PS4-02-27

Machine learning-enhanced ultra-deep sequencing for low-abundance circulating tumor DNA (ctDNA) in breast cancer

Burning Rock Dx

SABCS 2025 | Dec. 9-12, 2025
San Antonio, TX, USA

Xiaoling Li*[1], Yuwei Ni*[1], Si Zhang*[1], Chen Yang[1], Guangyou Li[1], Yunpeng Zhang[1], Xianrong Chen[1], Xingyu Yang[1], Jing Su[1], Bingsi Li†[1]

[1] Innovation Center, Burning Rock Biotech, Shanghai, China. * These authors contributed equally † Corresponding author

BACKGROUND

- Although traditional tissue biopsies provide valuable diagnostic insights, their invasive nature makes them unsuitable for long-term monitoring. Liquid biopsy, which analyzes circulating tumor DNA (ctDNA) in blood samples, offers a minimally invasive alternative with growing potential in breast cancer care.
- To develop an ultra-deep version of targeted sequencing for low-ctDNA scenarios, blood samples were collected from non-cancer controls and breast cancer patients.
- An ultra-deep targeted sequencing was established for low-ctDNA detection that use targeted high throughput hybridization based capture technology for detection of single nucleotide variants (SNVs), insertions and deletions (Indels), copy number variants (CNVs) and fusions in 168 genes respectively.

METHODS

Sample preparation and Data collection

- A UMI-aided ultra-deep sequencing assay with targeting panels for 168 pan-cancer genes (raw depth 35,000X, mean unique molecular depth 4000X at 20ng input, NovaSeq6000 platform) were employed in this study.
- Plasma cfDNA (20–60 ng) and WBC genomic DNA (100 ng) were used for performance evaluation.
- Limit of Blank (LoB) was defined from 51 cancer-free donors with paired WBC controls.
- Limit of Detection (LoD) were determined from a dilution series of contrived cell line samples (tumor allele frequencies(TAF): 0.25%, 0.125%, and 0.0625%. A pool of contrived cell line samples containing a total of 30 unique variants, including SNVs and Indels, were used.
- Accuracy study was compared to that of an independently validated NGS assay[1] using clinical samples from 44 clinical samples derived from breast cancer patients with an expanded panel.

Data Processing and Modeling

- The data processing pipeline incorporated the following steps to suppress technical and biological noise:

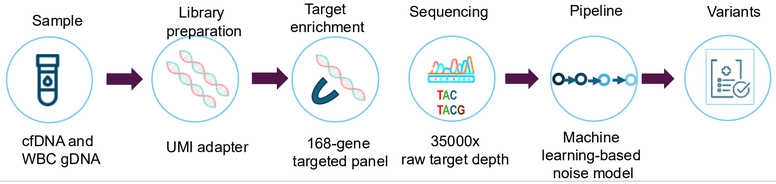
(1) Reads with identical UMIs and endpoints position were used to collapse and merge to generate high-accuracy consensus sequences, effectively remove random errors introduced during PCR amplification and sequencing .

(2) A machine-learning-driven noise model was developed to systematically identify and filter recurrent artifacts, including DNA damage patterns and sequencer-specific error signatures[2].

(3) To distinguish genuine somatic mutations from non-cancer biological aberrations (e.g. clonal hematopoiesis), paired WBC-sequencing was applied[3].

Figure 1. The workflow of ultra-deep sequencing and machine learning.

The sample was sequenced with our ultra-deep 168 targeted panel incorporating UMIs. Raw sequencing data were processed through our in-house bioinformatics pipeline. Technical and biological noise were suppressed by Machine learning-baseline noise model. From the analyzed data, we extracted SNVs, indels, CNVs, fusions and QC metrics.



RESULTS

Limit of Blank (LoB) study

- The LoB was defined for both the SNVs and indels across the entire gene panel. This panel covers whole coding regions of key breast cancer-associated genes, such as PTEN, ESR1, PIK3CA, BRCA1, BRCA2 and PALB2.
- This study demonstrated a near-zero false-positive rate across the entire assay reportable range(~10^{-7}), as shown in Table1.

Table 1. LoB Study Summary Results

Category	False Positive Rate
Panel-wide SNVs & Indels (400K)	$4.88*10^{-7}$ (Per Base)
Panel-wide CNSs (22 genes)	0 (Per Gene)
Panel-wide Fusion (3 genes)	0 (Per Gene)

Limit of Detection (LoD) study

- The median LoD, defined as the median value of the LoDs for SNV or Indel variants, was rigorously assessed across different variant classes at defined input masses.
- For panel-wide SNVs, the median LoD is 0.25% with a 20ng input, and this threshold could be further reduced to 0.15% with a 50ng input, as shown in Table 2.
- For panel-wide indels, the median LoD improved from 0.25% at a 20ng input to 0.14% at 50ng input, as shown in Table 2.
- Each variant's LoD was established using contrived samples and defined as the lowest TAF that achieved a ≥95% detection probability through Probit regression.
- Each contrived sample was processed with 20 replicates at both low and high cfDNA input amounts.

Table 2. The Median LoD determination at low and high cfDNA input in contrived cell line samples

Category	Raw depth	Median Lod,20ng (AF)	Median Lod,50ng (AF)
Panel-wide SNVs	35,000X	0.25%	0.15%
Panel-wide Indels	35,000X	0.25%	0.14%

Accuracy Study

- The detection of alterations was performed on 44 clinical breast cancer patients with expanded panel.
- The respective Positive Percent Agreement (PPA) of SNVs and indels was 83.0% and 85.7%. The Negative Percent Agreement (NPA) of SNVs and indels was both >99.9% (Table 3).

Table 3. Concordance Study Summary

Alternation Type	OncoCompass Target(+) Comparator(+)	OncoCompass Target(+) Comparator(-)	OncoCompass Target(-) Comparator(+)	OncoCompass Target(-) Comparator(-)	Possible Variants (n)	Patients (n)	PPA (95% CI)	NPA (95% CI)
Panel-wide SNVs	39	22	8	17379755	394996	44	83.0% (69.1%-92.4%)	99.9% (99.9%-100%)
Panel-wide Indels	6	0	1	17379817	394996	44	85.7% (42.1%-99.6%)	99.9% (99.9%-100%)

CONCLUSIONS

- In this study, we demonstrated that the accuracy and consistency of the advanced ultra-deep mutation sequencing technologies.
- Combined with machine-learning tools, this approach offers a practical, non-invasive diagnostic tool for mutation profiling, applicable even when tumor levels are low.

References

1. Deveson, I.W., Gong, B., Lai, K. et al. Nat Biotechnol 39, 1115–1128 (2021).
2. Newman AM, Lovejoy AF, Klass DM, et al. Nat Biotechnol. 2016;34(5):547-555.
3. Razavi P, Li BT, Brown DN, et al. Nat Med. 2019;25(12):1928-1937.

Conflict of Interest

All authors are employees of Burning Rock Dx.

Acknowledgement

This study was funded by Burning Rock Dx. Study design and execution assistance was provided by Medical Affairs group and Operation group.

© 2025 Burning Rock Dx