Robust bioinformatic method for estimating tumor content in cfDNA using denoised target sequencing

Pillar Biosciences, Inc. 9 Strathmore Rd Natick, MA 01760
(800) 514-9307 | www.pillarbiosci.com

Andrew B. Conley, Lavanya Rishishwar, Nachen Yang, Michael Liu, Yi Cao, Tejashree Modak, Jordan Aldersley, Sean Polvino, Yue Ke, Kirsty McFarland, Zhaohui Wang, Alex V. Kotlar

ABSTRACT

Introduction. Minimal residual disease (MRD) testing quantifies the presence of cancer in post-treatment patients, playing a crucial role in improving survival. By monitoring a patient's circulating tumor DNA (cfDNA), MRD testing can identify cancer presence before radiologic recurrence, providing early detection and intervention opportunities. Effective MRD testing relies on robust bioinformatic analysis, capable of accurately distinguishing low-frequency tumor molecules from technical noise. In this study, we introduce a robust method for estimating tumor content (TC) from targeted sequencing data obtained from fixed and personalized panels, integrated into PiVAT® (Pillar Biosciences Variant Analysis Toolkit), Pillar's genome analysis software.

Methods. We developed a maximum likelihood method for TC estimation paired with bootstrapping to ascertain statistical confidence of the estimates. This algorithm is integrated in and leverages existing quality control and refinement steps of PiVAT®. We tested dilutions from two different SeraCare® standard positive control samples, 31 normal samples as well as on clinical patient samples from a retrospective ovarian cancer study. Dilutions of positive samples were performed from 0.1% to 0.003125%, halving the concentration in each dilution, for a total of 34 samples. The positive controls were run on two fixed gene panels: a small panel targeting 10-13 variants, and a larger panel targeting 24-26 variants within the two positive samples. For the retrospective study, 73 banked plasma and whole blood samples from 10 ovarian cancer patients (5 recurrence cases) were analyzed using tumor-informed panels, with up to 2 years of follow-up visits.

Results. We correctly called the MRD status of almost all the diluted standard samples, with strong bootstrap support. We failed to call 2 & 3 of the 4 replicates at 0.003125% & 0.00625% respectively, owing to <4 positive variants identified in the samples. All the normal samples were correctly predicted as MRD negative. In the retrospective study, we correctly identified MRD+ samples 167 days before clinical diagnosis of recurrence.

Conclusion. Our findings highlight the effectiveness of Pillar's new tumor content estimation algorithm in accurately estimating tumor content and predicting MRD status in post-treatment cancer patients using targeted sequencing data. By successfully distinguishing low-frequency tumor signals from technical noise, this approach offers a valuable tool for early detection and intervention. The encouraging results obtained from the retrospective ovarian cancer study emphasize the potential of PiVAT® in predicting cancer recurrence months in advance. Larger studies would be needed to establish the clinical utility of this approach.

EXPERIMENT DESIGN

Two experiments:

1. Testing on standard reference positive and normal samples
2. Retrospective analysis of 10 ovarian cancer patients

Experiment #1: Standard reference samples

- **DNA Input:** 10-66ng
- Samples tested on a small panel (targeting 10-13 variants in CMM and MMv2 respectively) and a larger panel (24-26 variants in CMM and MMv2 respectively).

	Sample	Expected TC (%)	DNA Input (ng) 10	20	30	66	Total
Normals	AM	0	1	2			**3**
	Healthy donors	0	12	6	6		**24**
	GIAB	0		1	1		**2**
	SCC WT	0			2		**2**
Positive samples	Seraseq® ctDNA Complete™ Mutation Mix (CMM); *# variants =10-24*	0.003125			4		**4**
		0.00625			4		**4**
		0.0125			4		**4**
		0.025			4		**4**
		0.05		5	9	5	**19**
		0.1			4		**4**
	Seraseq® ctDNA Mutation Mix v2 (MMv2) *# variants =13-26*	0.05			5		**5**
		0.1		5			**5**
Total		**Normals**	**13**	**9**	**9**	**0**	**31**
		Positives	**0**	**10**	**34**	**5**	**49**

Table 1. Summary of samples used in the study.

Experiment #2: Retrospective study

- 10 ovarian cancer patients, (5 recurrence, 5 non-recurrence cases)
  - Plasma and whole blood samples were stored in a biobank prior to the study
- MRD panel designed for each patient using WES data
- Average days for follow-up visit = 701 days; average blood samples (per patient) = ~7 samples
- Total samples collected = 73

EXPERIMENT DESIGN

Table 2. Retrospective study cohort summary. Observed characteristics of the cohort of 10 ovarian cancer patients whose samples were analyzed in this study.

Parameter	Metric	Range	Parameter	Metric	Range
Number of patients	Total	10	Follow-up time (days)	Mean	701
	Recurrence	5		Min / Max	191 / 1,238
	Non-recurrence	5		Timepoints	73
Onset age	Min / Max	41.0 / 67.0	Prior treatment of CA125 (U/ml)	Min / Max	11.7 / 6030.2
	Med [IQR]	53.0 [47.5;62.0]		Med [IQR]	2306.4 [469.8;3681.5]
	Mean (std)	54.1 (8.7)		Mean (std)	2407.6 (2020.3)
Stage	I~II	3	Prior treatment of HE4	Min / Max	0 / 948.3
	III~IV	7		Med [IQR]	319.9 [188.0;627.6]
BMI (kg/m2)	Min / Max	20.5 / 31.6		Mean (std)	401.4 (310.5)
	Med [IQR]	23.6 [22.1;26.2]	CA199 prior to treatment	Min / Max	0 / 25.5
	Mean (std)	24.6 (3.4)		Med [IQR]	6.9 [3.7;11.3]
Family history	No	7		Mean (std)	8.2 (7.1)
	Yes	3	Fagotti score	Min / Max	0 / 10.0
Max tumor size (cm)	Min / Max	3.5 / 30.0		Med [IQR]	6.0 [1.0;8.0]
	Med [IQR]	6.0 [5.0;9.0]		Mean (std)	4.7 (4.0)
	Mean (std)	8.9 (7.6)	PCI score	Min / Max	2.0 / 25.0
Neoadjuvant chemotherapy	No	7		Med [IQR]	14.0 [4.0;16.5]
	Yes	3		Mean (std)	11.9 (7.9)

IQR: inter-quartile range; std: standard deviation; CA125: protein cancer antigen 125, blood biomarker; HE4: human epididymis protein 4, blood biomarker; CA199: protein cancer antigen 125, blood biomarker; Fagotti score: likelihood score for predicting optimality of cytoreductive surgery; PCI score: numerical score to capture the extent of tumor growth

RESULTS

Maximum Likelihood (ML) estimation of tumor content

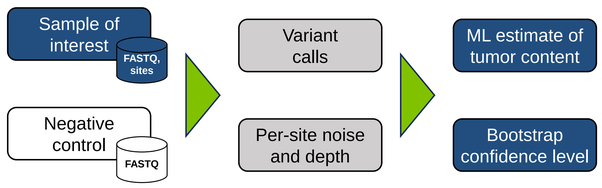

Figure 1. Overview of PiVAT's tumor estimation algorithm. To estimate the tumor content of a sample, PiVAT requires the amplicon sequencing data (FASTQ), the expected site information (sites), and, optionally, a negative control sample. Availability of negative control sample was found to improve the calling performance. The samples are processed using PiVAT to generate variant calls and per-site noise and depth information. The tumor estimation algorithm utilizes this information to determine the tumor content range that maximizes the likelihood of explaining the observed data. Further, bootstrapping analysis is also performed to add a confidence level to the observed estimate. A positive MRD result is supported by a significant bootstrap confidence value.

Performance Summary with Standard Reference Samples

Table 3. Summary performance of MRD calling with standard reference calling. The negative percent agreement (NPA) and positive percent agreement (PPA) across all standard reference sample is shown. PiVAT performs well for most of the samples, with 2 false negatives in the 0.003125% (~31 ppm) and 0.00625% (~62 ppm).

NPA = 100 - 100 * FP / (FP + TN)
PPA = 100 * TP / (TP+FN)

Metric	Expected TC (%)	DNA input (ng) 10	20	30	66	Average
NPA	0	100%	100%	100%	NT	100%
PPA	0.003125	NT	NT	75%	NT	75%
	0.00625	NT	NT	75%	NT	75%
	0.0125	NT	NT	100%	NT	100%
	0.025	NT	NT	100%	NT	100%
	0.05	NT	100%	100%	100%	100%
	0.1	NT	100%	100%	NT	100%
	Overall	NT	100%	94%	100%	96%

RESULTS

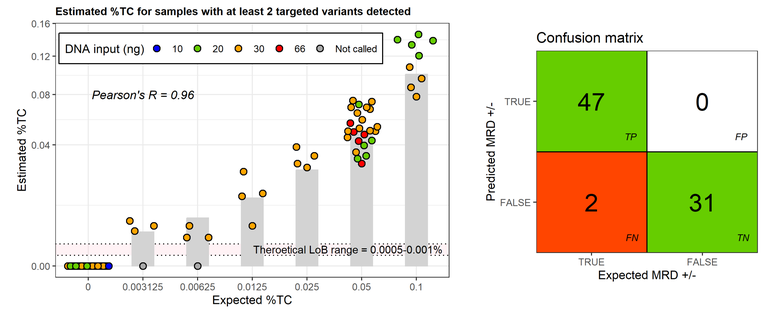

Figure 2. Observed tumor content estimates for standard reference samples from PiVAT. (left) Observed tumor content as estimated from PiVAT are regressed against expected tumor content (TC). The Circles are color coded based on the DNA input. Two false negative results are observed in low input TC samples, owing to detection of only a single variant in each. (right) Confusion matrix summarizing the performance of PiVAT across the standard reference samples.

Performance Summary with Retrospective Study

Table 3. Patient specific timepoints collected over the retrospective study. The table shows per-patient panel size (number of variant targeted), follow-up time (in days), intervention timepoints recorded, and number of blood draws performed ("sampling"). The patient ID are ordered identical to Figure 3.

	Patient ID	Panel size	Follow-up time (days)	Timepoints Intervention	Timepoints Sampling
Recurrence	002	30	540	3	6
	211	28	544	2	6
	222	28	842	3	6
	207	27	657	2	6
	213	28	976	3	12
No recurrence	205	30	191	2	3
	210	29	322	2	4
	212	24	700	2	5
	204	28	997	2	10
	217	26	1,238	3	15
	Average	**28**	**701**	**Total**	**73**

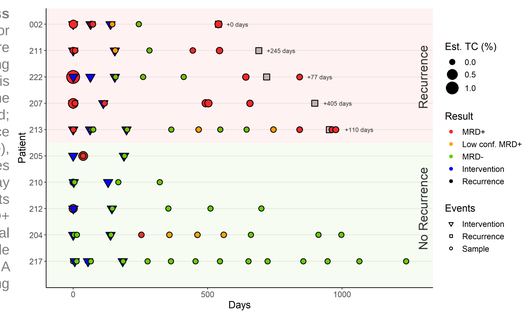

Figure 3. Tumor content estimates across timepoints in the retrospective study. Tumor estimates for each patient at each timepoint are shown as circles plotted across the tracking duration for the patient. The size of the circle is scaled to the estimated TC %. The color of the circle represents a significant MRD+ call (red; significant bootstrap confidence), low confidence MRD+ call (orange; low bootstrap confidence), and MRD- (green). Blue inverted triangles represent recorded intervention events and gray square represent clinical recurrence. For patients with clinical recurrence (pink background), MRD+ calls were observed up to 405 before clinical diagnosis. For all but one non-recurrence sample group, MRD- calls are made after intervention. A possible intervention point was missed being recorded for sample 204.

CONCLUSIONS

- We introduce in this study a new maximum likelihood method for estimating tumor content in MRD samples from amplicon sequencing data and expected variant sites
- In contrived samples, our approach able to detect tumor sequence below 0.01% VAF with high specificity.
- We demonstrate in this study that with the combination of our chemistry and algorithm, we can accurately call MRD status in a clinical retrospective study.
- Larger studies would be needed to establish the clinical utility of this approach.