

Leveraging Artificial Intelligence to Structure Next-Generation Sequencing Data for Precision Oncology Research: A Real-world Pancreatic Ductal Adenocarcinoma Liquid Biopsy Experience

Jacob Gabayan¹, Yousuf Altameemi² and Ali Muhammad MD, MBA^{3*}

¹AI Automation Specialist and Clinical Research Fellow, Beverly Hills Cancer Center and Clinical Research Advisors LLC, California, USA

²Clinical Research Coordinator, Beverly Hills Cancer Center and Clinical Research Advisors LLC, California, USA

³Co-Founder, Managing Member and Director of Clinical Research, Clinical Research Advisors LLC, California, USA

Abstract

Precision oncology generates increasingly complex molecular datasets that can be difficult to organize and analyze efficiently. We evaluated 212 real-world pancreatic ductal adenocarcinoma liquid biopsy datasets and explored AI-assisted workflows to transform unstructured next-generation sequencing outputs into structured research datasets. The analysis characterized 99 unique biomarkers and genes and demonstrates the potential operational value of AI-enabled data structuring for research, biomarker assessment and future clinical trial matching.

Keywords: Pancreatic ductal adenocarcinoma • Artificial Intelligence • Next-generation Sequencing • Liquid Biopsy • Comprehensive Genomic Profiling • Precision Oncology • Biomarkers • Clinical Research • Real-world Data • Beverly Hills Cancer Center • Clinical Research Advisors LLC

Introduction

NGS testing has transformed oncology; however, manual extraction of molecular findings remains labor intensive. This project sought to characterize the PDAC genomic landscape while examining how AI-assisted tools can support organization, standardization and analysis of high-volume genomic information.

Methods

Retrospective review of PDAC records identified within the Beverly Hills Cancer Center and its Clinical Research Advisors satellite sites' liquid biopsy dataset. Records containing pancreatic cancer terminology were reviewed and aggregated. AI-assisted workflows were conceptually utilized to standardize biomarker data, support dataset structuring and facilitate downstream analysis. The analyzed database contained 212 PDAC-related genomic data sets (Table 1 and Table 2).

AI-assisted extraction and standardization outputs were reviewed against the source genomic reports before analysis. Biomarker names, specimen-level observations, duplicate records and derived counts were checked using a predefined quality-control process, with discrepancies resolved through manual review [1-4].

***Address for Correspondence:** Ali Muhammad MD, MBA, Co-Founder, Managing Member and Director of Clinical Research, Clinical Research Advisors LLC, California, USA; Tel: +13474465612 E-mail: Dr.Ali@ClinicalResearchAdvisors.com

Copyright: © 2026 Muhammad A. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: 06 October, 2026, Manuscript No. jcc-26-194849; **Editor assigned:** 08 Oct. 26, 2026, PreQC No. P-194849; **Reviewed:** 12 October, 2026, QC No. Q-194849; **Revised:** 19 October, 2026, Manuscript No. R-194849; **Published:** 26 October, 2026, DOI: 10.37421/2577-0535.2026.15.388

Results

The molecular landscape was dominated by canonical driver alterations including KRAS, TP53, SMAD4 and CDKN2A. AI-assisted organization enabled rapid aggregation of molecular observations into searchable datasets and visual analytics. Figure 1 presents the most frequently observed biomarkers/genes in the PDAC subset (Figure 1) [5-6].

Operational Impact of AI

AI-assisted approaches reduced dependence on manual abstraction, improved consistency of nomenclature, enabled scalable data review and provide a framework for future clinical trial selection on institutional level and potential clinical trial matching decisions. Rather than replacing physician judgment, AI functioned as a force multiplier supporting efficient data organization and analysis [7-9].

Limitations

This retrospective analysis was derived from a single network dataset and is subject to limitations associated with accession-level analyses, free-text diagnoses and heterogeneous testing patterns. Clinical outcomes, survival endpoints and treatment response data were not evaluated at this point although with the encouraging results of this project we are planning future analysis of clinical outcomes, survival endpoints and treatment response rates related to different therapeutic indications and treatment approaches.

Discussion and Conclusions

The future of precision oncology depends not only on improved diagnostics but also on improved information management. This report suggests that AI-assisted data structuring can complement traditional oncology research by enabling efficient handling of increasingly complex genomic datasets.

Table 1: Dataset Summary.

Metric	Value
PDAC-related datasets	212
Unique biomarkers/genes	99

Table 2: Top Biomarkers/Genes.

Biomarker/Gene	Count
DNMT3A	170
TP53	140
MSI-High	120
Tumor Mutational Burden (TMB)	90
TET2	90
KRAS	80
Tumor Fraction	80
ATM	70
PPM1D	40
UGT1A1	40

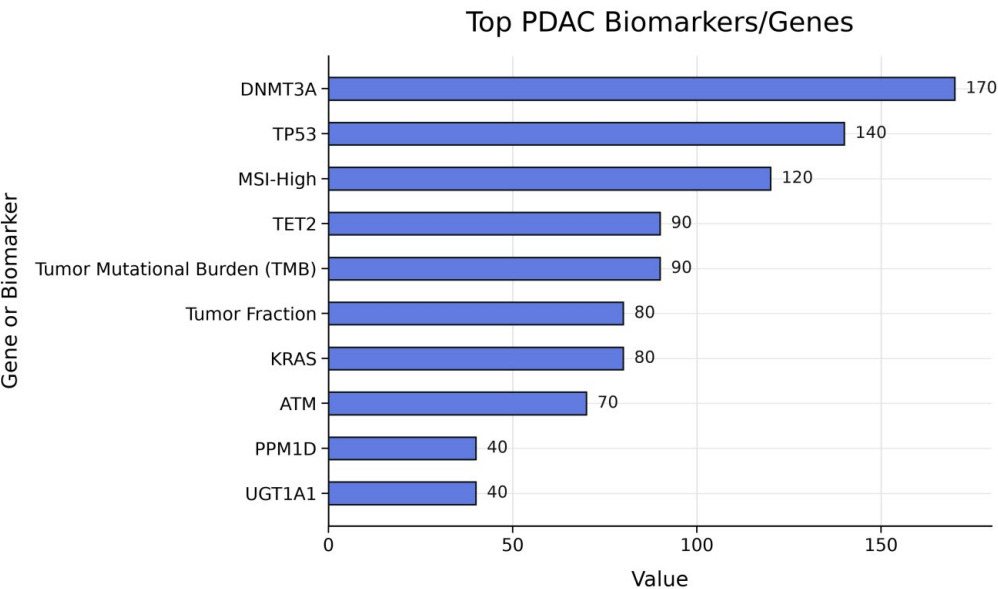


Figure 1. Frequency distribution of the most commonly observed biomarkers/genes within the PDAC subset of the Beverly Hills Cancer Center and Clinical Research Advisors affiliated sites liquid biopsy dataset.

Research Locations

As a private, academic, community-based cancer center, Beverly Hills Cancer Center provides state-of-the-art, multidisciplinary cancer care while supporting leading clinical trials and oncology research. By integrating specialized treatment services with clinical research operations, Beverly Hills Cancer Center offers patients access to innovative investigational therapies and advances the development of new approaches to cancer treatment.

Clinical research is conducted through Beverly Hills Cancer Center, the main Beverly Hills research location of Clinical Research Advisors LLC and five affiliated oncology satellite sites located in West Hollywood, Encino, Korea Town, West Covina and Downtown Los Angeles. This integrated research network is supported by multidisciplinary investigators and experienced clinical research professionals who coordinate patient identification, informed consent, protocol-directed assessments, treatment and follow-up across the greater Los Angeles area.

Conflict of Interest

The authors declare no potential conflicts of interest related to this work.

References

1. Siegel, Rebecca L., Kimberly D. Miller and Ahmedin Jemal. "Cancer statistics, 2020." *CA Cancer J Clin* 70 (2020): 7-30.
2. Mateo, J, et al. "Precision medicine in oncology." *Nat Rev Clin Oncol*.
3. Meric-Bernstam, F, et al. "Clinical implementation of genomic profiling in oncology."
4. Bailey, Peter, David K. Chang, Katia Nones and Amber L. Johns, et al. "Genomic analyses identify molecular subtypes of pancreatic cancer." *Nature* 531 (2016): 47-52.
5. Waddell, Nicola, Marina Pajic, Ann-Marie Patch and David K. Chang, et al. "Whole genomes redefine the mutational landscape of pancreatic cancer." *Nature* 518 (2015): 495-501.

6. Raphael, Benjamin J, Ralph H. Hruban andrew J. Aguirre and Richard A. Moffitt, et al. "Integrated genomic characterization of pancreatic ductal adenocarcinoma." *Cancer Cell* 32 (2017): 185-203.
7. Lee, Jee-Soo, Sung Sup Park, Young Kyung Lee and Jeffrey A. Norton, et al. "Liquid biopsy in pancreatic ductal adenocarcinoma: Current status of circulating tumor cells and circulating tumor DNA." *Mol Oncol* 13 (2019): 1623-1650.
8. Truhn, Daniel, Chiara ML Loeffler, Gustav Müller-Franzes and Sven Nebelung, et al. "Extracting structured information from unstructured histopathology reports using generative pre-trained transformer 4 (GPT-4)." *J Pathol* 262 (2024): 310-319.
9. Balasubramanian, Jeya Balaji, Daniel Adams, Ioannis Roxanis and Amy Berrington de Gonzalez, et al. "Leveraging large language models for structured information extraction from pathology reports." *J Pathol Inform* (2025): 100521.

How to cite this article: Jacob Gabayan, Yousuf Altameemi and Ali Muhammad MD, MBA. "Leveraging Artificial Intelligence to Structure Next-generation Sequencing Data for Precision Oncology Research: A Real-world Pancreatic Ductal Adenocarcinoma Liquid Biopsy Experience" *J Cancer Clin Trials* 15 (2026): 388.