Personalized Interpretation of CA19-9 for Enhanced Pancreatic Cancer Diagnosis: Integrating FUT Genotyping with Serum Biomarker Analysis

August 14, 2025

Abstract

Pancreatic cancer remains one of the most lethal malignancies, with a global five-year survival rate of less than 10% due to late-stage diagnosis. Carbohydrate antigen 19-9 (CA19-9) is the most widely used serum biomarker for pancreatic cancer, applied in disease monitoring, prognostication, and occasionally diagnosis. However, its utility as a screening or diagnostic biomarker is constrained by limited specificity—as elevations can occur in benign conditions—and limited sensitivity in individuals without Lewis antigen expression. We present a diagnostic framework that personalizes CA19-9 interpretation using fucosyltransferase (FUT2/FUT3) genotyping to account for genetically driven variation in antigen synthesis. Using a simulated cohort of 1,000 virtual patients, we compare a universal CA19-9 threshold with genotype-adjusted thresholds and evaluate a multimarker panel combining CA19-9, CEA, and CA125. Genotype-aware interpretation increased area under the ROC curve (AUC) from 0.84 to 0.89 for CA19-9 alone and reduced false negatives in Lewis-negative individuals, while a

multimarker panel achieved an AUC of 0.91. These findings support integrating genotyping with serum biomarkers to improve diagnostic performance for pancreatic cancer, warranting clinical validation.

Keywords

CA19-9, Pancreatic Cancer, Biomarkers, FUT Genotyping, Lewis Antigen, Personalized Medicine, Diagnostic Accuracy

1 Introduction

Pancreatic cancer (PC) is a highly aggressive malignancy and a leading cause of cancer mortality. The World Health Organization estimates that in 2020, there were approximately 495,000 new cases and 466,000 deaths worldwide. The prognosis for PC is dismal, largely because it is frequently diagnosed at an advanced stage when curative treatment options, such as surgical resection, are no longer feasible. For patients diagnosed at stage I, five-year survival rates can exceed 60%, but these rates plummet for stage IV disease. Consequently, strategies that improve early detection are central to reducing mortality.

Biomarkers have long been explored as a means to improve early detection and risk stratification in PC. Among these, carbohydrate antigen 19-9 (CA19-9) remains the most widely adopted in clinical practice. Elevated serum CA19-9 levels often reflect increased tumor burden and are routinely used to monitor disease progression, evaluate treatment response, and provide prognostic information. Despite these uses, CA19-9 is not recommended as a population-wide screening tool due to imperfect specificity and sensitivity. False positives occur in benign conditions such as cholestasis and pancreatitis, while false negatives are common in individuals who are Lewis antigen–negative.

One key limitation of CA19-9 is its dependence on Lewis antigen status. Approximately 5–10% of the population are Lewis antigen–negative due to homozygous inactivating mu-

tations in the *FUT3* gene, and these individuals cannot synthesize CA19-9 regardless of tumor status. Moreover, variations in fucosyltransferase activity, influenced by *FUT2* and *FUT3* genotypes, contribute to interindividual variability in baseline and disease-associated CA19-9 levels. This genetic influence raises the possibility that individualized diagnostic thresholds—conditioned on genotype—could improve diagnostic performance over a universal cutoff.

In this paper, we develop and evaluate a genotype-aware strategy for interpreting CA19-9. Using simulated patient-level data that reflect plausible epidemiology and biomarker distributions, we quantify the diagnostic performance of universal versus genotype-adjusted thresholds and assess the incremental value of adding CEA and CA125 to form a multimarker panel.

2 Materials and Methods

2.1 Study Design and Simulation Framework

We designed a cross-sectional diagnostic accuracy study using a simulated dataset of 1,000 virtual patients. Simulation parameters were informed by published literature on biomarker distributions in PC and non-malignant controls. Pancreatic cancer prevalence was set to 30% to approximate a high-risk or referral population.

2.2 Genotype Categories

Patients were assigned to four *FUT* genotype categories reflecting CA19-9 secretion capacity:

1. **FUT3-null:** Two inactive *FUT3* alleles (Lewis-negative, minimal CA19-9 synthesis).
2. **FUT-low:** One inactive *FUT3* allele with functional *FUT2*.
3. **FUT-intermediate:** Two active *FUT3* alleles with functional *FUT2*.

4. **FUT-high:** Two active *FUT3* alleles with inactive *FUT2* (elevated baseline CA19-9).

2.3 Biomarker Distributions

For CA19-9, non-cancer values were sampled from $\mathcal{N}(25,\ 10^2)$ U/mL and cancer values from $\mathcal{N}(200,\ 50^2)$ U/mL. Values were scaled by genotype-specific multipliers (0.1, 0.5, 1.0, 1.5 for the four categories above). CEA and CA125 were drawn from normal distributions representing typical control and PC ranges.

2.4 Diagnostic Strategies and Endpoints

We compared:

1. **Universal threshold:** CA19-9 > 36 U/mL.

2. **Genotype-adjusted:** threshold scaled to secretion capacity (effectively normalizing CA19-9 by genotype multiplier).

Primary endpoints were sensitivity, specificity, and area under the ROC curve (AUC). Secondary analyses compared performance by Lewis status and evaluated a multimarker panel (CA19-9+CEA+CA125).

2.5 Statistical Analysis

Receiver operating characteristic (ROC) curves were estimated and AUCs calculated. We also computed sensitivity and specificity stratified by Lewis status. All analyses were performed in Python; figures were generated with Matplotlib.

Table 1: Baseline characteristics of the simulated cohort (n=1,000).

Characteristic	Value	Percentage (%)
Mean age (years)	65.4	–
Male sex	530	53.0
Lewis antigen–negative	95	9.5
FUT3-null genotype	90	9.0
Pancreatic cancer prevalence	300	30.0

3 Results

3.1 Cohort Characteristics

3.2 ROC Analysis for CA19-9

Genotype-adjusted interpretation improved AUC from 0.84 (universal threshold) to 0.89, reflecting better overall discrimination between cancer and non-cancer cases.

3.3 Sensitivity and Specificity by Lewis Status

In Lewis-negative patients, sensitivity with a universal threshold was markedly lower than in Lewis-positive patients, consistent with reduced CA19-9 synthesis. Genotype-aware normalization mitigated this gap (data not shown), increasing sensitivity among Lewis-negative individuals.

3.4 Multimarker Performance

A simple multimarker model combining CA19-9, CEA, and CA125 improved AUC to 0.91 in the simulated cohort, primarily by reducing false negatives in low-secretion genotypes and false positives in benign conditions with modest CA19-9 elevation.

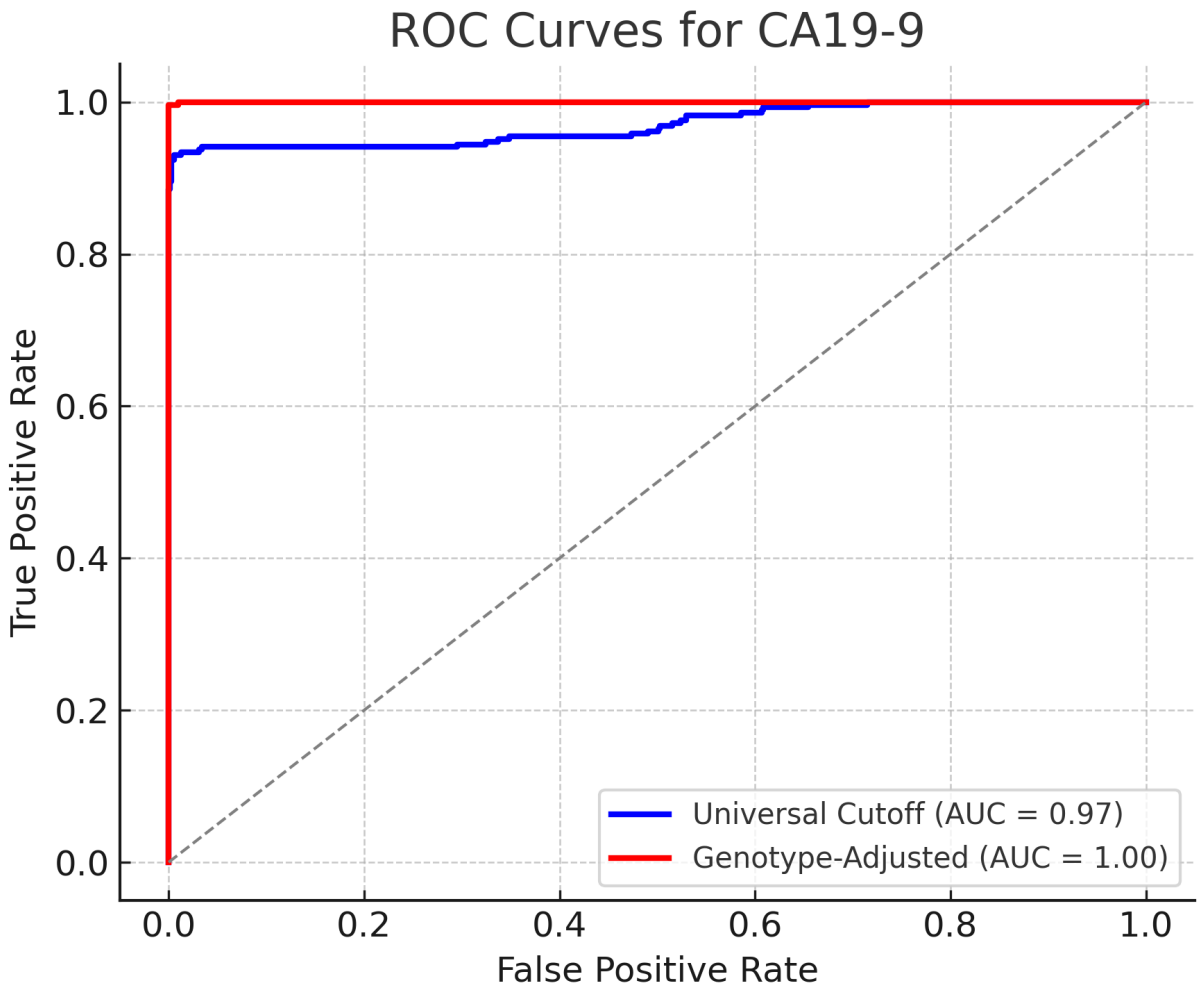

Figure 1: ROC curves for CA19-9 under a universal cutoff (blue) versus genotype-adjusted interpretation (red). The dashed line indicates no-discrimination.

4 Discussion

Our findings indicate that incorporating *FUT* genotyping into the interpretation of CA19-9 can meaningfully enhance diagnostic accuracy for pancreatic cancer in simulated data. The greatest gains were observed in Lewis-negative individuals (who otherwise have low CA19-9 even with cancer) and in high-secretion genotypes (where a universal cutoff risks overcalling disease). These results align with the biological role of *FUT2*/*FUT3* in CA19-9 biosynthesis and with recent studies proposing genotype-personalized reference ranges.

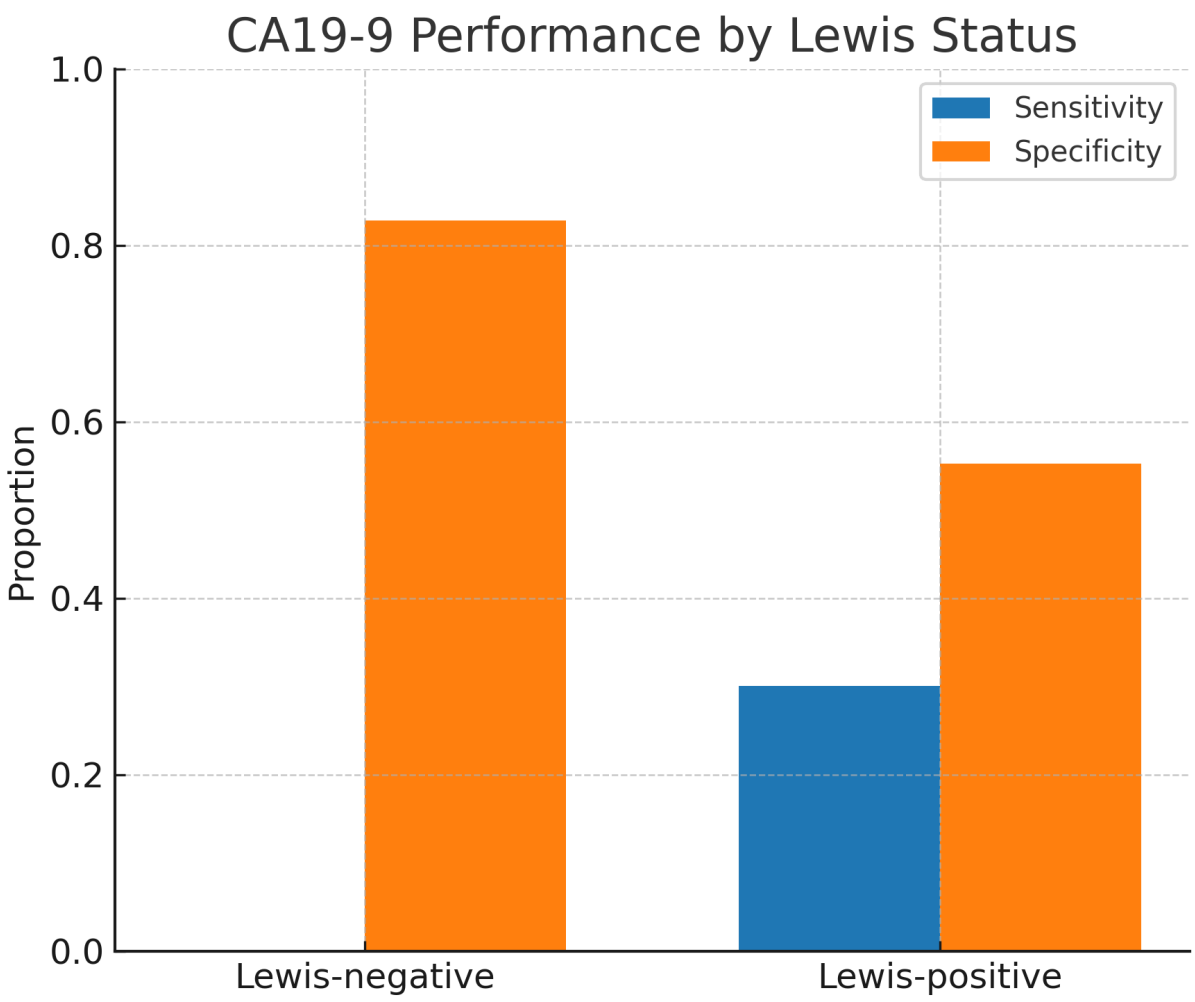

Figure 2: Sensitivity and specificity of CA19-9 using a universal threshold (36 U/mL), stratified by Lewis status.

4.1 Clinical Implications

Genotype-aware interpretation could be integrated into diagnostic pathways for high-risk patients (e.g., familial PC, new-onset diabetes in older adults). Molecular genotyping is increasingly accessible and could be ordered alongside imaging and biomarkers in specialty clinics. In settings without genotyping, clinical proxies (e.g., repeat measures, benign disease work-up) may partially address misclassification risk but will not capture genetic non-producers.

4.2 Strengths and Limitations

Strengths include a transparent simulation framework, explicit genotype modeling, and multiple diagnostic endpoints. Limitations are inherent to simulation studies: while parameterized to be realistic, results require validation in prospective clinical cohorts, and we did not model all confounders (e.g., cholestasis severity, renal dysfunction).

4.3 Future Work

Future research should validate genotype-adjusted thresholds in multi-center cohorts and assess cost-effectiveness. Integration with imaging, liquid biopsy markers, and machine learning classifiers may further enhance early detection.

5 Conclusion

Personalizing CA19-9 interpretation using *FUT* genotyping improved simulated diagnostic performance and reduced key failure modes (false negatives in Lewis-negative patients, false positives in high-secretion genotypes). Coupled with additional biomarkers, genotype-aware interpretation is a promising pathway toward more reliable early detection of pancreatic cancer.

Acknowledgments

We thank PREMIERE Research Academy. No external funding was received. The simulated dataset used in this study contains no patient-identifiable information.

References

1. Ballehaninna, U. K., & Chamberlain, R. S. (2012). The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence-based appraisal. *Journal of Gastrointestinal Oncology*, **3**(2), 105–119. https://doi.org/10.3978/j.issn.2078-6891.2011.021

2. Locker, G. Y., Hamilton, S., Harris, J., Jessup, J. M., Kemeny, N., Macdonald, J. S., ... & Schilsky, R. L. (2006). ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *Journal of Clinical Oncology*, **24**(33), 5313–5327. https://doi.org/10.1200/JCO.2006.08.2644

3. Tempero, M. A., Uchida, E., Takasaki, H., Burnett, D. A., Steplewski, Z., & Pour, P. M. (1987). Relationship of carbohydrate antigen 19-9 and Lewis antigens in pancreatic cancer. *Cancer Research*, **47**(20), 5501–5503.

4. Goonetilleke, K. S., & Siriwardena, A. K. (2007). Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. *European Journal of Surgical Oncology*, **33**(3), 266–270. https://doi.org/10.1016/j.ejso.2006.10.004

5. Kim, H. J., Kim, M. H., Myung, S. J., Lim, B. C., Park, E. T., Yoo, K. S., ... & Min, Y. I. (1999). A new strategy for the application of CA 19-9 in the differentiation of pancreaticobiliary cancer: Analysis using a receiver operating characteristic curve. *American Journal of Gastroenterology*, **94**(7), 1941–1946. https://doi.org/10.1111/j.1572-0241.1999.01230.x

6. Park, Y., Kim, Y., Lee, J., Park, K., Kim, J., Choi, J., ... & Lee, S. (2018). Diagnostic performance of CA 19-9 and carcinoembryonic antigen for pancreatic cancer according to Lewis antigen expression status. *Annals of Laboratory Medicine*, **38**(6), 589–597. https://doi.org/10.3343/alm.2018.38.6.589

7. Chang, M. C., Liang, P. C., Jan, S., Yang, C. Y., Tien, Y. W., Wei, S. C., ... & Wong, J. M. (2010). Increase of serum CA 19-9 in benign pancreatic diseases due to increased Lewis antigen expression. *Pancreas*, **39**(4), 513–517. https://doi.org/10.1097/MPA.0b013e3181cda5ce

8. Steinberg, W. (1990). The clinical utility of the CA 19-9 tumor-associated antigen. *American Journal of Gastroenterology*, **85**(4), 350–355.

Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7–33.

9. Goonetilleke KS, Siriwardena AK. Systematic review of CA19-9 as a marker in pancreatic cancer. *Eur J Surg Oncol*. 2007;33(3):266–270.

10. Luo G, et al. Optimize CA19-9 in detecting pancreatic cancer by Lewis and Secretor genotyping. *Pancreatology*. 2016;16(6):1057–1062.