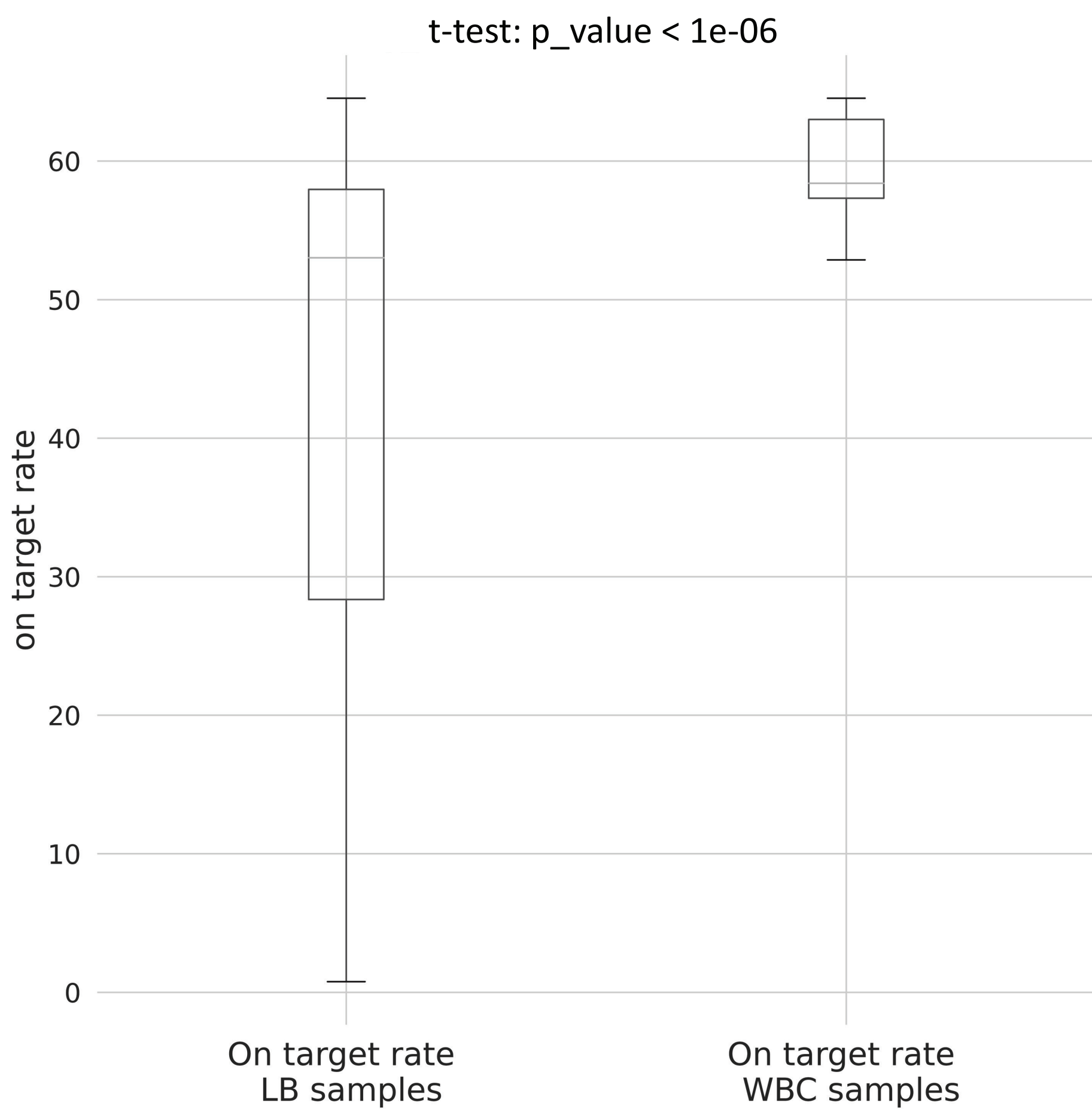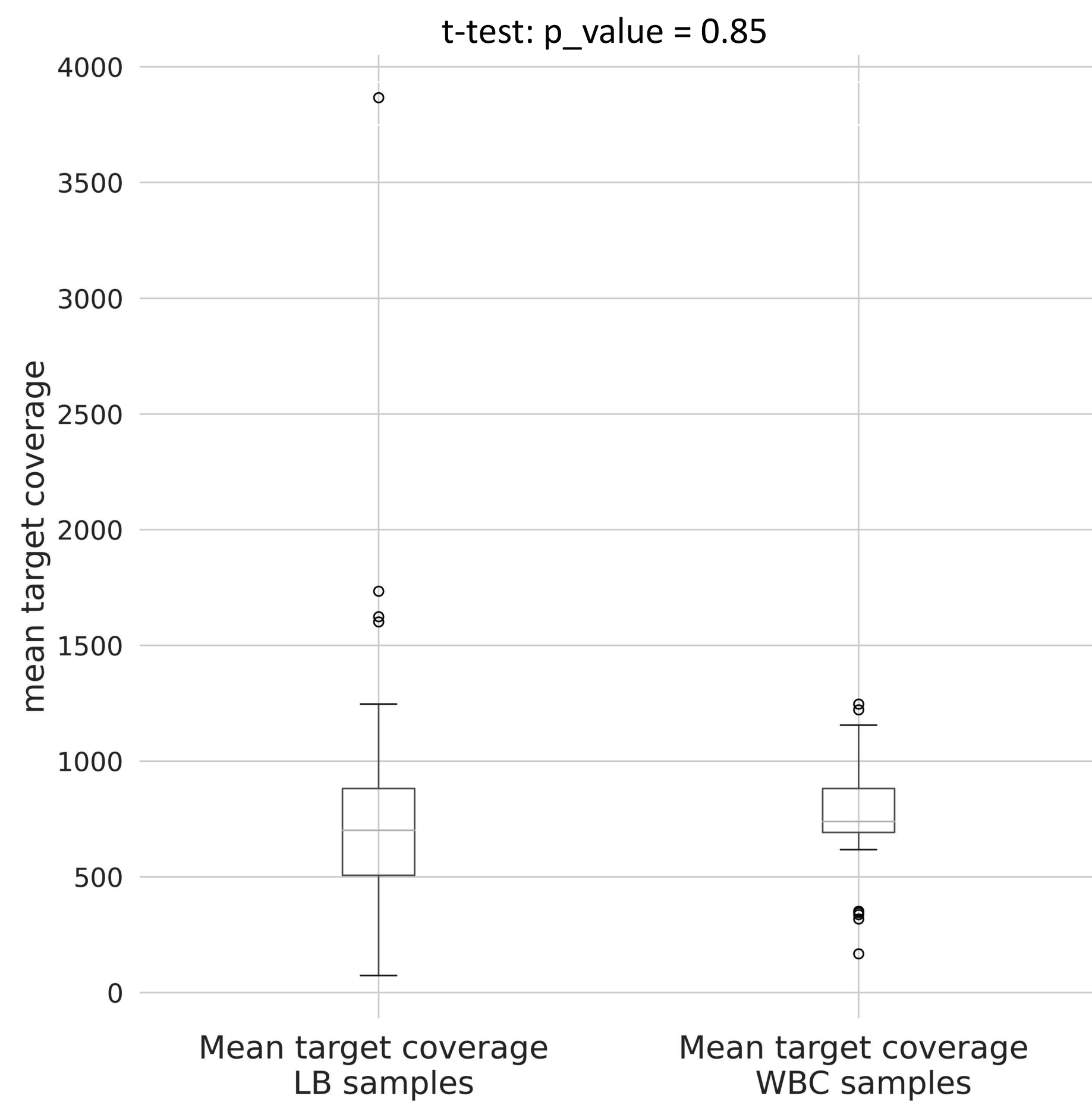**Fig. S1 Study design chart.** The discovery cohort consists of 55 patients with stage I, II and IIIA HCC and 55 healthy volunteers. Ultra-deep sequencing using a panel of the 13 genes found most often mutated among PwHCC were performed on WBC gDNA, FFPE-tumor tissue gDNA and plasma cell-free DNA to identify TDMs and the challenges in classifying the two groups of patients. Mutation fragment length profiles were used as input features to build a machine learning model. The model's performance was subsequently validated in an independent cohort of 55 HCC patients and 53 healthy individuals recruited from a different site.
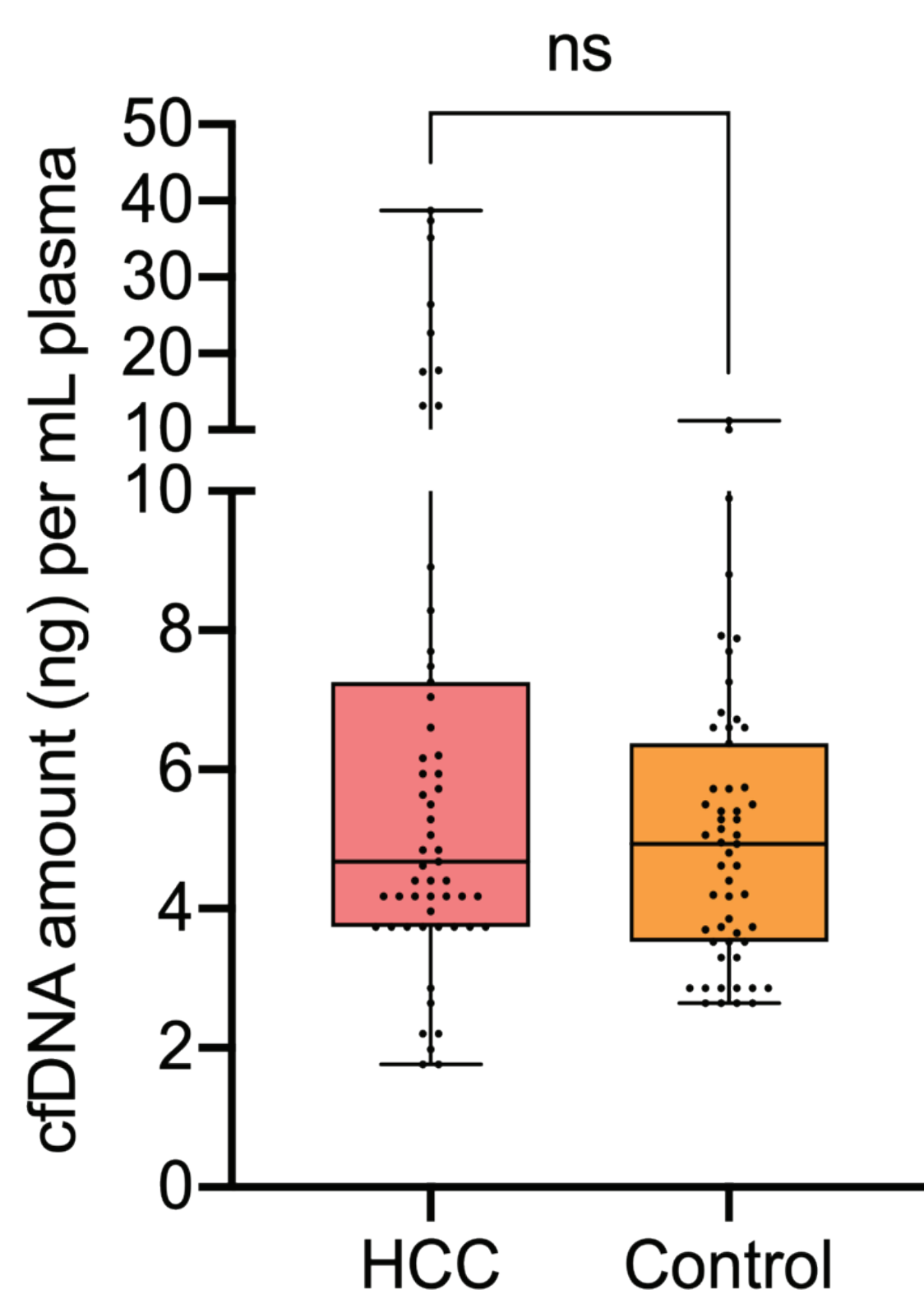
A

t-test: p_value < 1e-06

on target rate

On target rate
LB samples

On target rate
WBC samples

B

t-test: p_value = 0.85

mean target coverage

Mean target coverage
LB samples

Mean target coverage
WBC samples

C

cfDNA amount

ns

cfDNA amount (ng) per mL plasma

HCC   Control

Figure S2 Comparison of on target rate and depth coverage between cfDNA and paired WBC gDNA sequencing. **(A)** and **(B)** On target rate **(A)** and mean depth coverage **(B)** of paired liquid biopsy cfDNA and WBC gDNA samples from 55 HCC patients in the discovery cohort. (C) Bar graph showing the amount of cfDNA (ng per ml) of plasma from HCC and healthy control samples. ns: not significant, Mann–Whitney U test.
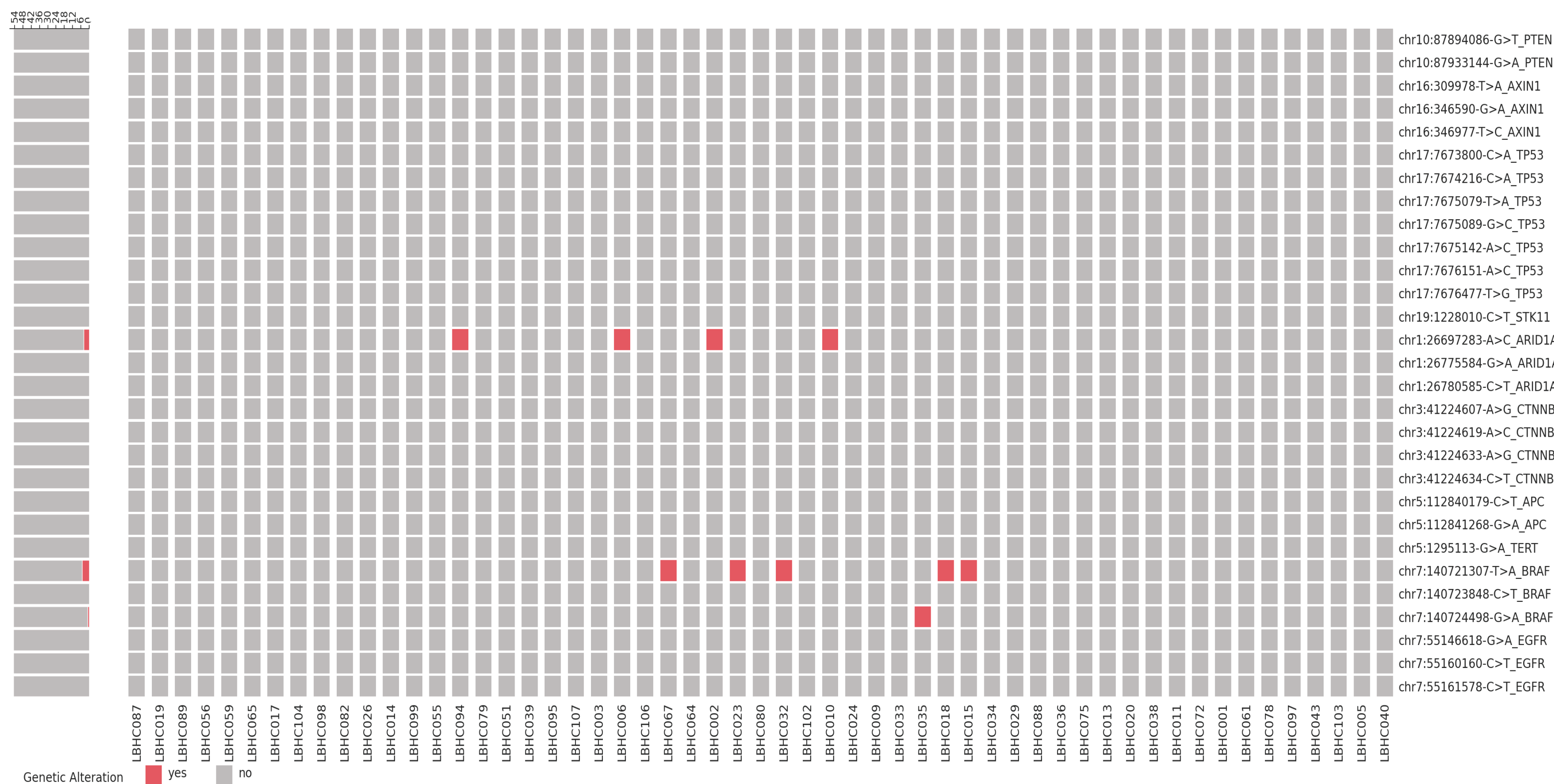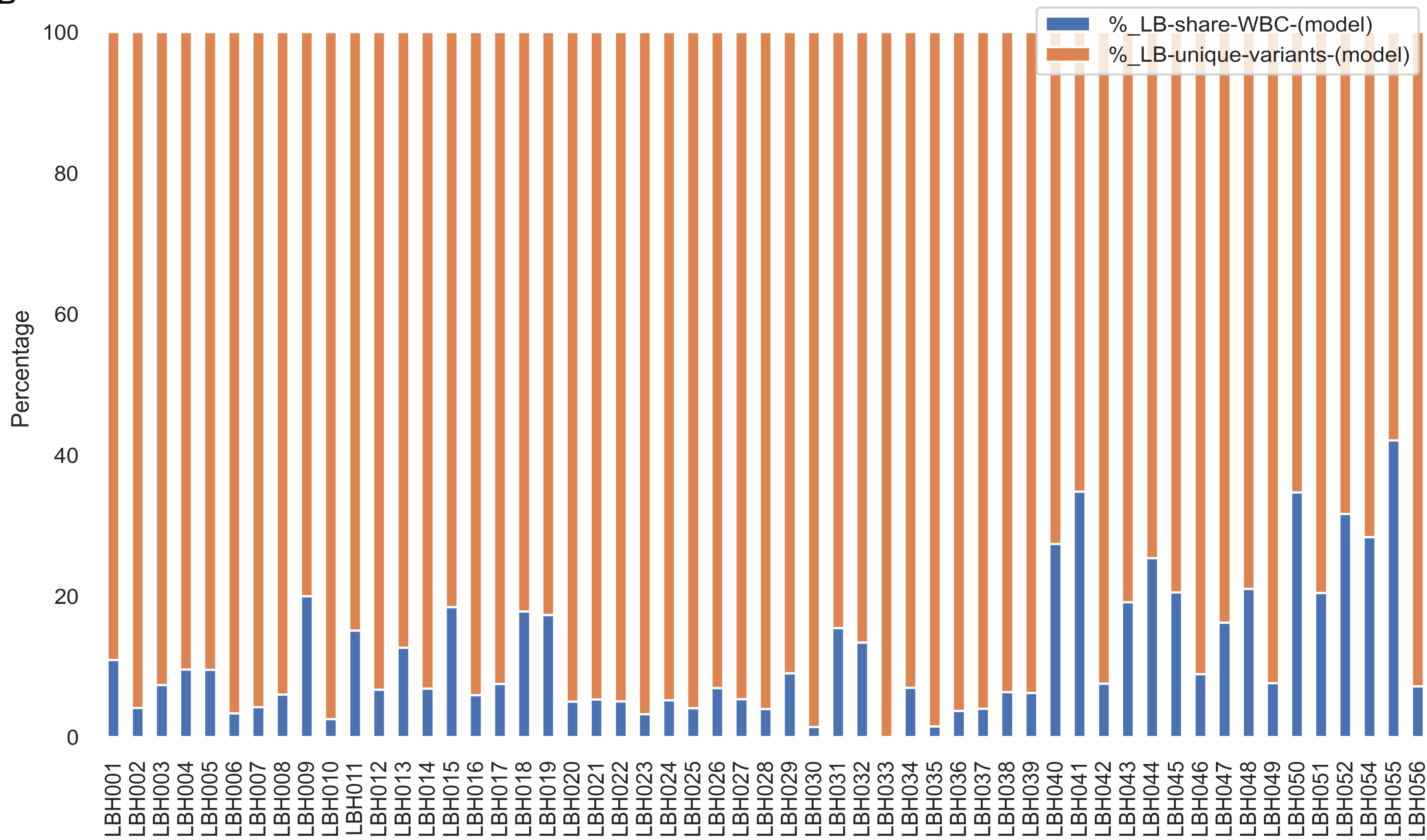
Fig. S3 Tumor-derived mutations in plasma samples of PwHCC overlapped with mutations detected in plasma of healthy participants. (A) Oncoprint plots of distributions of TDMs in 55 healthy individuals. Each row represents a TDM with mutation labeled on the right side and left side shows the occurrences of each mutation. Each column represents a patient. (B) Percentages of mutations shared between liquid biopsies and paired WBCs (LB-share-WBC, blue) and mutations uniquely found in liquid biopsy (LB-unique-variants, orange) for each patient after implementing our statistical model.
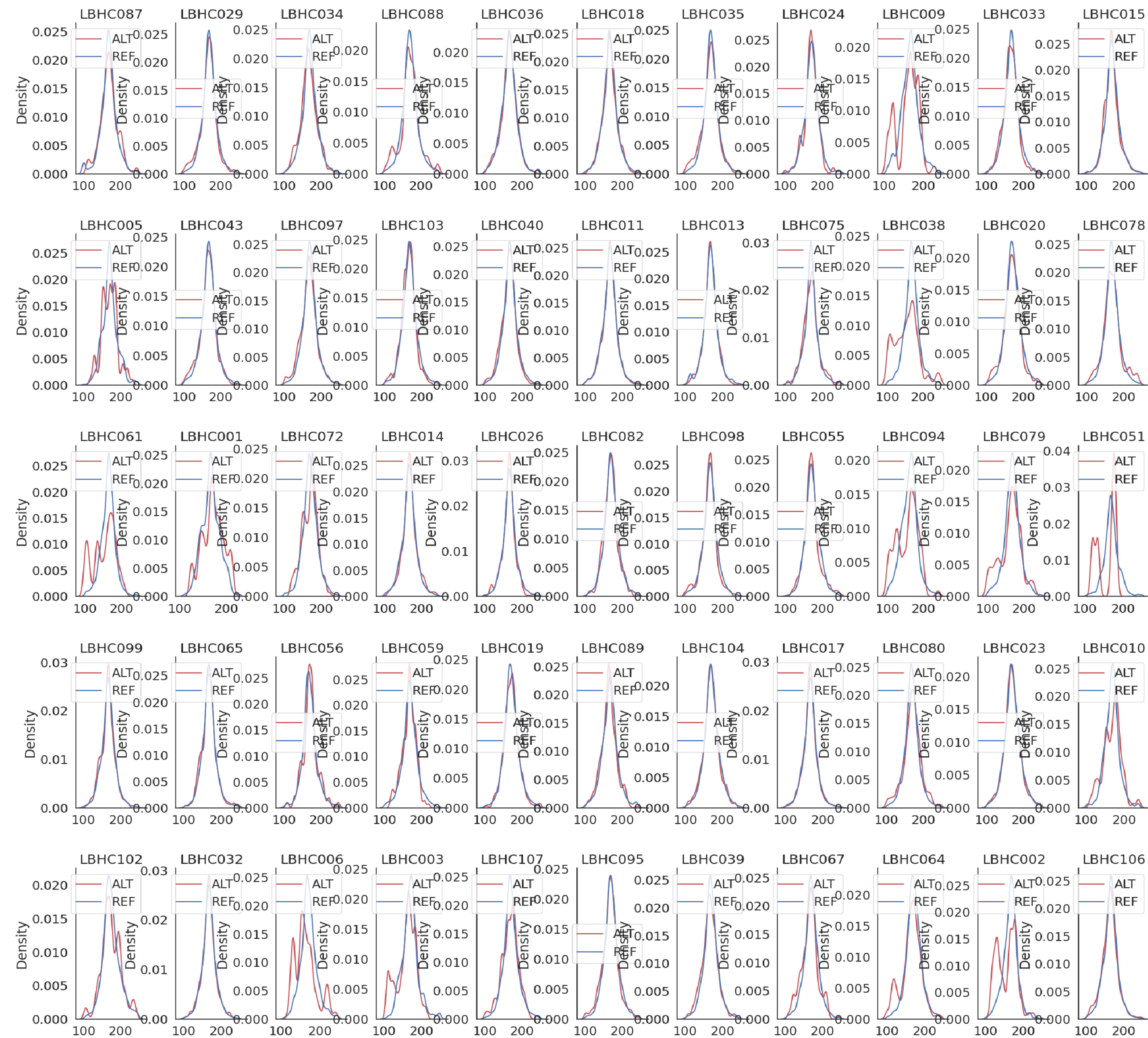
A

B
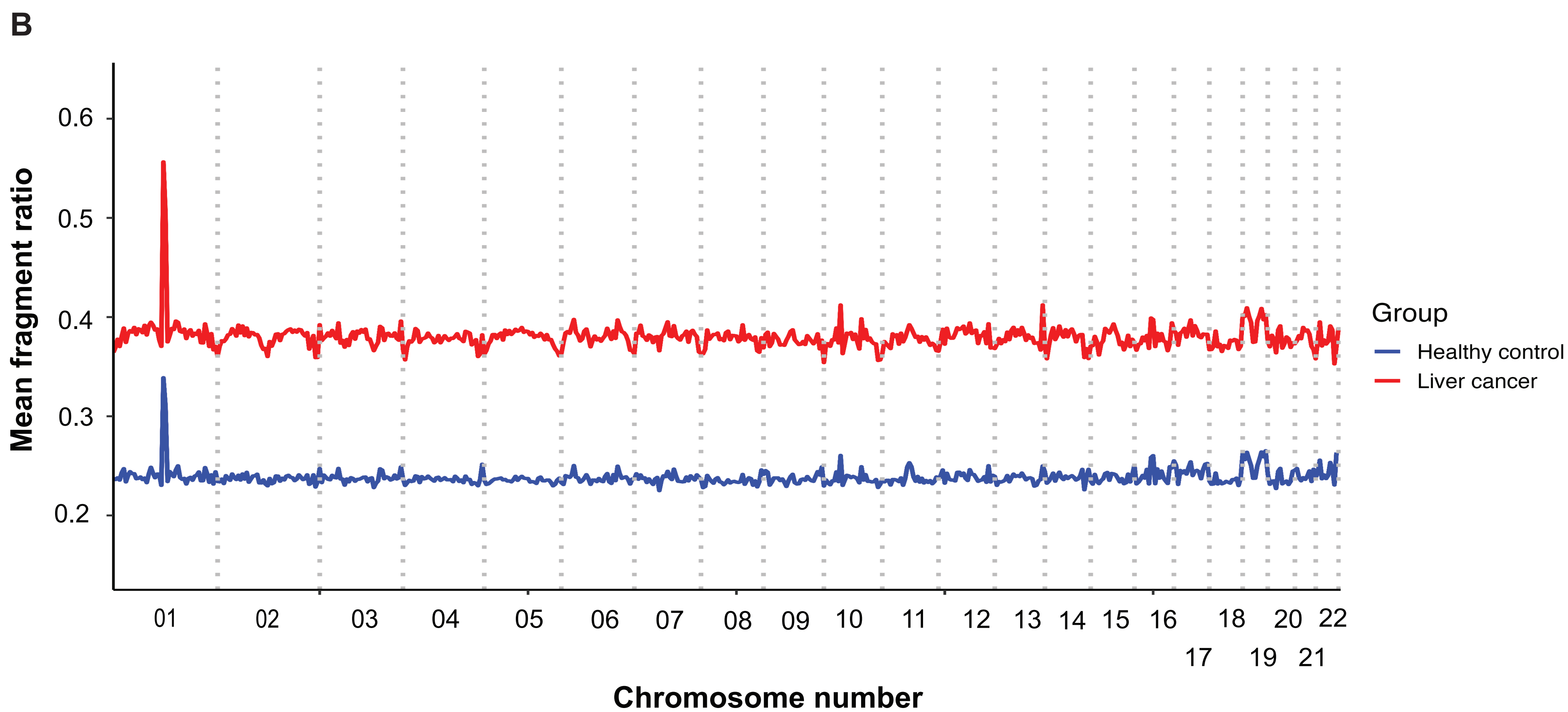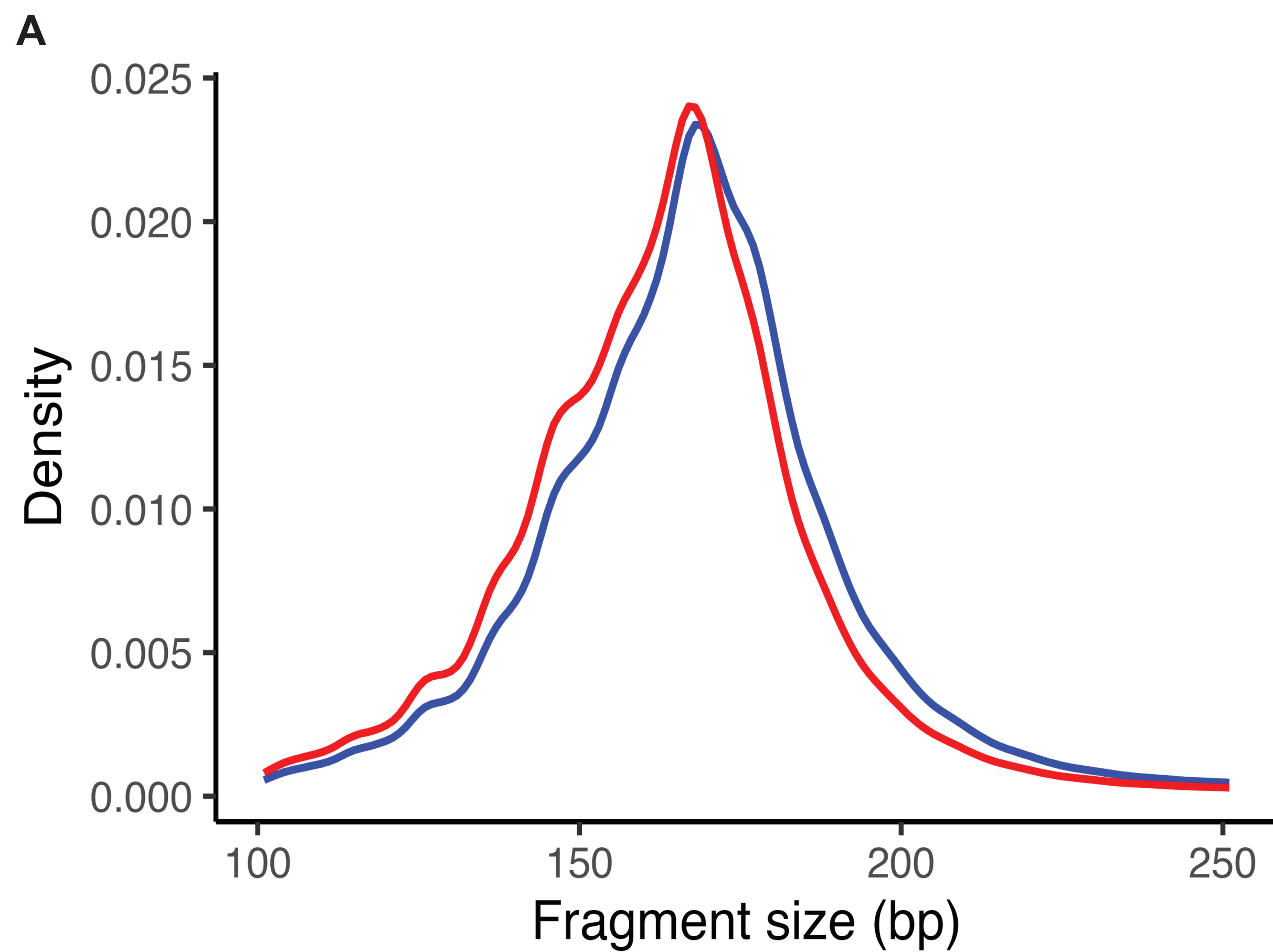
Fig. S4 PwHCC patients displayed cfDNA fragment length profiles distinct from healthy individuals. (A) Length distribution of cfDNA fragments in plasma samples of 55 PwHCC and 55 healthy individuals. **(B)** The mean ratio of short ($\leq$ 150 bp) to long ($>$ 150bp) fragments across 22 chromosomes at 5Mb resolution for 55 PwHCC and 55 healthy individuals.
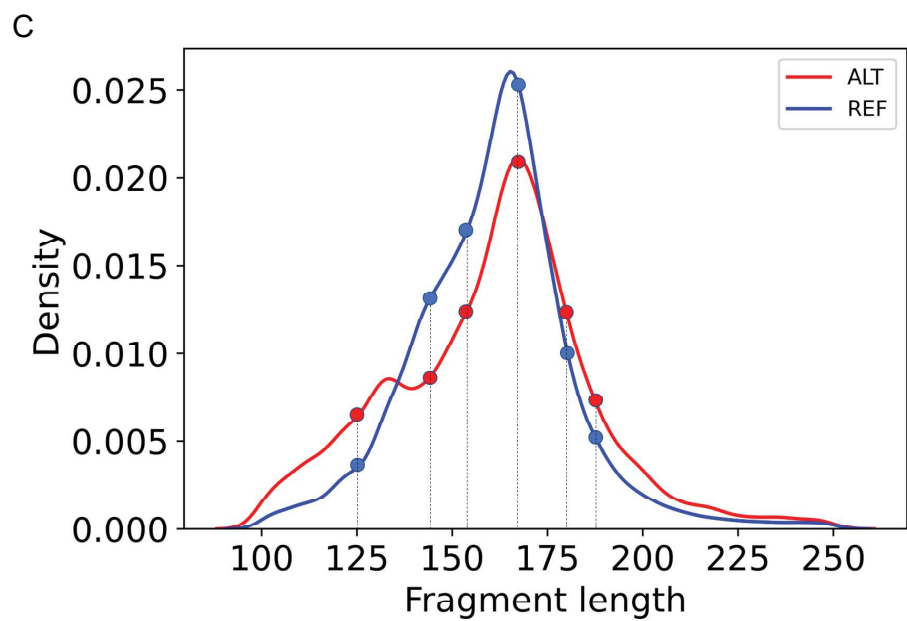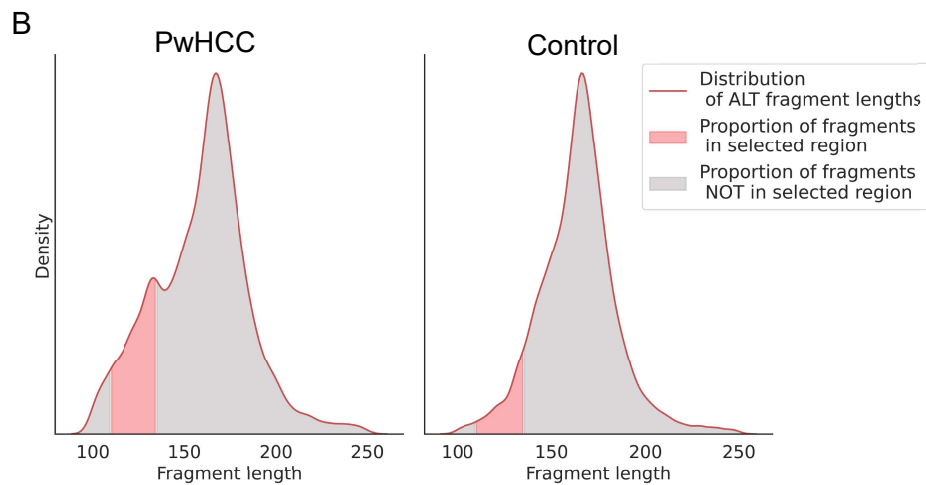
A
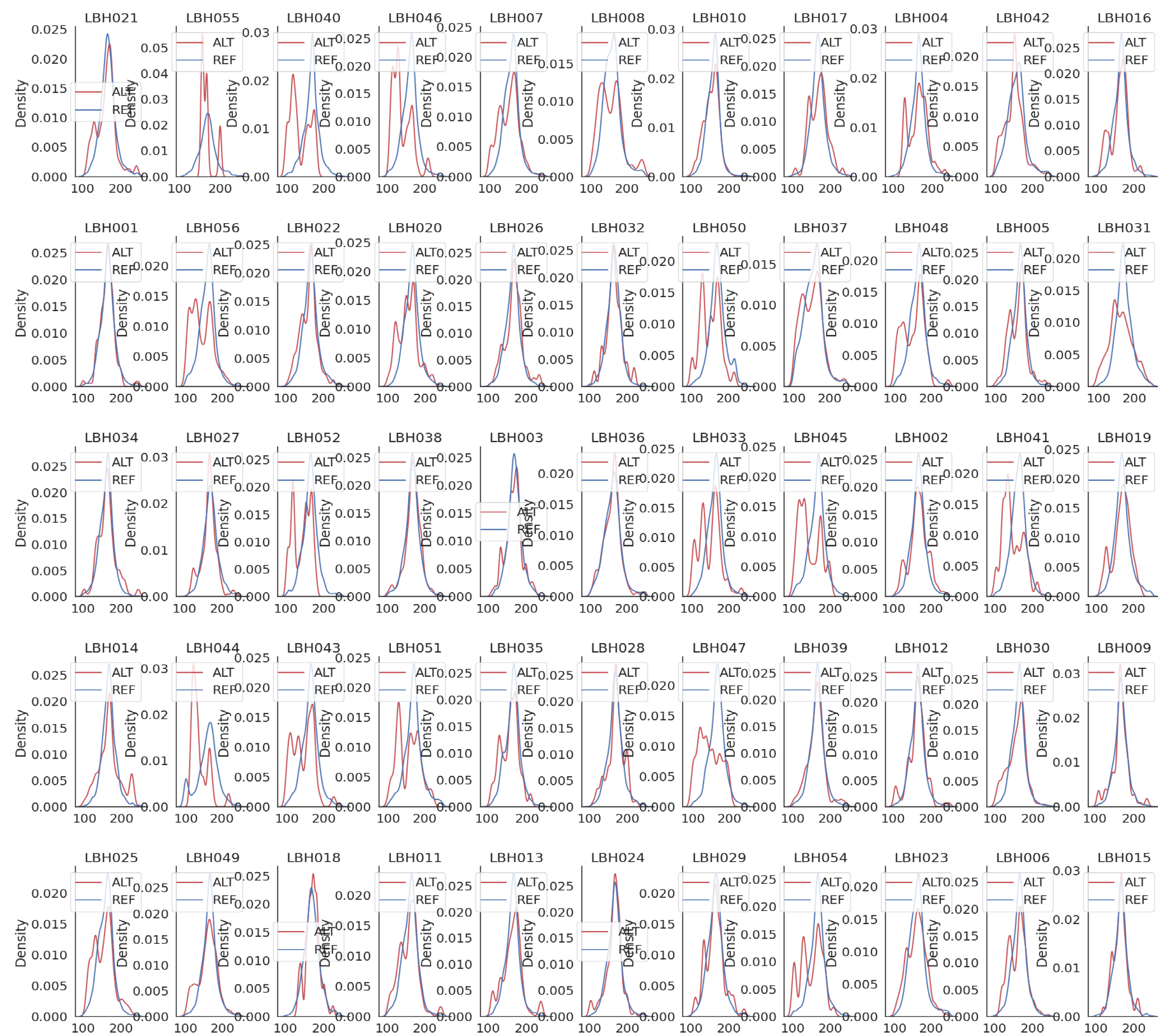
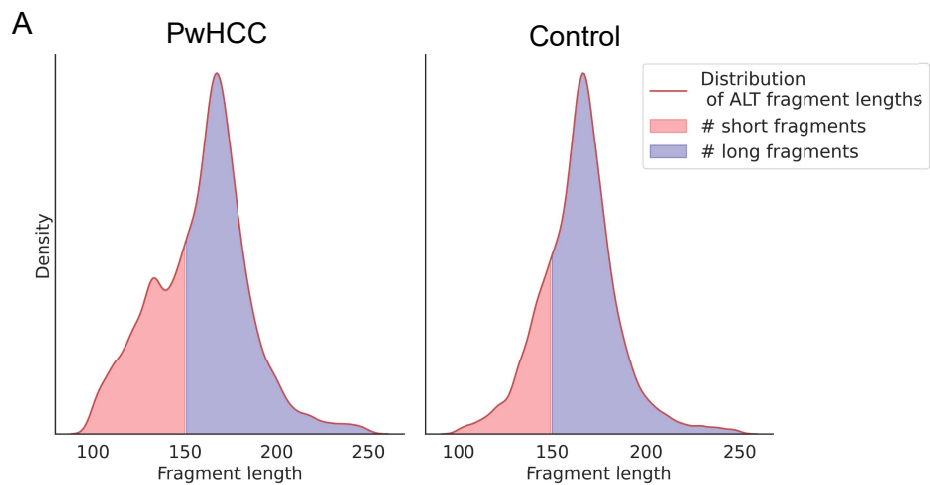B

Fig. S5 Distribution of fragment length of sequencing reads carrying LB-unique mutations (ALT) in all 55 PwHCC **(A)** and 55 healthy individuals **(B)**, as compared to their corresponding reference reads (REF) in the discovery cohort.

Fig. S6 Graphic explanation of three features generated from analyzing fragment length distribution of LB-unique mutations.

(A) Feature 1: The fraction of short-to-long reads that carry LB-unique mutations compared to their corresponding reference reads (RF) in PwHCC (left) and healthy controls (right). **(B)** Feature 2: Analysis of fragment length distribution of all LB-unique fragments enriched in specific regions (e.g., 110-135 bp). PHRED-scaled p-value obtained from Fisher's exact test used to compare the distribution of ALT-fragments in selected regions. **(C)** Feature 3: The probability of observing an ALT-fragment of size s (dot points) in all ALT fragments by proportions was calculated (λ) and the sum of $\lambda(s)$ over a sliding and non-overlapping window of 10 consecutive values of s was computed to yield a 15-dimensional feature vector.