AACR Special Conference: Precision Prevention, Early Detection, and Interception of Cancer

Improved sensitivity of a multi-analyte early detection test based on mutation, methylation, aneuploidy, and protein biomarkers

Vladimir Gainullin, Leonardo Hagmann, Kevin Arvai, Amira Best, Christopher L. Nobles, Hee Jung Hwang, Larson Hogstrom, Mael Manesse, Fanglei Zhuang, Vuna Fa, Xi Chen, Viatcheslav E. Katerov, Jorge Garces, Tomasz M. Beer, Hatim T. Allawi, Abigail McElhinny, Frank Diehl*, Gustavo C. Cerqueira

Exact Sciences Corporation, Madison, WI * Corresponding Author: fdiehl@exactsciences.com

Presented at the AACR Special Conference: Precision Prevention, Early Detection, and Interception of Cancer, November 17-19, 2022. Austin, TX

Background

- We recently demonstrated the performance of a multi-analyte early cancer detection test based on the combination of methylation, aneuploidy, protein, and mutation biomarkers in a large feasibility study (1). The performance of two different biomarker combinations was evaluated on cancers from up to 15 organ sites and all stages.
- This presentation provides an updated multi-analyte biomarker performance for this feasibility study including two changes. First, 127 samples from the test set that were previously excluded from the overlap analysis have now been qualified to be included. Second, the previously utilized OR-logic that was used to combine the methylation, aneuploidy and proteins classifier results was replaced with an overarching machine learning (ML)-based classifier. The ML classifier was applied in the overall performance assessment shown in **Table 1**.
- Free and total prostate-specific antigen (PSA) measurements were investigated as independent markers in parallel to the multi-analyte testing to increase the overall detection of clinically-actionable prostate cancers. First, an optimal threshold for total PSA (tPSA) was defined to reduce overdiagnosis. Second, the results from the tPSA measurements were compared to the results from the 4-biomarker class test.
- Lastly, to further refine the mutation detection, a ML-based mutation caller was developed and compared to a threshold calling model that is currently applied in the 4-biomarker class test.

Methods

- Methylation, aneuploidy, protein, and mutation measurements were performed as described previously (1-3) with the introduction of a newly developed ML-based overarching classifier that combines the methylation, aneuploidy, and protein results (**Fig. 2**). The mutation testing results were generated using a threshold-based mutation caller applied previously and combined with the 3 biomarker ML Classifier via OR-logic. The test performance of both biomarker combinations is shown in **Table 1**.
- A retrospectively-assembled, case-control feasibility study was set up to train, validate, and test the performance of two different marker combinations. The cancers were derived from all stages and up to 15 organ sites. The non-cancer control cohort was comprised of age-matched healthy individuals and an enriched fraction of samples from individuals with non-cancer diseases and benign tumors. The initial training and validation set (n=2386) included breast, bladder, colon, esophageal, kidney, liver, lung, ovarian, pancreatic, prostate, stomach, and uterine organ sites. The independent testing set (n=1259) included cancer from the same organ sites as well as three additional hematological cancer types: Non-Hodgkin's lymphoma, multiple myeloma, and myelodysplastic syndrome (**Figure 1**).
- Total and free PSA were quantified using the same high-throughput platform that was applied to quantify the other protein markers. PSA testing was performed on the testing set. A larger set of 834 cancers and 742 non-cancers was utilized to support the threshold setting study as well as to confirm the tPSA performance (data not shown).
- The development of a candidate ML-based caller for the mutation biomarker leveraged 186 young healthy donors processed in triplicate, as well as commercially available contrived ctDNA standards that carry well-documented mutations. The model was cross validated (4-fold) in the testing cohort and 85 cancer and 47 non-cancer cases not contained in the 4-biomarker testing set. During cross validation benign tumors were excluded from training but included in testing.

Results

Feasibility study design of a 3 and 4 Biomarker class test and resulting improved test performance

Figure 1

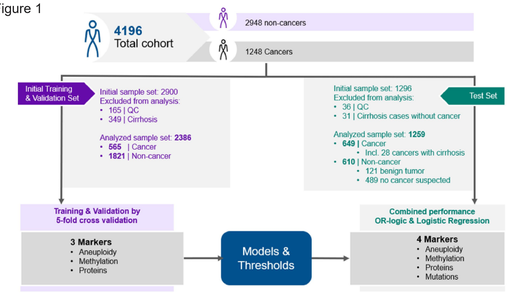
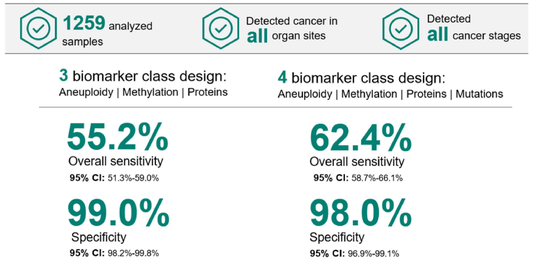
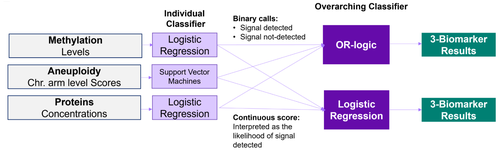


Table 1



Improved overarching classifier to combine individual methylation, aneuploidy and protein classifiers

Figure 2



- Methylation, aneuploidy and protein results were assessed using **individual classifiers** resulting in either a **binary call**, i.e. detected / not detected or a **continuous score**, i.e. a value between 0 and 1 that is be interpreted as the likelihood of a cancer-specific signal.
- The results from the individual classifiers were combined using two different **overarching classifier** approaches:
  - **OR-logic,** i.e. if any of the individual binary calls has a "detected" status the final call will also be "detected".
  - **Logistic Regression (LR),** i.e. the aggregate score produced by each biomarker classifier was used as features to train a new logistic regression model, which assigns a final call of "signal detected" or "signal not detected".

Table 2

3-Biomarker	OR-logic	Logistic Regression
Sensitivity	**53.9%** 95% CI: 50.1%-57.8%	**55.2%** 95% CI: 51.3%-59.0%
Specificity	**98.7%** 95% CI: 97.8%-99.6%	**99.0%** 95% CI: 98.2%-99.8%
False positives	**8 / 610**	**6 / 610**

- The two overarching classifier approaches (**Fig. 2)** were applied to the testing set (n=1259) shown in **Fig. 1**.
- The LR-based approach resulted in a mean sensitivity of 55.2% for detection of cancer signal: a 1.3% increase compared to the OR-logic (**Tab. 2**).
- Two of eight false positives reported by the OR-logic model were eliminated by utilizing the overarching classifier based on LR (**Tab. 2**).
- The results from the 3-Biomarker LR classifier were then combined with an OR-logic mutation detection caller using a threshold calling model, resulting in a small gain of sensitivity (**Tab. 1**).

Evaluation of tPSA a marker for prostate cancer (PCa) detection in combination with a 4-Biomarker class test design

Figure 3a

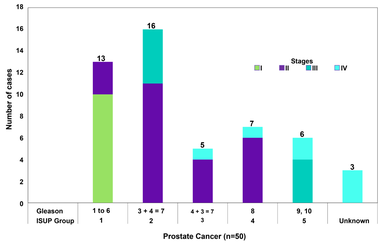
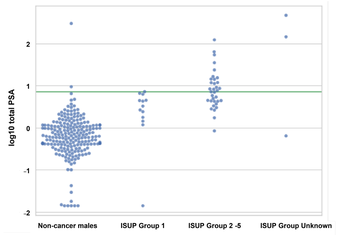


Figure 3b



- **Figure 3a** shows the distribution of PCa cases (n=50) utilized for threshold setting by stage and scores according to Gleason and the International Society of Urological Pathology (ISUP).
- **Figure 3b** shows the log tPSA concentrations of non-cancer male cases (n=214) as well as 50 PCa cases across different ISUP groups.
- To limit detection of low-grade PCa cases (ISUP 1) a tPSA threshold was established using a 99% specificity target considering the concentration values of these male donors and the 13 PCa ISUP group 1 cases. Based on this specificity target a tPSA threshold of 7.2 ng/ml was defined for plasma collected in LBgard® tubes.
- Free PSA was not considered for subsequent evaluations (data not shown).

Evaluation of tPSA a marker for prostate cancer (PCa) detection (cont.)

Table 3

Characteristics of PSA study	
Cases with 4-Biomarker data	n = 1259 (649 cancers; 610 non-cancers)
PCa case w/ 4-Biomarker data	n=26
Established tPSA threshold	7.2 ng / ml

	PCa Sensitivity / Overall Specificity	
	PCa ISUP 1-5 n=26	PCa ISUP 2-5 n=14
4-Biomarkers without tPSA	19.2% (5/26) 98.0%	35.7% (5/14) 98.0%
tPSA alone	26.9% (7/26) 99.8%	50.0% (7/14) 99.8%
4-Biomarkers with tPSA	34.6% (9/26) 97.9%	64.3% (9/14) 97.9%
Cases with tPSA > 7.2 ng / ml		
Non-cancer cohort	0.2% (1/610) 0.7% (1/149) male only	
Cancer cohort without PCa	1.0% (6/623) 2.2% (6/273) male only	

- tPSA performance was also evaluated on the testing set (n=1259) including 26 PCa and 623 other cancer cases.
- The non-cancer group included one case with a tPSA above 7.2 ng/mL. The measured value was 298 ng/mL and could potentially be a true PCa case.

Evaluation of a LR-based mutation classifier

Table 4

Mutations Only		Threshold model	Logistic Regression model
Sensitivity		42.1% (273/649)	42.4% (275/649)
Specificity	No cancer suspected	99.2% (485/489)	99.6% (487/489)
	Benign tumors	97.5% (118/121)	97.5% (118/121)
	Total non-cancers	98.9% (603/610)	99.2% (605/610)

- Mutation calling was performed using a rules-based (threshold) calling model. A new LR model was evaluated in order to improve the mutation calling sensitivity and specificity.
- 4-fold cross-validation was performed for the new LR model as outlined in the methods section. Subsequently, the performance was compared between the LR and threshold-based models.
- As shown in **Tab. 4**, the number of false positives was reduced in the no cancer suspected group. The LR calling model resulted in a 1.5% sensitivity improvement for Stage I & II cancers (data not shown).
- Further evaluations and refinement will be performed before including LR mutation calling model into the 4-Biomarker class test.

Conclusions

- Performance of an aneuploidy, DNA methylation, mutation, and protein biomarker class test that is intended for multi-cancer early detection was further improved.
- Inclusion of a ML classifier (ensemble-stacking approach) to combine the results from three individual biomarker classes is one of the contributing factors resulting in improved sensitivity and specificity.
- Total PSA at a newly established threshold was evaluated for its ability to improve the detection of clinically actionable prostate cancers. The performance was also assessed in combination with the 4 Biomarker class test.
- A new mutation caller algorithm, also based on a ML classifier, was established and evaluated. Mutation detection specificity and sensitivity, particularly for early-stage cancers, improved slightly. More rigorous testing will be required prior to integrating this approach into the overall 4-biomarker class calling algorithm.
- Additional approaches, such as feature optimization and engineering, as well as non-linear models, did not yield additional improvements over the existing methylation caller.
- As more comprehensive case control data is becoming available these approaches will be extensively re-evaluated to establish a robust classifier that will withstand the true test in a real-world average risk screening trial.

References and Acknowledgements

1. Douville C, Nobles C, Hwang HJ, et al. Multi-cancer early detection through evaluation of aneuploidy, methylation, mutation, and protein biomarkers in plasma [poster]. ESMO 2022. September 9-13, 2022. Paris, France.
2. Douville C, Cohen JD, Ptak J, et al. *Proc Natl Acad Sci U S A*. Mar 3 2020;117(9):4858-4863.
3. Cohen JD, Li L, Wang Y, et al. *Science*. Feb 23 2018;359(6378):926-930.

Medical writing and editorial support was provided by Carolyn Hall, PhD, and Feyza Sancar, PhD (Exact Sciences, Madison, WI). This study was sponsored by Exact Sciences Corp., Madison, WI.