported as NSCLC variables, this validates the methods and technology used by NetraAI. PIGX emerged as a novel target due to its previously unexplored role in cancer biology.

Further investigation into protein-protein interaction (PPI) networks revealed a significant connection between PIGX and BACE1, a protein implicated in NSCLC brain metastasis. This opens new avenues for understanding molecular mechanisms underlying cancer progression and metastasis.

- BACE1 functions to process proteins responsible for proper functioning of neuronal tissue.
- BACE2 correlates with worse prognosis in several cancers; BACE1 is implicated in the tumor microenvironment (TME) and has pro-tumorigenic effects.
- The association between PIGX and BACE1 emphasizes the critical role of membrane proteins and their post-translational modifications in disease etiology.
- PIGX is also related to PIGN which is associated with genomic instability and regulates spindle assembly checkpoint proteins in leukemia transformation and progression.
- PIGN encodes for a protein involved in the biosynthesis of GPI anchors (glycosylphosphatidylinositol) and has immune related roles in cancer biology



CONCLUSIONS & SIGNIFICANCE

SUMMARY OF KEY FINDINGS:

- PIGX is a novel potential biomarker for NSCLC. Connections of PIGX to PIGN and BACE1 underscore the complexity of the disease and open up new avenues for research.
- The biology of membrane proteins and their post-translational modifications, implicated in various diseases prompts further exploration.

SIGNIFICANCE OF FINDINGS:

- NetraAI can extract meaningful insights from small datasets, offering an alternative approach to stratify patients while not reinforcing our current knowledge and beliefs about patient populations.
- PIGX and PIGN disruptions may represent an opportunity for novel drug discoveries and may also represent an opportunity for the creation of novel immunotherapies that can penetrate the TME.

FUTURE USE AND APPLICATIONS:
The NetraAI approach has the potential to advance current research paradigms and patient-specific treatments contributing to precision medicine, ensuring that patients receive the most effective care based on their specific genetic make up and disease etiology.

REFERENCES

Qorri B, Geraci J, Bessi Q, et al. Small pattern datasets reveal genetic drivers of non-small cell lung cancer subtypes using machine learning for hypothesis generation. *Expert Med*. 2023;4(4):401-410. <https://doi.org/10.1080/17473669.2023.2201113>. Toyokuni S, Sato A, Kim T, et al. Protein-protein interaction network analysis reveals novel drug targets for cancer treatment. *Front Oncol*. 2022;12:958585. <https://doi.org/10.3389/fonc.2022.958585>. <https://pubmed.ncbi.nlm.nih.gov/36188032/>