



ARTICLE OPEN

Biallelic loss of RB1 in hepatocellular carcinoma as synthetic lethal target for artificial intelligence-guided therapy

Jihyun An¹, Wonkyung Kim^{2,3}, Xiumei Zhang⁴, Philip Chikontwe⁵, Jin-Sung Ju², Bora Oh⁶, Ji-Hye Oh⁷, Eun Ju Yang⁴, Hyun Namgung⁸, Dakeun Lee^{9,10}, Yue Pu⁴, Li-Jie Chen⁴, Sang Hyun Park^{8,11}✉, Joong Sup Shim¹²✉, Chang Ohk Sung^{13,7}✉ and Ju Hyun Shim^{12,13,14}✉

The retinoblastoma (*RB1*) gene is a critical tumor suppressor that regulates cell cycle progression and genomic stability. Although *RB1* alterations have been reported in hepatocellular carcinoma (HCC), the biological and clinical consequences of biallelic *RB1* inactivation (RB1-Bi) remain poorly defined. We performed a comprehensive allele-specific genomic analysis of HCC patients from the TCGA-LIHC ($n = 355$) and in-house AMC ($n = 206$) cohorts, collectively comprising the AMC-TCGA discovery cohort. In this combined cohort, RB1-Bi was identified in 14.6% of tumors, was enriched in poorly differentiated HCCs and was independently associated with significantly reduced overall survival (adjusted hazard ratio 3.32, 95% CI 1.93–5.72, $p < 0.001$). Additionally, a deep learning-based histopathology model using hematoxylin and eosin-stained slides (i.e., FR-MIL model) accurately predicted RB1-Bi status (F1 score 84.39% [95% CI, ± 0.02]), making it readily identifiable in routine clinical practice. The prevalence and prognostic impact of RB1-Bi, as well as FR-MIL model performance, were consistent across independent validation cohorts, including advanced-stage tumors and external institutions. High-throughput drug screening in isogenic HCC models revealed that RB1-Bi HCC cells were particularly sensitive to inhibitors targeting mitotic regulators (e.g., AURKA, PLK1, KSP) and DNA damage response pathways (e.g., PARP inhibitors). Synthetic lethal interactions between RB1-Bi and these compounds were demonstrated in vitro and in vivo, and combination treatment with mitotic and PARP inhibitors had synergistic effects with acceptable tolerability. We conclude that RB1-Bi represents a clinically actionable biomarker that identifies a high-risk HCC subtype with specific therapeutic vulnerabilities, offering new opportunities for precision medicine.

Signal Transduction and Targeted Therapy (2026)11:336

; <https://doi.org/10.1038/s41392-026-02815-0>

INTRODUCTION

Hepatocellular carcinoma (HCC), the most prevalent form of primary liver cancer, is the fourth leading cause of cancer-related death and ranks sixth in global cancer incidence.¹ Despite recent advances in systemic therapies, therapeutic outcomes remain suboptimal for a substantial proportion of patients.² Although combinations such as atezolizumab plus bevacizumab have improved survival in advanced disease, durable responses are observed only in a subset of patients, and most individuals eventually develop treatment resistance.³ Moreover, only a minority of HCC tumors harbor molecular alterations that are actionable by currently approved or investigational targeted agents, and validated predictive biomarkers to guide personalized treatment are largely lacking. This highlights a critical unmet need

to identify biologically defined subgroups of HCC that may benefit from tailored therapeutic strategies.⁴

Deregulation of the cell cycle through functional inactivation of the retinoblastoma (*RB1*) pathway is a canonical hallmark of cancer.⁵ *RB1* plays a central role in controlling the G1/S cell cycle checkpoint through regulation of E2F transcription factors, thereby regulating cell cycle progression.⁶ In HCC, genomic studies have identified nonsilent *RB1* mutations in approximately 3–10% of cases and copy number deletions in 10–19%, with frequencies varying across cohorts and sequencing platforms.^{7–10} A meta-analysis of multiple HCC genomic datasets has further identified *RB1* as a putative driver gene, particularly enriched in younger patients.¹¹ Functionally, *RB1* loss is associated with uncontrolled cell cycle progression, aggressive tumor behavior,

¹Gastroenterology and Hepatology, Hanyang University College of Medicine, Guri, Republic of Korea; ²Department of Biomedical Sciences, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea; ³Department of Pathology, Brain Korea 21 Project, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea; ⁴Cancer Centre, Faculty of Health Sciences, MOE Frontiers Science Centre for Precision Oncology, University of Macau, Taipa, Macau SAR, China; ⁵Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA; ⁶Asan Institute for Life Science, Asan Medical Center, Seoul, Republic of Korea; ⁷Bioinformatics Core Laboratory, Convergence Medicine Research Center, Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea; ⁸Graduate School of Artificial Intelligence, Pohang University of Science and Technology (POSTECH), Pohang, Republic of Korea; ⁹Department of Pathology, Ajou University School of Medicine, Suwon, Republic of Korea; ¹⁰Department of Biomedical Sciences, Ajou University, Suwon, Republic of Korea; ¹¹Department of Computer Science and Engineering, Pohang University of Science and Technology, (POSTECH), Pohang, Republic of Korea; ¹²Asan Liver Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea; ¹³Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea and ¹⁴Digestive Diseases Research Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

Correspondence: Sang Hyun Park (sanghyunpark@postech.ac.kr) or Joong Sup Shim (jsshim@um.edu.mo) or Chang Ohk Sung (co.sung@amc.seoul.kr) or Ju Hyun Shim (s5854@amc.seoul.kr)

These authors contributed equally: Jihyun An, Wonkyung Kim, Xiumei Zhang, Philip Chikontwe, Jin-Sung Ju

Received: 23 August 2025 Revised: 17 April 2026 Accepted: 14 May 2026

Published online: 20 August 2026

and poor clinical outcomes across several malignancies, including HCC.^{7,12–14} In addition, *RB1* deficiency has been implicated in resistance to common targeted therapies used against solid malignancies, underscoring its biological and clinical relevance.^{15,16}

According to the classical “two-hit” tumor suppressor model, biallelic inactivation of *RB1* is required for complete loss of function and tumorigenesis.¹⁷ However, most prior genomic studies in HCC have relied on non-allele-specific analyses, without distinguishing between monoallelic and biallelic alterations.^{8,18} As a result, the true prevalence of biallelic *RB1* inactivation, as well as its distinct biological consequences and therapeutic vulnerabilities, remains poorly defined in HCC. From a precision oncology perspective, failure to resolve allelic status may obscure clinically actionable subgroups and limit the identification of biomarker-driven therapeutic strategies.^{19,20} Therefore, a systematic allele-specific characterization of *RB1* alterations, coupled with integrative functional analyses, is essential to delineate the clinical significance of biallelic *RB1* loss and to uncover potential opportunities for targeted intervention.

In this study, we performed a comprehensive allele-specific analysis to define the genomic landscape and molecular consequences of biallelic *RB1* inactivation in HCC and to evaluate its clinical implications. We further developed and validated an artificial intelligence–based approach using digital pathology to reliably identify tumors with biallelic *RB1* inactivation in routine clinical settings, thereby overcoming practical limitations of genomic testing. Finally, through systematic high-throughput drug screening, we identified candidate compounds that selectively induce lethality in *RB1*-biallelic tumors, providing a potential therapeutic strategy for this aggressive molecular subtype. Taken together, our findings establish biallelic *RB1* loss as a biologically and clinically distinct entity and support the development of biomarker-driven precision therapy in HCC.

RESULTS

Overview of discovery and validation cohorts

For the discovery cohort (Fig. 1a, left panel), we integrated two comprehensive genomic datasets: 206 in-house HCC patients from Asan Medical Center (AMC discovery cohort) and 355 HCC cases from The Cancer Genome Atlas (TCGA-LIHC), collectively termed the AMC-TCGA discovery cohort.^{7,8,21} Overall, 70.4% of patients were male ($n = 395$), and the median age was 58 years. Hepatitis B virus (HBV) infection was the most common underlying etiology of chronic liver disease, accounting for 45.0% of cases (Supplementary Table 1). All cases underwent comprehensive multiomics profiling, including whole-exome sequencing (WES), RNA sequencing, and microarray-based single-nucleotide polymorphism (SNP) array analysis (Supplementary Fig. 1a). To minimize technical bias and batch effects, harmonized bioinformatic pipelines were applied across cohorts, including the GATK pipeline for WES and the TCGA RSEM pipeline for transcriptome quantification (Supplementary Fig. 1b). Key findings identified in the AMC-TCGA discovery cohort were subsequently validated across multiple independent in-house cohorts and publicly available datasets spanning diverse molecular platforms, ranging from bulk genomic profiling to spatial and single-cell transcriptomics, as well as heterogeneous clinical contexts, including varying disease stages, therapeutic exposures, and multicenter patient populations (Fig. 1a, right panel; Supplementary Table 2, Supplementary Table 3, and Supplementary Table 4). This study was conducted in accordance with applicable ethical guidelines for human subject research and was approved by the Institutional Review Board of Asan Medical Center (IRB No. 2021-1393).

Functional categorization of *RB1* alterations

Patients were classified into three groups according to *RB1* copy number and pathogenic mutation status as follows: 1) wild-type

RB1 (RB1-WT), 2) monoallelic inactivation of *RB1* (RB1-Mono), and 3) biallelic inactivation of *RB1* (RB1-Bi). To represent the clonal status of *RB1* variants, variant allele fractions (VAFs) were adjusted by tumor purity.

RB1-Bi was defined as 1) homozygous deletion of *RB1*; 2) biallelic pathogenic mutations of *RB1* (adjusted VAF ≥ 0.75); and 3) single copy loss of *RB1* with a concomitant pathogenic mutation that results in loss of heterozygosity (MutLOH) (Fig. 1b). The RB1-WT group was defined as HCC without any pathogenic alteration of *RB1*. The rest of the HCCs were classified as the RB1-Mono group, which refers to tumors with *RB1* shallow deletions without an accompanying pathogenic mutation and those with heterozygous pathogenic mutations of *RB1* without any loss of the concomitant allele (adjusted VAF < 0.75). To reveal the unique characteristics of RB1-Bi carriers, we divided all patients into two subgroups, an RB1-Bi group (with biallelic inactivation of *RB1*) and an RB1-nonBi group (with monoallelic *RB1* inactivation or wild-type *RB1*).

Figure 1c illustrates representative cases of *RB1* wild-type, pathogenic mutations in both alleles, and MutLOH, with corresponding copy number changes and in silico protein structure predictions. Tumors with homozygous deletions of *RB1* in the RB1-Bi group exhibited significantly reduced *RB1* expression compared to the RB1-nonBi group (Fig. 1d), indicating the functional impact of deep copy number alterations. When we mapped the distribution of pathogenic mutations within the *RB1* gene in the AMC-TCGA discovery cohort, we found that most pathogenic variants in the RB1-Bi group were located in the N-terminal region and the pocket A domain. These mutations are likely to promote uncontrolled cell division by inducing misfolding-associated degradation of the Rb protein or by disrupting transcriptional repression of E2F (Fig. 1e and Supplementary Table 5).^{22,23}

Prevalence and prognostic significance of RB1-Bi in discovery and validation cohorts

Among the 561 HCC patients in the AMC-TCGA discovery cohort, 82 (14.6%) were classified as RB1-Bi and 120 (21.4%) as RB1-Mono (Fig. 1f and Supplementary Fig. 2). No pathogenic germline alterations or gene fusions involving *RB1* were identified. Across the validation cohorts, the frequency of RB1-Bi ranged from 4.2% to 18.2% (Fig. 1f). Notably, the GSE164121 cohort, comprising 55.6% non-early-stage HCC cases, and the AMC immune checkpoint inhibitor (AMC-ICI) cohort showed RB1-Bi frequencies of 13.6% and 18.2%, respectively, which were comparable to the frequency observed in the AMC-TCGA discovery cohort. We further assessed the prevalence of RB1-Bi across 10,116 patients representing 33 cancer types in the TCGA pancancer dataset and found that HCC ranked sixth in RB1-Bi frequency (Supplementary Table 6 and Fig. 1g).

Complete baseline and clinical outcome data (including overall survival [OS] and disease-free survival [DFS]) were available for 324 patients in the AMC-TCGA discovery cohort; missing data were confined to the TCGA-LIHC subset (Supplementary Table 7). In patients with complete data ($n = 324$), the RB1-Bi group exhibited the poorest prognosis in terms of both OS and DFS compared with the RB1-Mono and RB1-WT groups (log-rank $p < 0.001$ for both; Fig. 1h, left panel). In multivariable Cox regression analyses adjusting for demographic and tumor-related variables (Fig. 1h, right panel; Supplementary Table 8 and Supplementary Table 9), RB1-Bi remained an independent predictor of adverse outcomes. The adjusted hazard ratios for RB1-Bi were 3.32 (95% CI, 1.93–5.72; $p < 0.001$) for OS and 3.15 (95% CI, 2.12–4.68; $p < 0.001$) for DFS. When the RB1-Bi group was stratified by alteration subtype, no significant differences in OS or DFS were observed between tumors with homozygous deletions and those with other biallelic alterations (log-rank $p = 0.630$ for OS; $p = 0.920$ for DFS; Supplementary Fig. 3), suggesting that pathogenic mutations

Fig. 1 Characterization of HCC with biallelic inactivation of *RB1* (RB1-Bi). **a** Overview of the study design and workflow integrating the AMC-TCGA discovery and validation cohorts. AMC Asan Medical Center, TCGA-LIHC The Cancer Genome Atlas Liver Hepatocellular Carcinoma, CNV copy number variation, SNP single-nucleotide polymorphism, AI artificial intelligence, MSKCC Memorial Sloan Kettering Cancer Center, ICI immune checkpoint inhibitor. **b** Schematic representation of monoallelic and biallelic alterations in the *RB1* gene. **c** Representative examples of copy number variation and protein structural changes at the *RB1* locus across tumors with different alteration patterns: wild-type (WT; H090284), pathogenic mutations affecting both alleles (H080357), and single-copy loss with a concomitant pathogenic mutation resulting in loss of heterozygosity (MutLOH; TCGA-DD-AADM). **d** A representative example of homozygous *RB1* deletion (TCGA-BC-A10Z) is shown on the left. A comparison of *RB1* gene expression between tumors homozygous for *RB1* deletions ($n = 58$) and the RB1-nonBi group ($n = 479$) is shown on the right (two-sided Wilcoxon rank-sum test). **e** Density (top) and distribution (bottom) of *RB1* pathogenic mutations per 150 base pairs of coding sequence in the RB1-Bi ($n = 24$) and RB1-nonBi ($n = 5$) groups. **f** Prevalence of RB1-Bi in the AMC-TCGA discovery cohort and seven validation cohorts. **g** Proportion of cases according to *RB1* alteration status across 10,116 tumors from 33 cancer types. *Proportions in HCC samples were derived from the AMC-TCGA discovery cohort. **h** Kaplan–Meier curves comparing overall survival (top) and disease-free survival (bottom) according to *RB1* alteration group (log-rank test). Forest plots showing hazard ratios (HR) from univariable (left) and multivariable (right) Cox proportional hazards models. Variables with significance in univariable analysis ($p < 0.05$) were included in the multivariable model. * indicates $p < 0.05$ in the multivariable model. CI confidence interval, BMI body mass index, HBV hepatitis B virus, HCV hepatitis C virus, CTP Child-Turcotte-Pugh, AJCC American Joint Committee on Cancer, E-S Edmondson-Steiner, AFP alpha-fetoprotein

differences were observed between the RB1-Bi and RB1-nonBi groups in the immunotherapy-treated cohorts (AMC-ICI, MSKCC-ICI, and GO30140) or in the sorafenib-treated cohort (MSKCC-sorafenib) (Supplementary Fig. 5b, c). Collectively, these findings indicate that RB1-Bi is consistently associated with unfavorable survival outcomes across diverse clinical contexts, whereas the efficacy of currently available systemic therapies does not appear to be significantly modified by *RB1* alteration status.

Clinical and genomic characteristics of RB1-Bi tumors

We compared the clinical features and somatic mutations of patients with top-ranked cancer driver genes between the RB1-Bi and RB1-nonBi groups in the AMC-TCGA discovery cohort (Fig. 2a). No significant differences were observed in demographic or tumor-related characteristics, except for a higher prevalence of poorly differentiated tumors in the RB1-Bi group (48.8% vs. 34.6%, $p = 0.019$; Fig. 2b). RB1-Bi alterations were positively associated with *TP53* mutations and negatively associated with *CTNNB1* mutations, while the distribution of other mutations was generally comparable between the two groups (Fig. 2c). No specific COSMIC mutational signature (v2) was linked to *RB1* alteration status (Supplementary Fig. 6).²⁵

Functionally, RB1-Bi tumors were enriched for the products of cell cycle-related pathways, including E2F targets, G2M checkpoint, mitotic spindle, and DNA repair gene sets, as defined by the Hallmark database, across diverse tumor stages. (Fig. 2d and Supplementary Fig. 7). Key genes involved in cell cycle progression and proliferation (e.g., *CCND1*, *CCNE1*, *E2F1*, and *MKI67*), as well as replication stress-related genes (e.g., *CDC6*, *CLSPN*, *POLQ*, and *RAD51*), were also significantly upregulated (Fig. 2e).²⁶ Notably, pathways associated with the DNA damage response were generally enriched in the RB1-Bi group (Fig. 2f). In line with this, the total frequency of copy number alterations was significantly elevated in RB1-Bi tumors (Fig. 2g), suggesting that biallelic *RB1* loss impairs DNA repair mechanisms, mitotic defects, genomic instability, and ultimately tumorigenesis.^{27,28} Transcriptomic classification further revealed a significant enrichment of RB1-Bi tumors in Boyault's G2 subclass, which is characterized by high proliferation, chromosomal instability, cell cycle activation, and poor differentiation (Fig. 2h).²⁹

In the context of the tumor immune microenvironment, the proportions of CD8⁺ and CD4⁺ T cells, natural killer (NK) cells, and monocytes were consistently reduced in the RB1-Bi group across multiple computational algorithms in the AMC-TCGA discovery cohort (Fig. 2h and Supplementary Fig. 8a). In parallel, immune-related gene sets—particularly those associated with CD8⁺ T-cell responses—were significantly downregulated (Fig. 2i). Quantification of CD8⁺ T-cell infiltration based on immunohistochemistry (IHC) demonstrated a marked reduction in CD8⁺ T-cell density in RB1-Bi tumors compared with RB1-nonBi tumors, both in the AMC discovery cohort and GSE164121 cohort (Fig. 2j and Supplementary

Fig. 8b–c). Notably, among the three identified immunotypes (immunoactive, exhausted, and immunodesert),³⁰ RB1-Bi HCC was significantly enriched in the immunodesert subtype (Fig. 2k), accompanied by reduced expression of immune checkpoint molecules, including *PD-1* and *PD-L1* (Supplementary Fig. 8d). In addition, interferon signaling and antigen processing and presentation pathways were downregulated in RB1-Bi HCC (Supplementary Fig. 8e). No microsatellite instability (MSI)-high cases were identified, and tumor mutational burden (TMB) did not differ between groups, indicating that the reduced immune infiltration in RB1-Bi tumors is independent of MSI status and TMB (Supplementary Fig. 9).

Spatial transcriptomic (ST) analysis using the Visium platform further demonstrated that RB1-Bi HCC exhibited markedly reduced *RB1* expression within tumor regions (Supplementary Fig. 10a–c). Spatial mapping revealed decreased infiltration of T and NK cells, accompanied by reciprocal enrichment of myeloid populations in RB1-Bi tumors. To corroborate these findings, we analyzed independent single-cell RNA sequencing data (GSE156625) and confirmed a reduced frequency of T and NK cells in HCC tumors with low *RB1* expression (Supplementary Fig. 10d–g).³¹ Collectively, these results suggest that impaired immune-related signaling pathways may contribute to the attenuated immune cell infiltration observed in RB1-Bi HCC.

Clonal status and chromosomal duplications in RB1-Bi HCCs

To evaluate the clonality and temporal dynamics of RB1-Bi events, we assessed the cancer cell fraction (CCF) of tumors harboring RB1-Bi alterations in the AMC-TCGA discovery cohort (Fig. 3a). High CCF values (mean > 0.9) were consistently observed across all three types of RB1-Bi events, supporting RB1-Bi as predominantly early clonal events. To further validate this pattern, clonality analysis using MutationTimeR demonstrated that the highest proportion of early clonal events (56.5%) among genes frequently altered in HCC was found in pathogenic *RB1* mutations (Fig. 3b). This finding supports the role of *RB1* mutations as early drivers in tumor evolution.

We next investigated whether early clonal RB1-Bi events influenced subsequent genomic instability by examining the ploidy of RB1-Bi tumors. These exhibited a median ploidy of 3.1 (interquartile range, 2.0–3.8), suggesting a close relationship between whole-genome doubling (WGD) and RB1-Bi status (Fig. 3c and Supplementary Table 12). WGD events were significantly more frequent in the RB1-Bi group than in the RB1-nonBi group (54.9% vs. 38.6%, $p = 0.007$), irrespective of RB1-Bi subtype (Fig. 3d). Temporal modeling using the PhyloGNDT algorithm revealed that both *RB1* pathogenic mutations and 13q deletions encompassing the *RB1* locus (indicative of homozygous *RB1* loss) occurred prior to WGD events (Fig. 3e). Analysis of a representative RB1-Bi case (H080357), with pathogenic *RB1* mutations in both alleles, confirmed that *RB1* alterations preceded

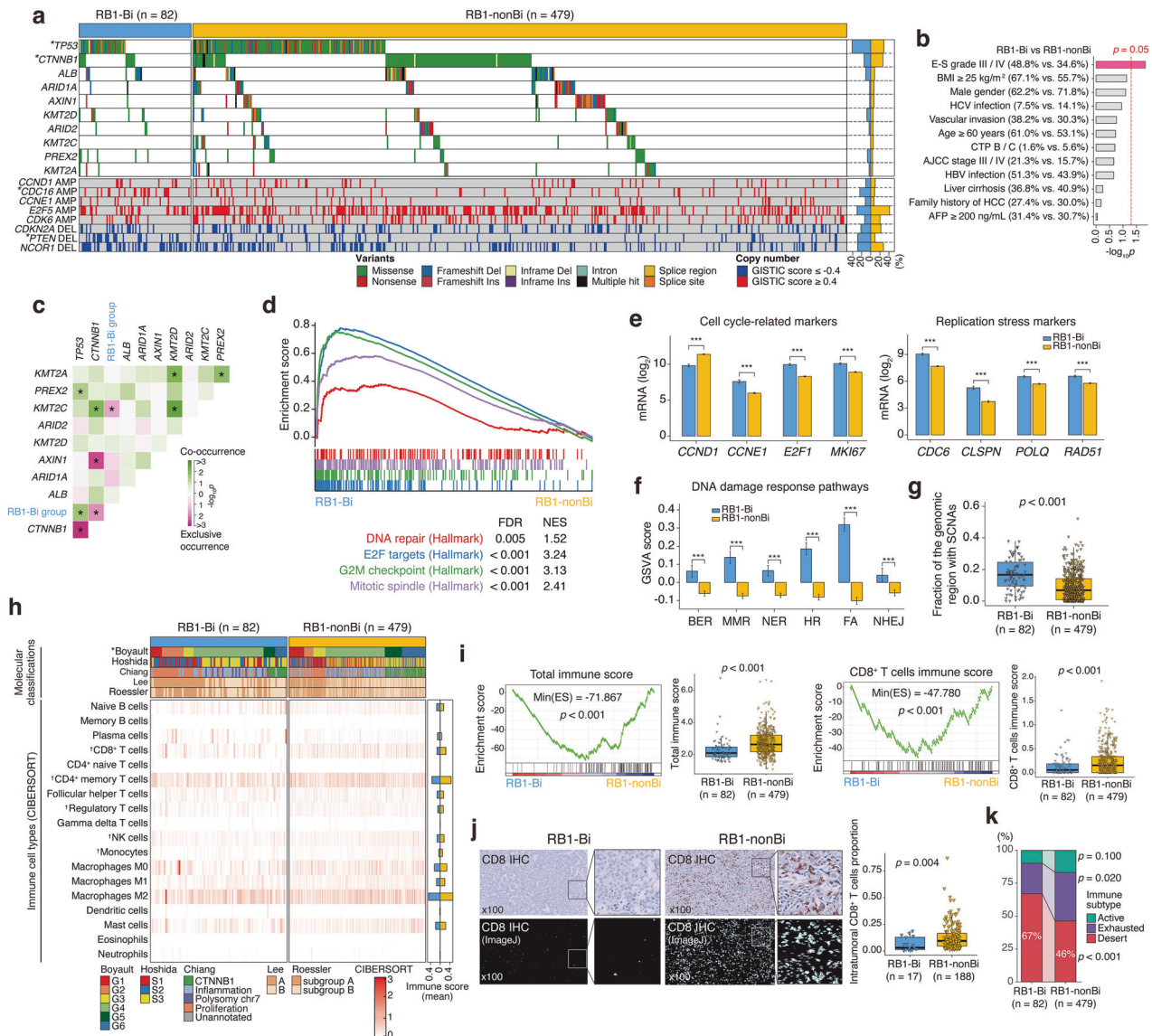


Fig. 2 Integrated analysis and functional implications of RB1-Bi HCCs. **a** Landscape of somatic alterations according to *RB1* alteration status. Genes harboring statistically significant mutation frequencies (MutSigCV2, $q < 0.1$) are displayed across individual tumor samples (columns) as a heatmap. The bottom eight rows display significant DNA copy number alterations in cell cycle-related genes. * indicates oncogenic alterations significantly enriched in RB1-Bi HCCs compared with RB1-nonBi tumors ($p < 0.05$ by Fisher's exact test). **b** Logistic regression analysis for clinical factors significantly associated with the RB1-Bi group. BMI body mass index, HBV hepatitis B virus, HCV hepatitis C virus, CTP Child-Turcotte-Pugh, AJCC American Joint Committee on Cancer, E-S Edmondson-Steiner, AFP alpha-fetoprotein. **c** Mutual exclusivity and co-occurrence analysis of somatic mutations assessed by Fisher's exact test. $p < 0.05$. **d** Significantly upregulated gene sets in the RB1-Bi group by gene set enrichment analysis using the Hallmark database. FDR, false discovery rate; NES, normalized enrichment score. **e** Expression of genes related to the cell cycle (*CCND1*, *CCNE1*, *E2F1*, and *MKI67*) and replication stress (*CDC6*, *CLSPN*, *POLQ*, and *RAD51*) according to *RB1* alteration status. P values were calculated using the Wilcoxon rank-sum test. $***p < 0.001$. **f** Comparison of gene set variation analysis (GSVA) scores in DNA damage response pathways in the RB1-Bi and RB1-nonBi groups. BER, base excision repair; MMR, mismatch repair; NER, nucleotide excision repair; HR, homologous recombination; FA, Fanconi anemia; and NHEJ, nonhomologous end joining. P values were calculated using the Wilcoxon rank-sum test. $***p < 0.001$. **g** Proportions of genomes with somatic copy number alterations (SCNAs) according to *RB1* alteration status. P values were calculated using the Wilcoxon rank-sum test. **h** Distribution of predefined HCC molecular classes and absolute immune cell scores estimated by CIBERSORT according to *RB1* alteration status. Differences in subtype distribution were assessed using Fisher's exact test, and differences in immune cell scores were evaluated using the Wilcoxon rank-sum test. FDR-adjusted q values were computed using the Benjamini-Hochberg method. $*p < 0.05$, $^{\dagger}FDR q < 0.05$. **i** Gene set enrichment plots (left) and box plots (right) of total immune and CD8⁺ T-cell immune scores in the RB1-Bi and RB1-nonBi samples. ES, enrichment score. **j** Representative immunohistochemical (IHC) staining images for CD8 in the AMC discovery cohort according to *RB1* alteration status are shown, with CD8⁺ T cells indicated by brown staining (left). CD8⁺ T cells were quantified using ImageJ-based cell counting, and the normalized ratio (right) represents the proportion of CD8⁺ cells relative to the total number of cells. Differences between groups were assessed using the Wilcoxon rank-sum test. **k** Immune subtype classification based on RNA expression profiles from the AMC-TCGA discovery cohort using the nearest template prediction (NTP) algorithm. The three immune subtypes between the RB1-Bi and RB1-nonBi groups were compared using Fisher's exact test

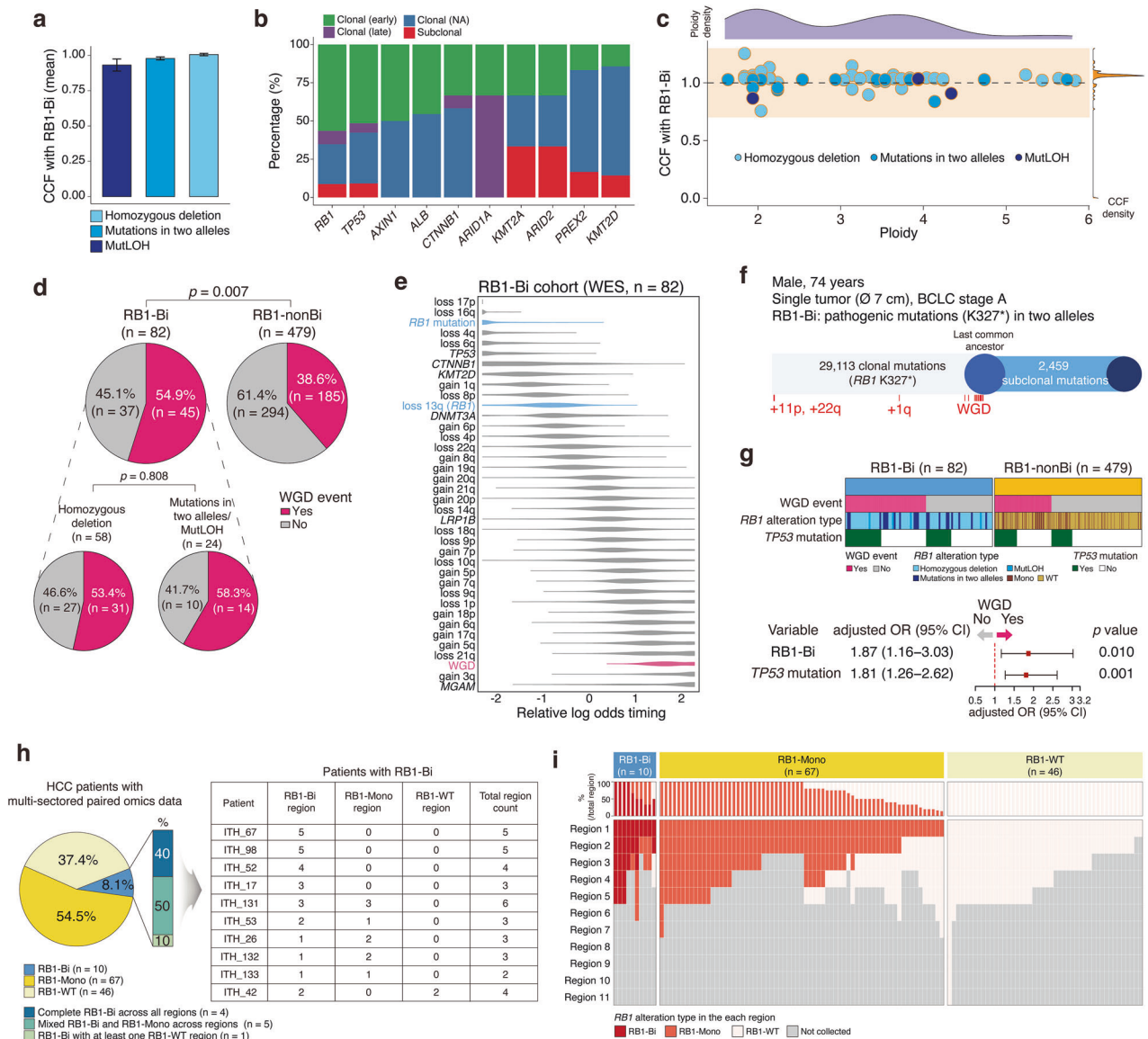


Fig. 3 Genetic clonality and whole-genome doubling in RB1-Bi HCCs. **a** Mean cancer cell fraction (CCF) of the RB1-Bi group ($n = 82$), stratified by RB1-Bi subtype. **b** Clonality assessment of the ten most frequent genomic alterations across the cohort using MutationTimeR, arranged in descending order of early clonal event frequency. NA, timing not assigned. **c** Relationship between ploidy and CCF for RB1-Bi events. Each dot represents an individual tumor, colored by RB1-Bi subtype. Marginal kernel density plots showing the distributions of ploidy and CCF. **d** Frequency of whole-genome doubling (WGD) events among tumors according to RB1 alteration status ($p = 0.007$, Fisher exact test), further stratified by RB1-Bi subtype. **e** Relative log odds of early timing for various genomic alterations in the RB1-Bi HCC cohort ($n = 82$), inferred using PhyloGNTD. *RB1* pathogenic mutations and 13q deletions encompassing *RB1* are shown in blue; WGD events are highlighted in pink. **f** Representative case of a 74-year-old male patient (H080357) with Barcelona Clinic Liver Cancer (BCLC) stage A HCC harboring a clonal biallelic *RB1* pathogenic mutation (K327*) and early chromosomal gains (+11p, +22q), followed by WGD and acquisition of 2459 subclonal mutations. **g** Co-occurrence of WGD and *TP53* mutations in RB1-Bi and RB1-nonBi tumors. The upper panels show individual tumor profiles, while the lower panel shows the results of logistic regression, indicating that RB1-Bi is independently associated with WGD after adjustment for *TP53* mutation status. OR, odds ratio; CI, confidence interval. **h** Analysis of *RB1* alteration status (RB1-Bi, RB1-Mono, and RB1-WT) using multisectored paired omics data from 123 HCC patients from the PLANet cohort. Ten patients (8.1%) exhibited at least one region with RB1-Bi, whereas 67 (54.5%) presented with RB1-Mono only (left). For the 10 patients with RB1-Bi tumors, detailed *RB1* alteration status across all sampled regions is shown (right). **i** Distribution of *RB1* alteration status across multiple regions in 123 HCC patients

WGD (Fig. 3f). Importantly, after adjusting for *TP53* mutations, which are also known to be associated with WGD,³² RB1-Bi status remained independently associated with WGD (odds ratio, 1.87; 95% CI, 1.16–3.03; $p = 0.010$; Fig. 3g).

Analysis of a publicly available multiregion genomic dataset comprising 123 HCC cases identified RB1-Bi in 10 tumors (8.1%; Fig. 3h).³³ Among these, 4 cases (40%) exhibited complete regional concordance of RB1-Bi, while 5 cases (50%) showed coexistence of RB1-Bi and RB1-Mono across spatially distinct

regions (Fig. 3i). No case demonstrated RB1-Bi confined to a single spatial region. These spatial patterns suggest early involvement of the *RB1* pathway with stepwise acquisition of biallelic inactivation during tumor evolution, rather than representing an exclusively late, region-restricted emergence of RB1-Bi.

Deep learning-based classification of RB1-Bi in HCC

Given that RB1-Bi identification requires comprehensive genomic profiling, we developed a histopathology-based deep learning

model to infer RB1-Bi status from routine whole-slide images (WSIs). We trained and evaluated the feature recalibration-based multiple instance learning (FR-MIL) model using the TCGA-LIHC dataset and externally validated it in three independent datasets (Fig. 4a–c, Supplementary Fig. 11, Supplementary Fig. 12, and Supplementary Table 13). To enhance interpretability, patch-level probability scores were overlaid onto WSIs to generate attention heatmaps. RB1-nonBi cases exhibited predominantly low-probability regions, whereas RB1-Bi cases displayed discrete high-probability foci that drove slide-level classification (Fig. 4d).

We first assessed model performance using receiver operating characteristic (ROC) and precision–recall (PR) analyses across independent datasets (Supplementary Fig. 13). The model demonstrated consistent discriminative ability, with area under the ROC curve (AUROC) values of 0.697 in the AMC discovery dataset, 0.800 in the GSE164121 dataset, and 0.661 in the Ajuu dataset. Given the low prevalence of RB1-Bi (5.1–8.3%), PR analysis provided complementary evaluation, yielding mean average area under the precision–recall curve (AUPRC) values ranging from 0.190 to 0.358, with precision consistently exceeding the baseline prevalence. When classification performance was evaluated (Fig. 4e), the FR-MIL model achieved an F1 score of 79.7% with a precision of 83.1% in the TCGA-LIHC internal validation dataset. Comparable performance was observed in two independent AMC cohorts (F1 range: 84.4–91.6% and precision range: 86.7–94.7%; Fig. 4e and Supplementary Fig. 13). The Ajuu dataset further demonstrated stable classification performance (F1: 82.8% and precision: 87.5%), supporting cross-institutional generalizability.

Because RB1-Bi tumors were more frequently poorly differentiated (Fig. 2b), we conducted grade-stratified analyses to determine whether model performance was independent of histologic grade (Supplementary Fig. 14). F1 scores remained stable across Edmondson–Steiner grades (Grade I–II: 83.4%; Grade III–IV: 87.5%; overall: 84.9%), with minimal variation in specificity (85.6–92.3%) and precision (86.4–88.1%). Although discriminative performance showed modest grade-related differences (AUROC: 0.689 vs. 0.564; AUPRC: 0.207 vs. 0.081 for grade I–II vs. III–IV, respectively), this likely reflects increased morphologic heterogeneity in poorly differentiated tumors rather than reliance on grade alone. Collectively, these findings support that the model captures RB1-Bi-associated histopathologic features beyond differentiation-related changes.

Screening for “synthetic lethal” drugs in RB1-deficient HCC models
To identify therapeutic vulnerabilities associated with RB1-Bi in HCC, we conducted a comprehensive synthetic lethal drug screening approach. We established *RB1* knockout (*RB1*^{−/−}) clones from three *RB1*^{+/+} HCC cell lines (Huh7, PLC/PRF/5, and HepG2) to model RB1-Bi and utilized Huh7 *RB1*-isogenic cell pairs for primary synthetic lethal drug screening (Fig. 5a and Supplementary Fig. 15). Additional *RB1*-isogenic pairs were employed for validation studies.

The drug screen encompassed three target-defined small molecule libraries: an epigenetics compound library (128 compounds), a kinase inhibitor library (430 compounds), and a highly selective inhibitor library (318 compounds), collectively targeting diverse cellular druggable proteins (Fig. 5b). Multilibrary hits included aurora kinase A (*AURKA*) and poly(ADP-ribose) polymerase (PARP) inhibitors (Fig. 5c–e). Although mTOR and BTK inhibitors emerged from multiple screens, they lacked clear synthetic lethal effects in validation experiments (data not shown). Polo-like kinase 1 (*PLK1*) and kinesin spindle protein (KSP) inhibitors had reproducible synthetic lethal activities as single-library hits. Based on these results, we selected four representative inhibitors—targeting *AURKA*, *KSP*, *PLK1*, and *PARP*—for subsequent functional validation (Fig. 5f and Supplementary Table 14). Consistent with the identified therapeutic vulnerabilities, RB1-Bi tumors in the AMC–TCGA discovery cohort exhibited significantly elevated expression of the corresponding target genes (*AURKA*,

PLK1, *KIF11*, and *PARP1*; all $p < 0.01$) (Supplementary Fig. 16). Similar trends were observed in the GSE164121 validation cohort, which included a higher proportion of non-early-stage HCC, as well as in our in-house ST data. Notably, enrichment of these target genes was largely confined to RB1-Bi tumors, whereas RB1-Mono cases exhibited intermediate or WT-like expression patterns (Supplementary Fig. 16b).

To validate the screening hits, we selected two independent *RB1*^{−/−} clones from each HCC cell line (Huh7, PLC/PRF/5, and HepG2) and tested their sensitivity to the four target inhibitors: alisertib (*AURKA* inhibitor), ispinesib (KSP inhibitor), volasertib (*PLK1* inhibitor), and talazoparib (PARP inhibitor). All four compounds had consistent synthetic lethal effects, showing enhanced cytotoxicity in *RB1*^{−/−} cells compared to their wild-type counterparts (Fig. 5g–j). Figure 5k shows representative images illustrating the synthetic lethal effects of the four drugs in *RB1*^{−/−} cells. To confirm on-target dependency, we performed siRNA-mediated depletion of *KIF11*, *AURKA*, and *PLK1* in *RB1*-isogenic cell pairs. Knockdown of these genes selectively reduced viability in *RB1*^{−/−} cells but not in *RB1*^{+/+} cells (Supplementary Fig. 17). In contrast, *PARP1* knockdown did not recapitulate the effect of PARP inhibition. This discrepancy is consistent with prior reports demonstrating that PARP inhibitors exert cytotoxicity not only through catalytic inhibition but also via PARP trapping, a mechanism not reproduced by siRNA-mediated depletion.³⁴ Finally, to determine whether drug sensitivity was directly attributable to *RB1* loss, we performed rescue experiments by reintroducing wild-type *RB1* into *RB1*^{−/−} cells. Stable expression of *RB1* restored resistance to all four agents, reversing the synthetic lethal phenotype observed in *RB1*-deficient cells (Fig. 5l and Supplementary Fig. 18).

Mitotic dysregulation and impaired canonical DNA repair in *RB1* deficiency

AURKA, *KIF11*, and *PLK1* are key regulators of G2/M transition and mitotic progression, whereas *PARP1* plays a central role in DNA damage repair (Fig. 5f). Based on these observations, we next investigated the mechanistic basis underlying synthetic lethality in *RB1*-deficient HCC cells.

To examine mitotic progression, we utilized *RB1*-isogenic Huh7 cell pairs. Following synchronization and release from S phase, *RB1*^{+/+} cells completed mitosis and returned to G1 within 12 h, whereas *RB1*^{−/−} cells exhibited an approximately 2-hour delay in mitotic progression (Fig. 6a). Phospho-histone H3 (p-HH3) staining further demonstrated an increased mitotic index in *RB1*^{−/−} cells, consistent with prolonged mitotic duration (Fig. 6b). Transcriptomic profiling and immunoblot analyses revealed coordinated dysregulation of mitotic machinery components—including mitotic kinases, spindle-associated proteins, and checkpoint regulators—in *RB1*^{−/−} cells (Fig. 6c, d). Consistently, upstream regulator and canonical pathway analyses of the AMC–TCGA discovery cohort demonstrated enrichment of E2F target genes, G2/M checkpoint signatures, and mitotic machinery programs in *RB1*-Bi tumors (Fig. 6e, f), with similar findings in ST analyses (Fig. 6g). Collectively, these results indicate that *RB1* loss promotes E2F-driven transcriptional deregulation and mitotic instability, potentially increasing replication stress and genomic vulnerability.

In addition to mitotic dysregulation, we next evaluated canonical DNA repair capacity in *RB1*-deficient cells. Using an I-SceI-based reporter system, we quantified nonhomologous end joining (NHEJ) and homologous recombination (HR) following double-strand break (DSB) induction (Fig. 6h). *RB1*-deficient HCC cells exhibited significantly reduced NHEJ and HR activities compared with their *RB1*-proficient counterparts, indicating impaired canonical DSB repair (Fig. 6i). Consistent with these functional findings, transcriptomic and ST analyses of human tumor samples demonstrated significant alterations in DNA damage response pathways in *RB1*-Bi tumors (Fig. 6e–g). Importantly, *RB1*-Bi tumors lacked pathogenic *BRCA1/2* variants and exhibited comparable *BRCA1/2* mRNA expression levels relative to *RB1*-nonBi tumors (Supplementary Fig. 19), indicating that

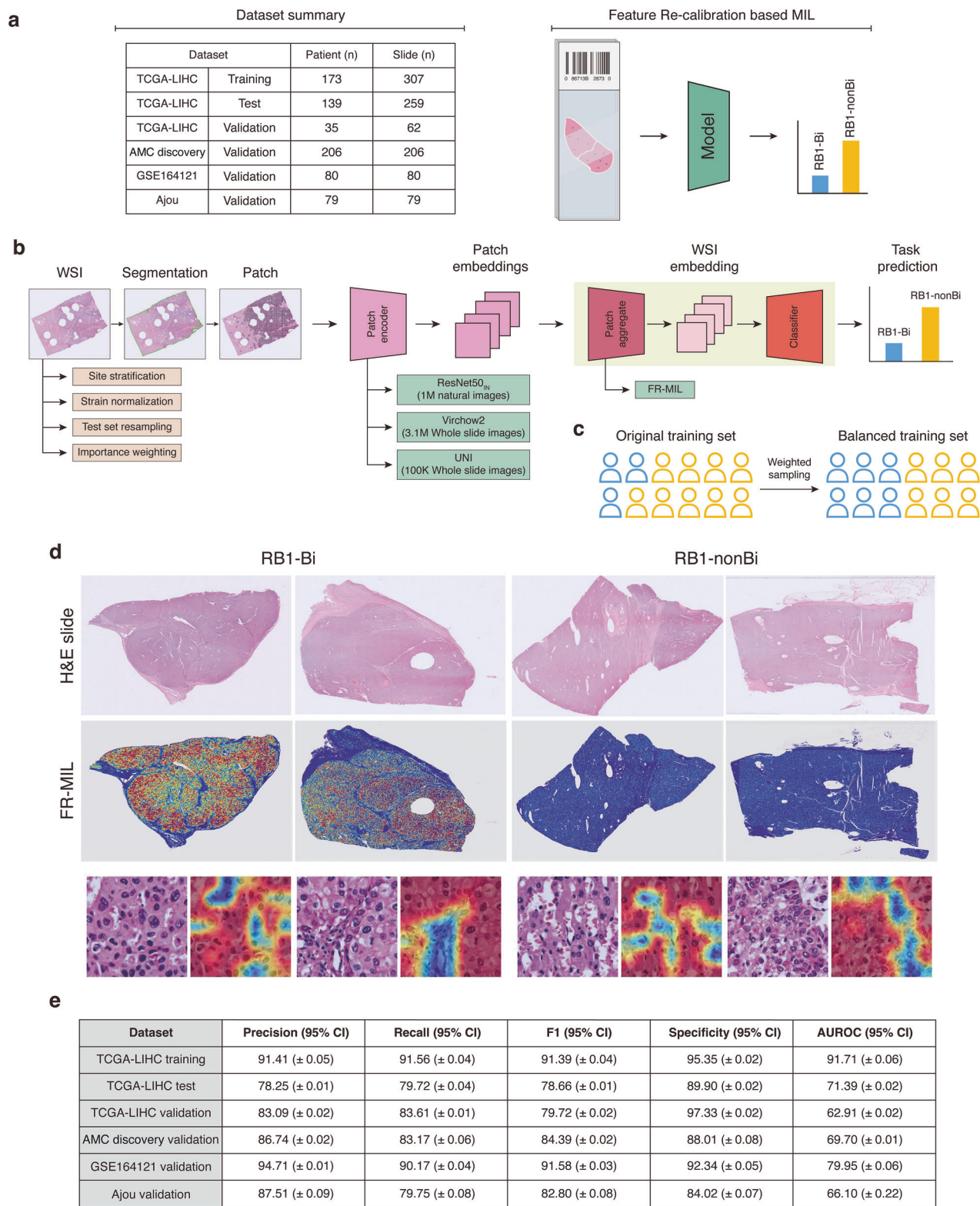
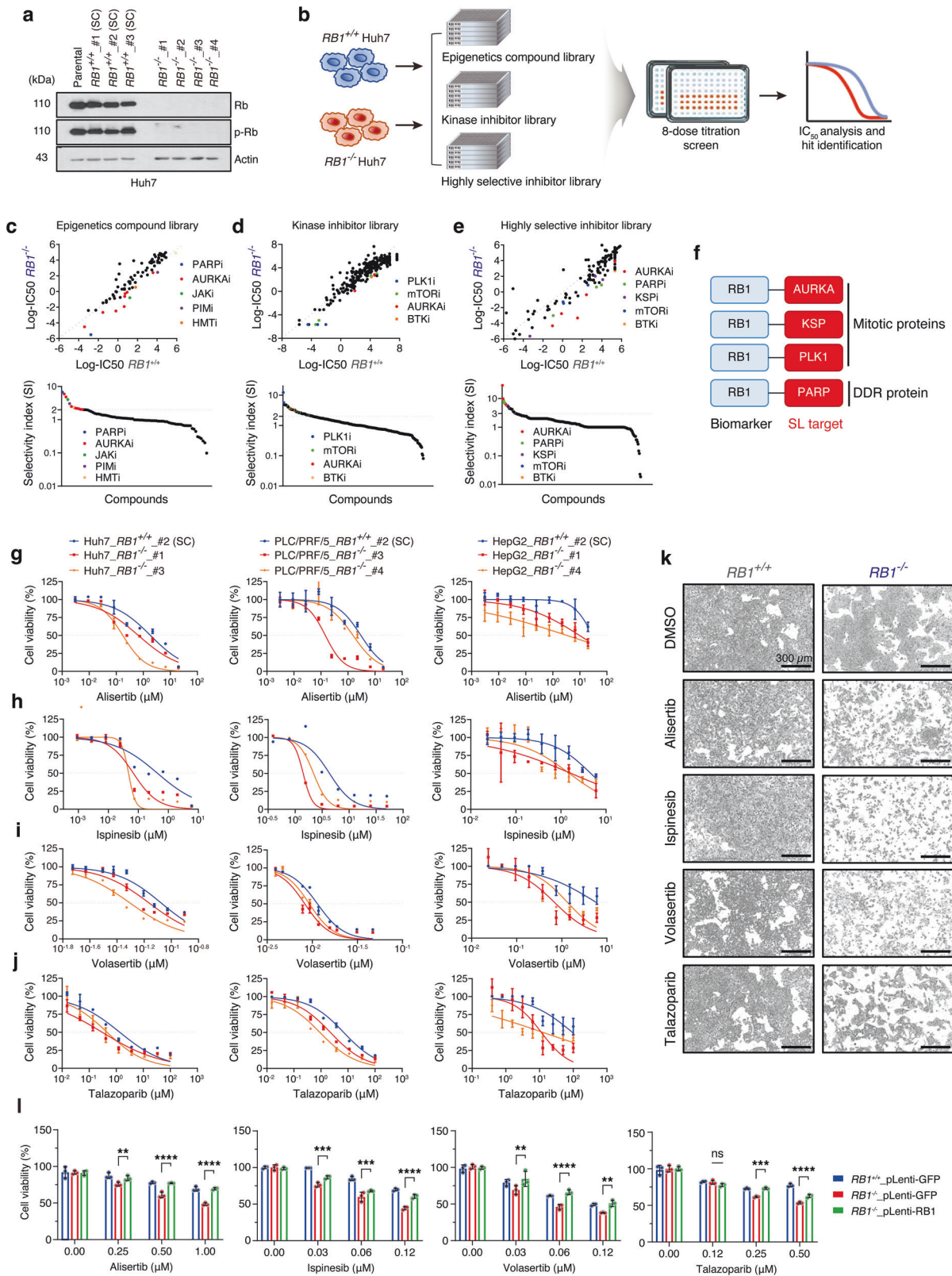


Fig. 4 Deep learning-based prediction of RB1-Bi HCC from whole slide images. **a** Overview of the study design. A total of 628 TCGA-LIHC whole slide images (WSIs) were split into training, test, and validation sets. A feature recalibration-based multiple instance learning (FR-MIL) model was trained and then externally validated using independent internal and external institutional HCC datasets. **b** FR-MIL workflow. Preprocessed WSIs were segmented into patches, from which features were extracted using pretrained models (ResNet50N, Virchow2, UNI), followed by patch aggregation and slide-level representation learning. The final classifier predicted RB1-Bi status. **c** Preprocessing strategies used to reduce batch effects and address dataset imbalance. **d** Selected examples of WSIs from RB1-nonBi and RB1-Bi tumors at three resolutions. Original H&E slides (top), FR-MIL heatmaps (middle), and high-resolution patch-level attention maps (bottom) display the key histological features used by the model for classification. Red areas indicate higher attention scores that are more critical for decisions on WSI classification. **e** Performance metrics of the FR-MIL model evaluated on the independent validation datasets



the observed DSB repair defects are independent of BRCA1/2 dysfunction. Taken together, these data show that RB1 deficiency is characterized by mitotic dysregulation and impaired canonical DNA repair, which may underlie the observed sensitivity to mitotic kinase inhibitors and PARP-targeted therapies.

Pharmacologic targeting of mitotic and DNA repair vulnerabilities in RB1-deficient HCC
Building upon the identified mitotic dysregulation and DNA repair defects in RB1-deficient cells, we next evaluated the phenotypic consequences of pharmacologic inhibition targeting these pathways.

Fig. 5 Screening of drugs for synthetic lethality (SL) to identify therapeutic vulnerabilities associated with RB1-Bi HCC. **a** Western blot validation of RB1 knockout ($RB1^{-/-}$) clones generated in Huh7 cells. **b** Schematic illustration of the SL screening strategy using three drug libraries: the Epigenetics Drug Library (128 drugs), Kinase Inhibitor Library (430 drugs), and Highly Selective Inhibitor Library (318 drugs). **c–e** Scatter plots showing the IC₅₀ values (top) and selectivity indices (SI, bottom) for SL candidates from the epigenetics drug library (**c**), kinase inhibitor library (**d**), and highly selective inhibitor library (**e**). Drugs with SI > 2 are highlighted by colored dots. **f** Summary of the potential SL targets in RB1-Bi HCC: AURKA, KSP, PLK1, and PARP. **g–j** Validation of the top candidate compounds—5-day exposure to alisertib (**g**), ispinesib (**h**), volasertib (**i**), and talazoparib (**j**)—in $RB1^{+/+}$ and $RB1^{-/-}$ isogenic cell lines across three different HCC models (Huh7, PLC/PRF/5, and HepG2). Dose–response curves represent the mean $\pm$ SD of cell viability. **k** Representative images showing SL effects in HepG2 $RB1^{+/+}$ and $RB1^{-/-}$ cells treated with DMSO or the indicated compounds, demonstrating increased sensitivity and cytotoxicity in RB1-deficient cells. Images were acquired using an IncuCyte ZOOM system. Scale bars: 300 μ m. **l** Effects of RB1 reintroduction in RB1-deficient HCC cells on the synthetic lethal effects of the four drugs. Stable cell lines expressing empty vector with CopGFP (pLenti-GFP) and RB1 with CopGFP (pLenti-RB1) in either $RB1^{+/+}$ or $RB1^{-/-}$ Huh7 cells were tested with different concentrations of four SL drugs for cell viability assays. Data represent the mean $\pm$ SD, and statistical significance was analyzed by Student's *t* test. *****p* < 0.0001; ****p* < 0.001; ***p* < 0.01. ns denotes not significant (*p* > 0.05)

All four synthetic lethal agents significantly increased the mitotic index, particularly in $RB1^{-/-}$ cells, inducing G2/M arrest (Fig. 7a–d and Supplementary Fig. 20). Spindle morphology analysis revealed distinct mechanistic patterns: alisertib (AURKA inhibitor) induced shortened bipolar spindles with reduced astral microtubules, whereas ispinesib (KSP inhibitor) and volasertib (PLK1 inhibitor) caused monopolar spindle formation with chromosome misalignment (Fig. 7e, f). In contrast, talazoparib (PARP inhibitor) did not alter spindle architecture, consistent with its primary role in DNA repair inhibition. Consistent with the DNA repair vulnerability observed in RB1-deficient cells, PARP inhibitor and PLK1 inhibitor significantly increased γ -H2AX levels and nuclear foci formation, indicating the accumulation of DNA damage (Fig. 7g, h). The KSP inhibitor also elevated γ -H2AX levels, likely reflecting secondary DNA damage resulting from mitotic failure rather than direct inhibition of DNA repair pathways. Comet assays further demonstrated that the PLK1 inhibitor and PARP inhibitor markedly increased DNA strand breaks in $RB1^{-/-}$ cells (Fig. 7i–k). These results delineate two mechanistically distinct but complementary RB1 synthetic lethal vulnerabilities: mitotic dysregulation targeted by AURKA, KSP, and PLK1 inhibition and DNA repair dependency targeted by PLK1 and PARP inhibition, with PLK1 functioning at the intersection of both processes.

Drug combination studies further supported this framework. Combinations targeting the same pathway showed minimal synergy, whereas cross-pathway combinations (AURKA inhibitor plus PARP inhibitor; KSP inhibitor plus PARP inhibitor) exhibited strong synergistic effects. Both fixed-ratio Chou–Talalay analysis and nonfixed-ratio Highest Single Agent (HSA) modeling confirmed robust synergy between cross-pathway combinations (Fig. 7l–n), indicating that simultaneous targeting of mitotic progression and DNA repair represents a rational therapeutic strategy in RB1-Bi HCC.

To experimentally evaluate whether RB1 alteration status influences responsiveness to currently available first-line therapies, we performed drug sensitivity assays using isogenic $RB1^{+/+}$ and $RB1^{-/-}$ HCC cell models (Supplementary Fig. 21a). Dose–response analyses across three HCC cell lines (Huh7, PLC/PRF/5, and HepG2) demonstrated no consistent differences in lenvatinib sensitivity between $RB1^{+/+}$ and $RB1^{-/-}$ cells, with comparable IC₅₀ values observed across genotypes. To assess responsiveness to immune checkpoint blockades, we conducted peripheral blood mononuclear cell (PBMC)-mediated immune cytotoxicity assays (Supplementary Fig. 21b–d). Tumor cells were cocultured with activated PBMCs in the presence or absence of atezolizumab, durvalumab, or pembrolizumab. ICIs did not significantly enhance PBMC-mediated cytotoxicity in either $RB1^{+/+}$ or $RB1^{-/-}$ cells.

In vivo validation of RB1 synthetic lethal targets

To evaluate synthetic lethal effects in vivo, we established subcutaneous xenograft models using RB1-isogenic PLC/PRF/5 tumors in immunocompromised mice. Based on preliminary toxicity studies defining maximum tolerated doses and administration schedules (see “Methods”), we assessed the therapeutic efficacy of four candidate synthetic lethal compounds (Fig. 8a). All

four agents selectively demonstrated significant antitumor activity in $RB1^{-/-}$ xenografts, with no observable systemic toxicity, as indicated by stable body weight and preserved general condition (Fig. 8b–m and Supplementary Fig. 22). Immunohistochemical analyses revealed increased p-HH3 levels in treated $RB1^{-/-}$ tumors, consistent with enhanced mitotic arrest. In addition, the KSP inhibitor, PLK1 inhibitor, and PARP inhibitor—but not the AURKA inhibitor—elevated γ -H2AX levels, indicating the accumulation of DNA damage. Gene set enrichment analysis (GSEA) further supported these mechanistic findings, demonstrating enrichment of G2M checkpoint and mitotic arrest signatures in drug-treated RB1-deficient tumors.

We next evaluated the in vivo efficacy and tolerability of cross-pathway drug combinations (AURKA inhibitor plus PARP inhibitor, KSP inhibitor plus PARP inhibitor) in $RB1^{-/-}$ xenografts. For combination studies, each drug was administered at half the dose used in single-agent experiments (Fig. 8n). Both combination regimens produced significantly greater tumor growth inhibition than their respective single-agent controls (Fig. 8o, p). Combination treatment further enhanced pharmacodynamic markers, including increased p-HH3 and γ -H2AX levels, consistent with augmented mitotic disruption and DNA damage, respectively (Fig. 8o, p and Supplementary Fig. 23a, b). Importantly, no overt toxicity was observed. Body weight remained stable throughout treatment, and serum alanine aminotransferase (ALT) levels at the study endpoint were comparable to those in the single-agent groups and remained within the normal range across all groups (Fig. 8q and Supplementary Fig. 23c, d). These findings indicate that cross-pathway synthetic-lethal combinations achieve superior antitumor efficacy while maintaining acceptable in vivo tolerability.

DISCUSSION

Our comprehensive analysis demonstrates that biallelic RB1 alterations drive characteristically aggressive HCC behaviors with poor clinical outcomes across tumor stages, involving widespread chromosomal copy number alterations, poorly differentiated pathology, and defective DNA damage repair and replication stress responses. Through artificial intelligence (AI)-assisted histopathological analysis and systematic drug screening, we succeeded in identifying both diagnostic markers and synthetic lethal therapeutic targets specific to RB1-deficient HCC, thus offering translational approaches for this high-risk patient population.

The biological and prognostic significance of RB1 oncogenic mutations has been studied in several cancer lineages, including HCC, but the particular impact of RB1-Bi, as opposed to RB1-Mono, had not previously been established.^{7,13} Although haploinsufficiency of tumor suppressor genes may attenuate their physiological function of preventing abnormal cell proliferation, complete inactivation of both alleles is classically needed for the development of many types of cancer,³⁵ as reflected in the tumor-prone phenotype resulting from RB1 loss in the case of HCC. Our findings

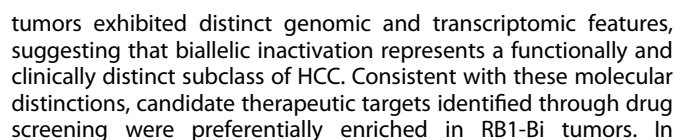


Fig. 6 Mitotic dysregulation and impaired DNA damage repair pathways in RB1-Bi HCC. **a** Flow cytometry analysis of the cell cycle profiles of $RB1^{+/+}$ and $RB1^{-/-}$ Huh7 cells synchronized by double thymidine block and release. Asynchronous cells are shown as a control. PE-A, phycoerythrin area. **b** Immunofluorescence analysis of mitotic indices of $RB1^{+/+}$ and $RB1^{-/-}$ Huh7 cells using a phosphorylated histone-H3 antibody. Representative images (DAPI: blue; phosphorylated histone H3: green) (left) and mitotic index quantification (right) are shown. Scale bars: 45 μ m. $n = 3$. Data are the mean $\pm$ SD, and statistical significance was analyzed by Student's t test. **** $p < 0.0001$. **c** Heatmap showing differential expression of mitosis-related and DNA repair pathway genes between $RB1^{+/+}$ and $RB1^{-/-}$ Huh7 cells. **d** Immunoblot of mitotic regulators and microtubule-associated proteins in $RB1^{+/+}$ and $RB1^{-/-}$ cells. $RB1$ loss is associated with increased levels of mitotic regulators (e.g., KIF5B, Cyclin B1, and Aurora B) and altered tubulin acetylation. **e** Network representation of downstream genes regulated by $RB1$, inferred by Ingenuity Pathway Analysis (IPA) upstream regulator analysis based on differentially expressed genes identified using the Wilcoxon rank-sum test. Nodes colored in orange and blue indicate predicted activation and inhibition, respectively. The edge colors and line styles correspond to the predicted regulatory relationships, as detailed in the prediction legend. **f** Canonical pathway enrichment analysis of genes linked to RB1-Bi tumors using IPA. Four significantly enriched canonical pathways are presented, with corresponding pathway activation Z scores and p values. **g** Spatial expression maps showing higher expression of mitotic regulators in an RB1-Bi tumor than in an RB1-nonBi tumor. Comparison of gene set variation analysis (GSVA) scores between RB1-Bi and RB1-nonBi tumors. The pathways were significantly enriched in RB1-Bi tumors based on fast gene set enrichment analysis (FGSEA). WP, WikiPathways; NES, Normalized Enrichment Score. **h** Schematic of the DNA double-strand break (DSB) repair reporter (DRR) assay. A DRR-GFP construct containing I-SceI recognition sites between NeomycinR-T2A and GFP sequences was used to induce site-specific DSBs via I-SceI endonuclease. Repair outcomes are monitored using donor DNA containing mCherry and BFP. Nonhomologous end joining (NHEJ) restores GFP expression, while homologous recombination (HR) incorporates mCherry, enabling fluorescent discrimination of repair pathways. BFP is constitutively expressed upon transfection and is thus used as an internal control for reporter-expressing cells. **i** Flow cytometry analyses of GFP (left)- and mCherry (right)-expressing cells as surrogate markers for NHEJ and HR efficiency, respectively. The right panels in each data point are quantitative measurements of DSB repair efficiency. DNA repair activity in the absence of I-SceI (I-SceI (-)) was used as a negative control

contrast, RB1-Mono cases exhibited intermediate or overlapping features with RB1-wildtype tumors, although the functional and pharmacologic significance of this subgroup remains to be experimentally clarified. Taken together, these findings support the biological and clinical relevance of RB1-Bi as a distinct context for understanding the oncogenic role of $RB1$ alterations and their associated therapeutic vulnerabilities.

To place RB1-Bi within the broader landscape of established HCC molecular subclasses, we examined its relationship with $TP53$ - and $CTNNB1$ -driven tumors. In our cohort, RB1-Bi frequently co-occurred with $TP53$ mutations while demonstrating near mutual exclusivity with $CTNNB1$ alterations. $TP53$ -mutant HCC is typically characterized by chromosomal instability, enhanced proliferative signaling, poor differentiation, and aggressive clinical behavior, whereas $CTNNB1$ -mutant tumors are defined by WNT/ β -catenin pathway activation, relative immune exclusion, and generally well-differentiated histology with lower proliferative indices.^{36,37} Although RB1-Bi partially overlaps with $TP53$ -mutant tumors, our data indicate that it defines a distinct molecular context characterized by pronounced mitotic dysregulation, activation of DNA damage response pathways, and an immune-depleted phenotype. Importantly, RB1-Bi tumors were largely absent in $CTNNB1$ -mutant HCC, suggesting biological divergence from the WNT-driven subclass. Collectively, these findings position RB1-Bi as a proliferation- and genomic instability-associated molecular subgroup that intersects with, but is not redundant to, $TP53$ -mutant HCC and remains clearly distinct from $CTNNB1$ -driven tumors.

The identification of RB1-Bi HCCs traditionally requires testing for widespread mutations and allele-specific copy number alterations across the $RB1$ gene through whole-exome or whole-genome sequencing. However, the high cost and technical complexity of these methods limit their widespread clinical adoption. In our study, we demonstrated that RB1-Bi HCC can be distinguished with high accuracy using AI analysis of pathology images alone, eliminating the need for labor-intensive region-of-interest specification or cell type annotation. Importantly, external validation across independent cohorts with diverse patient populations and institutional workflows demonstrated the robustness and generalizability of the model beyond a single race or slide source. Our FR-MIL model was trained using a weakly supervised approach that learns directly from whole slide images with only slide-level RB1-Bi labels. This study establishes a foundation for future clinical applications where patients with RB1-Bi tumors can be identified directly from routine pathological images. The annotation-free training approach minimizes human

error and reduces preparation time, thereby offering a more efficient and cost-effective solution for clinical implementation.

Despite its consistent association with poor survival across diverse clinical settings, RB1-Bi does not appear to significantly modify the therapeutic efficacy of currently approved systemic treatments, including targeted agents and ICIs, in either clinical or experimental contexts. Through comprehensive drug screening, we identified several mitotic inhibitors, including AURKA, KSP, and PLK1 inhibitors, that exhibited synthetic lethality with $RB1$ loss. This finding is consistent with the transcriptional profiles of patients with RB1-deficient HCC, which demonstrate marked upregulation of genes involved in G2/M transition and mitotic control, highlighting the therapeutic potential of targeting mitotic pathways in this molecular subtype.^{38,39} $RB1$ loss also compromises DNA damage repair processes independently of BRCA dysfunction, thereby sensitizing tumor cells to PARP inhibitor, as demonstrated in our drug screening analyses.^{40,41} Notably, combination drug testing revealed pronounced synergistic effects, with the AURKA inhibitor plus the PARP inhibitor and the KSP inhibitor plus the PARP inhibitor inducing the most potent synthetic lethal responses. These combinatorial effects were further validated in vivo, where dual targeting demonstrated superior antitumor activity compared with single-agent treatment, without compromising hepatic function or inducing overt toxicity. Notably, PARP inhibitor monotherapy has demonstrated only modest clinical benefit in $RB1$ -altered and other non-BRCA-driven malignancies.⁴² Therefore, our findings support positioning PARP inhibition within a rational synthetic-lethal combination framework rather than as a standalone strategy. However, the combinations evaluated in this study were selected based on independently identified vulnerabilities and were not derived from a systematic search for optimal partners for PARP inhibition in the RB1-Bi context. Future studies employing systematic drug-combination screening or CRISPR-based chemical genetic approaches under PARP inhibitor treatment will be important to identify optimal synergistic partners and to further refine biomarker-driven therapeutic strategies for RB1-Bi HCC.

Of the four synthetically lethal candidates identified in RB1-Bi HCC, only PARP inhibitors have received approval from the US FDA and EMA for treating BRCAness cancers. The other three inhibitors—AURKA, KSP, and PLK1 inhibitors—have been actively investigated in phase I to III clinical trials across various malignancies over the past decade.^{43–45} The major clinical challenge with these agents has been their lack of superior efficacy as monotherapies compared with standard comparator

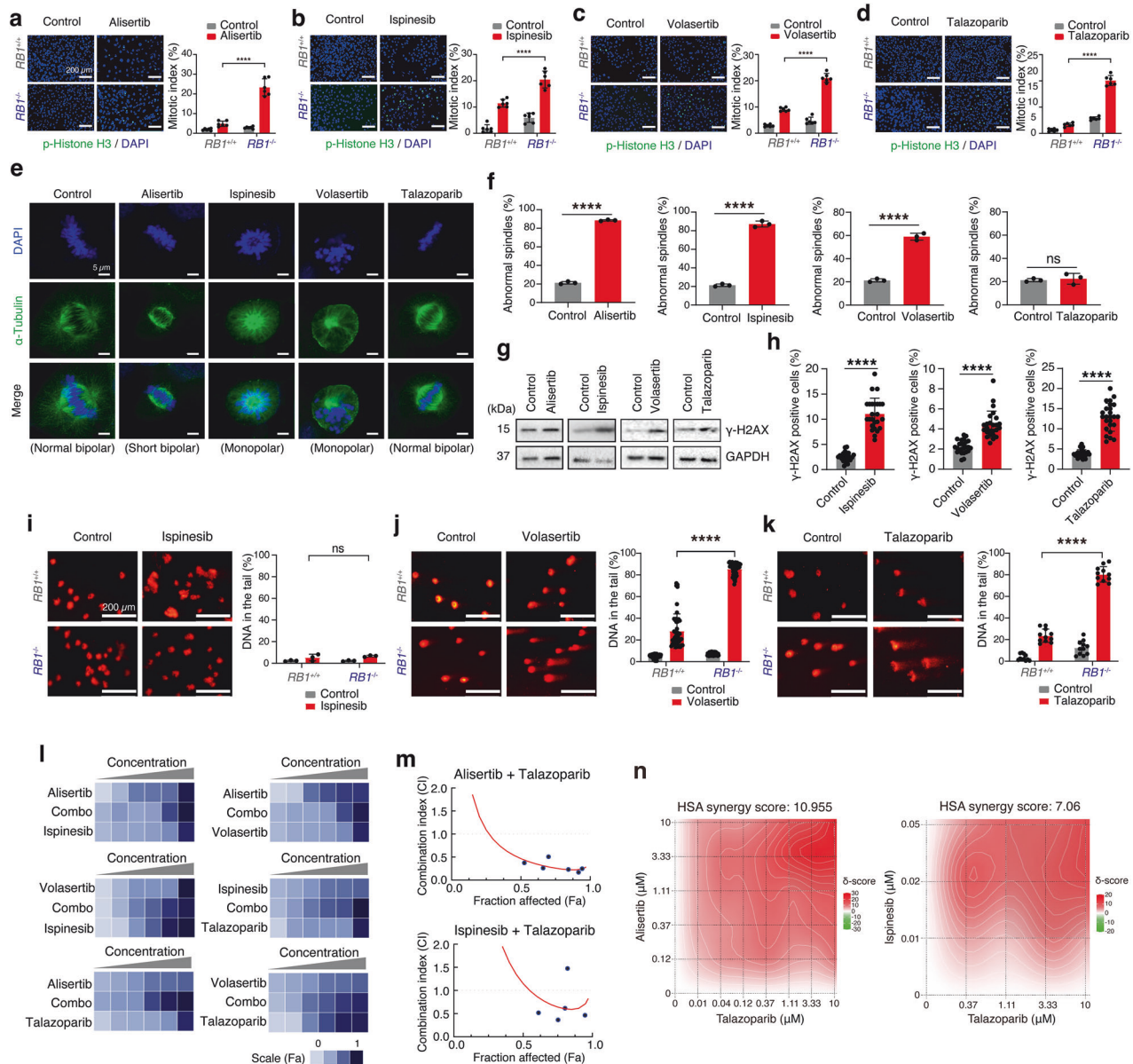
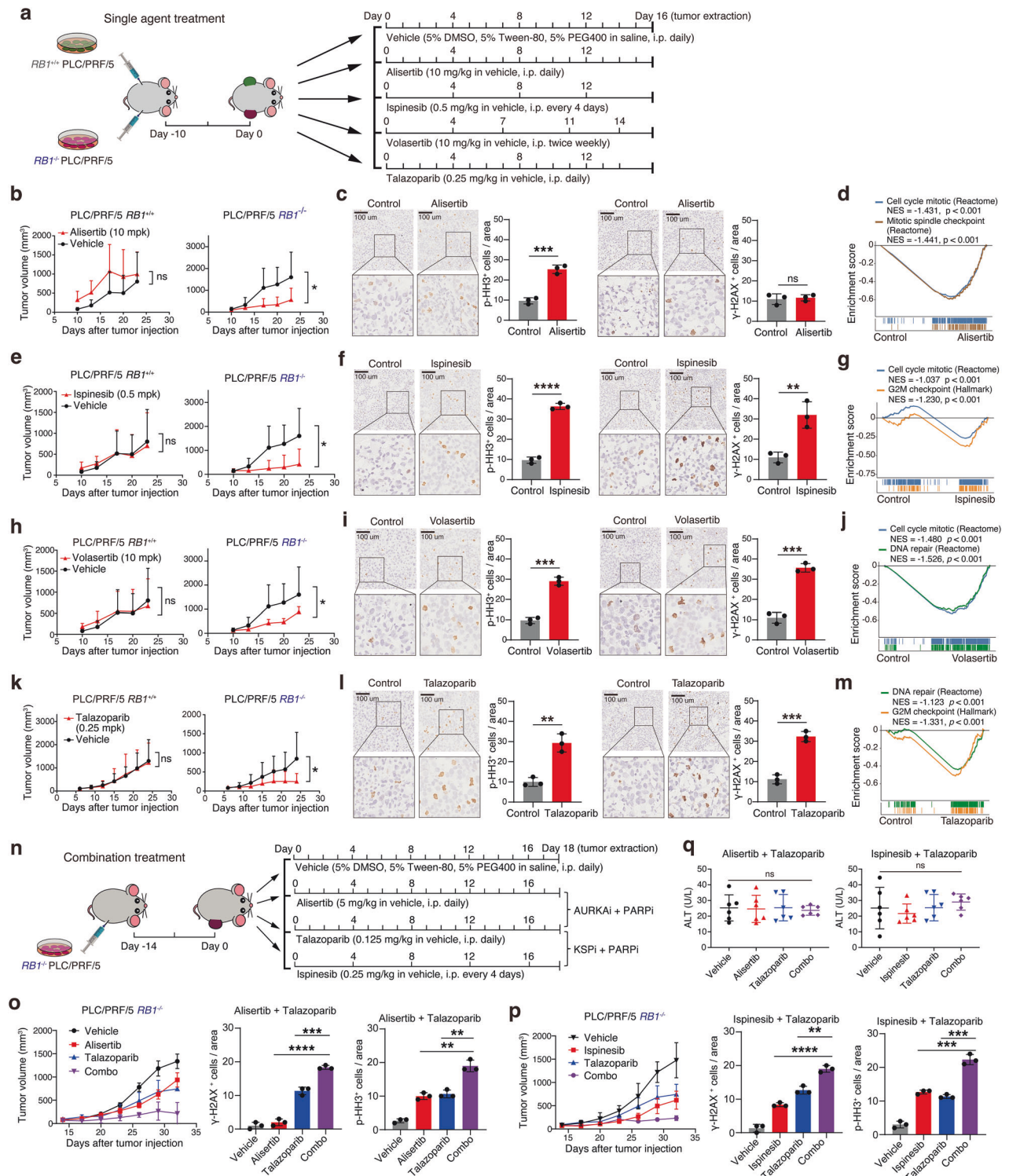


Fig. 7 Single-agent and combinatorial therapeutic vulnerability in RB1-deficient HCC. Immunofluorescence of p-Histone H3 (green) and DAPI (blue) in $RB1^{+/+}$ and $RB1^{-/-}$ Huh7 cells treated with the indicated compounds or DMSO control: alisertib (a), ispinesib (b), volasertib (c), and talazoparib (d). $RB1^{-/-}$ cells show significantly more extensive mitotic arrest in response to 48-h treatment. Quantified mitotic index shown on the right (**** $p < 0.0001$). e Immunofluorescence images of mitotic spindles stained with DAPI (blue) and α -tubulin (green) in Huh7 cells treated with each compound. Scale bars: 5 μ m. A total of 100 mitotic cells per condition were analyzed. Data represent the mean $\pm$ SD, and statistical significance was analyzed by Student's t test. **** $p < 0.0001$. ns denotes not significant ($p > 0.05$). f Quantification of abnormal spindle formation after drug treatment. $RB1^{-/-}$ Huh7 cells exhibited significantly more spindle defects after alisertib, ispinesib, and volasertib (**** $p < 0.0001$) but not talazoparib (ns: not significant). g Western blot analysis of γ -H2AX in $RB1^{-/-}$ Huh7 cells treated with the indicated drugs for 24 h. GAPDH was used as a loading control. h Quantification of γ -H2AX-positive cells by immunofluorescence after 48 h of treatment, showing significantly increased DNA damage following ispinesib, volasertib, and talazoparib exposure. γ -H2AX nuclear foci-positive cells and total cells were counted in each confocal image area and quantified. i–k Representative comet assay images (PI staining, red) and quantification of DNA damage (% DNA in tail) in $RB1^{+/+}$ and $RB1^{-/-}$ Huh7 cells treated with ispinesib (i), volasertib (j), or talazoparib (k) for 48 h. Scale bars: 200 μ m. **** $p < 0.0001$. l Drug combination matrices showing synergistic effects by Fa (fraction affected). Cells were treated with increasing concentrations of alisertib, ispinesib, talazoparib, or volasertib alone or in combination for 4 days. Cell viability was measured using AlamarBlue. Scale indicates Fa values using different color intensities (0: 100% viability; 1: 0% viability). m Chou-Talalay Fa-CI plot for alisertib plus talazoparib and ispinesib plus talazoparib. CI < 1: synergism; CI = 1: additivity; CI > 1: antagonism. Representative data from three independent experiments are shown. n Highest Single Agent (HSA) synergy analysis using non-fixed drug concentration ratios. HSA synergy scores of 5–10 indicate moderate to strong synergy, and scores > 10 indicate strong synergy

drugs, leading many trials to fail in meeting their primary end points required for regulatory approval. One potential reason for these developmental setbacks is that clinical studies have a pancancer or pan-patient design rather than adopting tumor

subtype-specific tailored approaches similar to those used successfully to develop PARP inhibitors. A notable example is the AURKA inhibitor trial that failed to meet its primary endpoint in a phase II ovarian clear cell carcinoma study in 2018.⁴⁶ Post hoc



analysis revealed that patients who responded favorably to AURKA inhibition were predominantly those with *ARID1A*-mutant subtypes, where the AURKA inhibitor was synthetically lethal in combination with *ARID1A* loss.⁴⁷ This finding underscores the critical importance of tumor subtype-specific trial design for developing synthetic lethal drugs. Therefore, biomarker-driven, tumor subtype-specific trials will be essential for developing effective HCC-targeted therapies. The identification of two distinct classes of *RB1* synthetic lethal drugs in our study provides a crucial

foundation for developing precision medicine approaches for HCC patients harboring *RB1* biallelic loss.

Several limitations of this study should be acknowledged. First, our *in vivo* therapeutic experiments were conducted using subcutaneous xenograft models in immunodeficient mice. While these models provide valuable proof-of-concept evidence, they do not fully recapitulate the liver-specific tumor microenvironment or immune interactions that critically influence HCC progression and treatment response. Given that *RB1*-Bi tumors are characterized by

Fig. 8 Monotherapies and combination treatments of synthetic lethal drugs on tumor growth of *RB1*-isogenic HCC xenografts in vivo. **a** Schematic illustration of the *RB1*^{+/+} and *RB1*^{-/-} PLC/PRF/5 HCC xenograft mouse models. Vehicle and specified drugs were administered according to the indicated dosing schedule. **b, e, h, k** Minor ticks denote the days of drug injection. Tumor growth inhibition in PLC/PRF/5 xenograft models treated with each compound: alisertib (**b**), ispinesib (**e**), volasertib (**h**), and talazoparib (**k**). Tumor volumes were measured in *RB1*^{+/+} (left) and *RB1*^{-/-} (right) mice treated with vehicle or the indicated drug. Statistical comparisons were performed using Student's *t* test. **p* < 0.05; **c, f, i, l** Immunohistochemical analysis of γ -H2AX (DNA damage marker) and phospho-histone H3 (p-HH3; mitotic marker) in tumor tissues from *RB1*^{-/-} xenografts. Representative IHC images and quantification of marker-positive cells are shown. Scale bars: 100 μ m. Quantification data for IHC are mean $\pm$ SD from three different tumor sections. Statistical comparisons were performed using Student's *t* test. ***p* < 0.01; ****p* < 0.001; *****p* < 0.0001; ns denotes not significant (*p* > 0.05). **d, g, j, m** Gene set enrichment analysis (GSEA) plots showing significantly upregulated pathways in *RB1*^{-/-} tumors following treatment with the indicated compounds. **n** Schematic illustration of the drug combination study in the *RB1*^{-/-} PLC/PRF/5 HCC xenograft mouse model. Vehicle and specified drugs were administered according to the indicated dosing schedule. Minor ticks denote the days of drug injection. **o, p** Mice bearing *RB1*^{-/-} PLC/PRF/5 HCC xenografts were treated with vehicle, alisertib (5 mg/kg), talazoparib (0.125 mg/kg), ispinesib (0.25 mg/kg), or the indicated two-drug combinations (Combo). Tumor volumes were measured in mice treated with vehicle or the indicated drugs. Because talazoparib is included in both combination regimens, the talazoparib single-agent control group is shared between the two panels. Immunohistochemical (IHC) analysis of γ -H2AX (DNA damage marker) and p-HH3 (mitotic marker) in tumor tissues from *RB1*^{-/-} xenografts. Quantification of marker-positive cells is shown. Quantification data for IHC are mean $\pm$ SD from three different tumor sections. Statistical comparisons were performed using Student's *t* test. ***p* < 0.01; ****p* < 0.001; *****p* < 0.0001. **q** Serum alanine transaminase (ALT) levels were quantified at the study endpoint and reported in units per liter (U/L). A common control group was used for the Alisertib, Ispinesib and Volasertib assays; therefore, the corresponding panels share the same control dataset. Statistical comparisons were performed using one-way ANOVA; ns, not significant (*p* > 0.05)

an immune-depleted microenvironment, evaluating synthetic lethal therapeutic strategies in immunocompetent settings, such as syngeneic mouse models, is particularly important, as treatment responses may be influenced by both tumor-intrinsic vulnerabilities and RB1-associated immune dynamics. In addition, patient-derived xenografts and organoid models may enhance translational relevance by preserving tumor heterogeneity and patient-specific characteristics. Second, although our combination strategies demonstrated favorable tolerability profiles in vivo, more comprehensive toxicological characterization, including hematologic, organ-specific, and long-term safety assessments, as well as formal pharmacokinetic analyses, will be essential prior to clinical trial initiation to ensure translational feasibility. Third, although our findings are supported by multicohort analyses and functional validation, prospective multicenter studies with coordinated patient accrual across broader geographic regions and longitudinal follow-up will be essential to further validate the predictive performance of the AI-based model and to refine RB1-Bi-guided therapeutic strategies in real-world clinical settings.

In conclusion, this analysis defines a distinct subset of biologically aggressive HCCs characterized by biallelic *RB1* alterations that exhibits sensitivity to two different classes of inhibitors: mitotic regulator inhibitors and DNA damage response pathway inhibitors. The identification of these dual vulnerabilities, combined with accurate patient stratification using digital pathology AI techniques, warrants future clinical trials focused on biomarker-driven patient selection and combinations of targeted drugs.

MATERIALS AND METHODS

Cohort descriptions

AMC-TCGA discovery cohort: The AMC-TCGA discovery cohort comprised two independent genomic datasets: 206 patients with HCC who underwent surgical resection at Asan Medical Center (AMC discovery cohort) and 355 HCC cases from The Cancer Genome Atlas (TCGA-LIHC) (Supplementary Table 1). Genomic and transcriptomic data from the AMC cohort were generated in-house and are publicly available through cBioPortal and the Gene Expression Omnibus (GEO) under accession number GSE124751. TCGA-LIHC processed data were obtained from the Broad GDAC Firehose, and raw sequencing data were accessed through the Genomic Data Commons (GDC) Data Portal under approved project access (#Project 7043).

GSE164121 validation cohort: An independent in-house cohort of 81 patients with HCC who underwent primary hepatic resection between 2009 and 2013 at Asan Medical Center was used as a validation cohort. Over half of the patients presented with intermediate-to-advanced-stage disease (BCLC stage B/C,

55.6%), and 39.5% of tumors were of non-HBV etiology (Supplementary Table 2). Tumor tissues were obtained from surgical specimens archived in the Asan Medical Center Bio-Resource Center (<http://brc.amc.seoul.kr>) under IRB approval (No. 2019-1321). Targeted sequencing of key driver genes and bulk transcriptomic profiling were performed. Among these samples, one RB1-Bi tumor and one RB1-WT tumor were selected for spatial transcriptomics (ST) analysis. Normalized gene expression data from this cohort are publicly available in the GEO under accession number GSE164121.

AMC-ICI cohort: The AMC-ICI cohort consisted of in-house patients with advanced/unresectable HCC who received ICI-based therapy. Tumor tissues were obtained following written informed consent under a protocol approved by the IRB of Asan Medical Center (IRB No. 2020-0982). Clinicopathologic and outcome data were collected for response and survival analyses (Supplementary Table 2). Tumor and matched normal samples underwent whole-exome sequencing (WES), and tumor samples were subjected to bulk RNA sequencing.

MSKCC-ICI and MSKCC-sorafenib cohorts: Publicly available genomic, transcriptomic and clinical datasets from patients with unresectable HCC treated with ICI-based regimens or sorafenib were obtained from the cBioPortal for Cancer Genomics (<https://www.cbioportal.org>).⁴ These datasets were primarily used for comparative survival analyses (Supplementary Table 3).

GO30140 cohort: GO30140 is a phase Ib study evaluating the safety and preliminary efficacy of atezolizumab plus bevacizumab in patients with unresectable HCC, which laid the foundation for the phase III IMbrave150 trial (Supplementary Table 4).⁴⁸ Raw sequencing data of WES and RNA-seq data from the GO30140 clinical trial were obtained from the European Genome-phenome Archive (EGA; EGAD00001008129).

PLANet multiregion cohort: To assess spatial genomic heterogeneity, we analyzed a publicly available multiregion HCC cohort.³³ In this prospective study, a grid-based sampling protocol was used to collect 2-11 spatially distinct tumor regions per patient. Whole-genome sequencing (WGS) was performed on 490 tumor sections from 123 patients and 125 matched adjacent normal tissues. Data were downloaded from the respective repositories as described in the original publications.

Single-cell transcriptomic cohort (GSE156625): Single-cell RNA sequencing data from HCC tumors (GSE156625; *n* = 4) originally described by Sharma et al.³¹ was used to characterize tumor microenvironmental features at single-cell resolution.

Ajou cohort: An independent cohort from Ajou University Hospital (Suwon, Republic of Korea) included 79 patients with HCC who underwent surgical resection between 2019 and 2025. WES was performed on tumor specimens with matched normal tissue. Whole-slide images from this cohort were used as an independent

external validation set to assess the performance of the FR-MIL model in predicting RB1-Bi status.

Somatic mutation profiling

The mutation annotation file (MAF) format data for the AMC and TCGA cohorts was used. We further filtered out variants that failed to satisfy the following criteria: 1) total depth ≥ 30 and allele depth > 3 for single nucleotide variants (SNVs) and 2) allele depth > 5 for small insertions and deletions (indels).⁴⁹ Intronic or silent variants were also filtered out. MutSigCV2 (v.1)⁵⁰ was used to identify significantly mutated genes with a cutoff false discovery rate (FDR) $q < 0.1$.

Profiling of germline variants

Germline variants were called with VarScan2⁵¹ v.2.3.9 from whole-exome sequencing Binary Alignment Map (BAM) files for matched tumor-normal pairs. All passed variants were annotated with the Genome Aggregation Database (GnomAD; <http://gnomad.broadinstitute.org>),⁵² 1000 Genomes project (1000 Genome; <http://www.1000genomes.org>),⁵³ Exome Aggregation Consortium (ExAC; <http://exac.broadinstitute.org>),⁵⁴ ClinVar (<https://github.com/macarthur-lab/clinvar>),⁵⁵ Combined Annotation-Dependent Depletion (CADD; <https://cadd.gs.washington.edu>),⁵⁶ and the Korean Reference Genome database (KRGDB; <http://coda.nih.go.kr/coda/KRGDB>).⁵⁷ Variants located in intronic or synonymous regions, those with variant allele frequencies (VAFs) $< 10\%$ in nontumor samples, or variants with an estimated minor allele frequency $\geq 1\%$ in population databases were excluded. Pathogenic germline variants were defined as those meeting at least one of the following criteria: (1) classified as pathogenic or likely pathogenic in ClinVar; (2) protein-truncating variants (including frameshift insertions/deletions, nonsense mutations, or splice-site alterations); or (3) heterozygous germline variants demonstrating loss of heterozygosity (LOH) in the tumor tissue, evidenced by increased VAF relative to matched normal samples, and predicted to be deleterious (CADD score ≥ 15).

Defining pathogenic variants of the RB1 gene

To identify pathogenic variants, we used ClinVar data (<https://github.com/macarthur-lab/clinvar>),⁵⁸ variant types and the functional impact of variants using sorting intolerant frame tolerant (SIFT; sift.jcvi.org),⁵⁹ polymorphism phenotyping (PolyPhen; genetics.bwh.harvard.edu/pph2)⁶⁰ and CADD (cadd.gs.washington.edu)⁵⁶. Pathogenic variants were defined according to the following criteria: (1) pathogenic or likely pathogenic variants in ClinVar; (2) protein-truncating variants such as frameshift insertions, frameshift deletions, and splice site or nonsense mutations; (3) SNVs supported by 2 or more algorithms with CADD scores ≥ 20 and deleterious effects according to SIFT and PolyPhen scores; and (4) indels with CADD ≥ 20 . All benign or likely benign variants in ClinVar were discarded irrespective of their SIFT, PolyPhen, and CADD scores.⁴⁹ We manually inspected all RB1 insertion and deletion variants to filter out false positives (e.g., alignment artifacts).⁶¹ Adjusted VAFs corrected for tumor purity obtained from Sequenza for RB1 mutations were calculated as follows:

$$\text{adjusted VAF (aVAF)} = \frac{\text{VAF}_t - (1 - \text{Purity}_t) * \text{VAF}_n}{\text{Purity}_t}$$

Copy number alterations, tumor cell purity, and ploidy

Segmented copy number data were inferred from CytoScan HD arrays (for the AMC discovery cohort)⁷ and SNP 6.0 arrays (for the TCGA-LIHC cohort from GDAC firehouse). For the 6 TCGA patients with no SNP 6.0 data (TCGA-CC-5259, TCGA-CC-A8HS, TCGA-DD-A4NE, TCGA-DD-A4NG, TCGA-G3-A5SK, and TCGA-XR-A8TC), segmented copy number data were inferred from the whole exome

sequencing BAM file for matched normal-tumor pairs using Sequenza v.3.0.0.⁶² Significant copy number alterations and corresponding G-scores were identified using GISTIC2 with the default option (q -value < 0.25).⁶³ Due to differences in SNP array platforms, GISTIC2 analyses were performed independently for the TCGA-LIHC and AMC discovery cohorts. RB1 GISTIC scores obtained from separate analyses demonstrated near-perfect concordance with those derived from a combined cohort analysis ($R > 0.999$, $p < 0.001$; Supplementary Fig. 2b).

Allele-specific copy number imbalances, tumor purity, and ploidy were estimated from WES data in both cohorts using the Sequenza algorithm. These WES-derived purity and ploidy estimates were subsequently used to calculate the cancer cell fraction (CCF) for each mutation. Copy number deletions were classified into two categories: (1) homozygous deletion (copy number $\log_2$ ratio ≤ -0.8) and (2) heterozygous deletion (one-copy loss; $-0.8 < \text{copy number } \log_2 \text{ ratio} \leq -0.4$). For assessment of clonality in HCC samples harboring RB1 copy number alterations, the CCF of each copy number alteration (CCF_{CNA}) was calculated by adjusting allele-specific copy number changes according to tumor purity estimated by Sequenza using the following formula:

$$\text{CCF}_{\text{CNA}} = \left(\frac{2 \cdot \text{BAF} - 1}{[(\text{major} - 1) - ((\text{major} - 1) \cdot \text{BAF}) - ((\text{minor} - 1) \cdot \text{BAF})]} \right) \cdot \frac{1}{p}$$

where BAF represents the B-allele frequency, major and minor denote allele-specific copy numbers, and p indicates tumor purity. RB1 copy number alterations were considered clonal when $\text{CCF}_{\text{CNA}} \geq 0.7$.

Microsatellite instability analysis

MSI status was evaluated using MSIsensor-pro (v1.2.0)⁶⁴ based on paired tumor-normal sequencing data. The MSIsensor score, representing the percentage of unstable microsatellite sites, was used to define MSI categories: samples with a score below 3.5 were classified as microsatellite stable, and those with a score of 10 or greater were defined as microsatellite instability-high (MSI-H).

Evolutionary timing analysis

The relative evolutionary timing of candidate driver mutations was inferred using MutationTimeR v1.00.2 (<https://github.com/gerstung-lab/MutationTimeR>).⁶⁵ The temporal ordering of somatic mutations relative to clonal and subclonal copy number states, as well as the timing of copy number gains, was determined. The analysis was performed with 200 bootstrap iterations. Mutations were classified into the following categories based on copy number context and CCF: "Clonal [early]" – mutations on ≥ 2 copies per cell; "Clonal [late]" – mutations on 1 copy per cell with no retained allele; "Clonal [NA]" – mutations on 1 copy per cell, either on an amplified or retained allele; and "Subclonal" – mutations on < 1 copy per cell.

To further investigate tumor evolutionary trajectories, we used PhyloPicNDT (<https://github.com/broadinstitute/PhyloPicNDT>).⁶⁵ CCF was estimated using somatic single-nucleotide mutation (SSM) VAFs, reformatted copy number data from ASCAT (<https://github.com/getzlab/ASCAT-parser.git>), and tumor purity inferred via the cluster function. Mutations were clustered based on similar CCFs to reconstruct subclonal architecture. Tumor phylogenies and mutation timings were inferred using the SinglePatientTiming and LeagueModel modules within PhyloPicNDT. Additionally, Palimpsest was used to visualize the molecular timing of evolutionary changes using whole-genome sequencing (WGS) data from representative cases.⁶⁶

Gene expression profiling and pathway analysis

The AMC-TCGA discovery cohort, GSE164121 cohort, and AMC-ICI cohort were processed and normalized using the TCGA rnaseqv2

pipeline with upper quartile normalization. GSEA⁶⁷ and gene set variation analysis (GSVA)⁶⁸ were performed to identify enriched molecular pathways using the Hallmark and Reactome gene set collections in the Molecular Signatures Database (MSigDB, version 7.2)⁶⁹ and WIKI pathway.⁷⁰ GSVA analysis was performed using the nine gene sets in DNA damage response (DDR) pathways, including BER (base excision repair), MMR (mismatch excision repair), NER (nucleotide excision repair), HR (homologous recombination), FA (Fanconi anemia), and NHEJ (nonhomologous end joining), downloaded from a human DNA repair gene database (https://www.mdanderson.org/documents/Labs/Wood-Laboratory/human-dna-repair-genes.html#DNA_pol). For the ST dataset, GSEA was performed using the fgsea R package (v1.20.0)⁷¹. In addition, pathway analysis to identify downstream target molecules for *RB1* upstream regulators was performed using Ingenuity Pathways Analysis (IPA) software (QIAGEN, Redwood City, CA; www.qiagen.com/ingenuity) with the Ingenuity Knowledge Base.

Profiling of tumor-infiltrating immune cells

Immune cell infiltration was computationally inferred from bulk RNA-sequencing data using five established deconvolution tools: MCP-counter, xCell, quantIseq, EPIC, and BayesPrism. Infiltrating immune cells were profiled using CIBERSORT absolute mode with LM22 (22 immune cell types) gene signatures from the normalized gene expression data for 561 HCCs.^{72,73} The total immune score was defined as the sum of the estimated immune scores for each cell type. MCP-counter⁷⁴ quantifies the absolute abundance of immune and stromal cell populations using transcriptomic markers specific to each cell type. xCell⁷⁵ estimates enrichment scores for immune and stromal compartments based on gene signature sets. quantIseq⁷⁶ infers tumor-infiltrating immune cell fractions from RNA-seq profiles using a predefined immune signature matrix. EPIC⁷⁷ estimates the cellular composition of tumor and nontumor compartments by modeling bulk transcriptomic profiles. BayesPrism⁷⁸ performs Bayesian deconvolution to jointly estimate cell-type fractions and cell-type-specific expression from bulk RNA-seq data using single-cell RNA-seq.

Immunohistochemistry (IHC)

Tumor tissues from the AMC discovery cohort and the GSE164121 cohort were subjected to immunohistochemical analysis. Formalin-fixed, paraffin-embedded (FFPE) tissue sections were stained for CD8 using a mouse monoclonal antibody (clone C8/144B; 1:400 dilution; catalog No. NCL-L-CD8-4B11; Novocastra/Leica Biosystems, CA, USA). Immunostaining was performed on a BenchMark XT automated immunostaining platform (Ventana Medical Systems, Tucson, AZ, USA) using the OptiView DAB IHC Detection Kit, according to the manufacturer's instructions.

Quantification of CD8-positive cells

The quantification of CD8-positive cells was performed using ImageJ (v1.54),⁷⁹ following a previously described protocol (Current Protocols). Briefly, representative tumor regions were selected from whole-slide images, and DAB-positive signals corresponding to CD8 staining were extracted. Images were processed by conversion to 8-bit grayscale and thresholded to distinguish positive staining from the background. Automated particle analysis was then applied to count CD8-positive cells, and the counts were normalized to the total number of cells or analyzed tissue area, as indicated.

Mutational signatures

DeconstructSigs evaluates the contribution of mutational signatures reported in the Catalog of Somatic Mutations in Cancers (COSMIC, version 2) (<https://cancer.sanger.ac.uk/cosmic/signatures>) to the mutational profile of each sample.⁸⁰ Mutational signatures were calculated considering all somatic mutations in a given sample. The signature scores obtained were

then analyzed in association with each group using the Wilcoxon rank-sum test.

Targeted next-generation sequencing

Targeted next-generation sequencing (NGS) was performed using the NextSeq platform (Illumina, San Diego, CA, USA) with a custom-designed hybrid capture panel (AMC_HBV_8custom_sequence_PI). Capture probes were designed by Celemics Inc. (Seoul, Republic of Korea) using the SureDesign platform based on the GRCh37 reference genome. Target enrichment was carried out using the Celemics Target Enrichment Kit (Celemics Inc.). The custom panel covered approximately 0.25 Mbp and consisted of 898 probes targeting a total of 33 genes, including the entire coding exons of the selected genes. For library preparation, 100 ng of FFPE-derived genomic DNA was fragmented by sonication to a mean size of approximately 200 bp, followed by size selection and purification using CeleMag™ Clean-up Beads (Celemics Inc.). Library quality and fragment size distribution were assessed using the Agilent TapeStation system (Agilent Technologies, Santa Clara, CA, USA). DNA libraries were prepared through sequential end repair, A-tailing, and ligation of Illumina-compatible adapters using the Celemics Library Preparation Kit. Indexed libraries were pooled to a total input of 1000 ng for hybrid capture using the AMC_HBV_8custom_sequence_PI custom panel. Captured target libraries were enriched using streptavidin-coated magnetic beads, quantified by quantitative PCR, and sequenced on the NextSeq platform (Illumina Inc.) using paired-end sequencing.

Sequenced reads were aligned to the human genome reference (GRCh37/Hg19) using BWA-MEM (v0.7.18)⁸¹ with default parameters. The aligned reads were processed with the GATK package (v4.5.0.0)⁸². Raw alignments were sorted using SortSam, and then duplicate reads were marked using MarkDuplicates. Base quality scores were recalibrated using BaseRecalibrator and ApplyBQSR. Somatic single-nucleotide variants (SNVs) and short insertions or deletions (indels) were identified using Mutect2 with matched normal samples. Putative somatic mutations were annotated using Variant Effect Predictor (v110.1)⁸³ and converted to MAF format using vcf2maf (v1.6.21)⁸⁴. Somatic copy numbers were estimated using CNVkit (v0.9.12)⁸⁵ with pooled normal references constructed from all available normal samples.

Public single-cell RNA sequencing data

Public single-cell RNA (scRNA) sequencing data from HCC samples were sourced from the GEO database (GSE156625)³¹. The normalized gene expression matrix and predefined cell types and clusters were downloaded. Preprocessing of single-cell proteomics data was performed using the Seurat (v4.4.0) R package.⁸⁶ The matrix was scaled by the ScaleData function. To address batch effects among the four samples, the R package Harmony (v1.2.3) was used.⁸⁷ Dimensionality reduction was achieved through principal component analysis (PCA), with the top 30 principal components applied to harmony with default parameters. The top 30 resulting harmony factors were then used for downstream uniform manifold approximation and projection (UMAP) reduction. To define the *RB1* status, hepatocytes from each sample were extracted and aggregated into pseudobulk cells. The mean *RB1* expression of hepatocytes was calculated per sample. Based on this average expression, samples were divided into two groups: RB1-Low ($n = 2$) and RB1-High ($n = 2$). The RB1-Low group was used as a surrogate for RB1-Bi, as it partially mimics the biological features of RB1-Bi tumors that showed markedly reduced *RB1* expression in bulk data.

Spatial transcriptomic profiling (10x Genomics Visium)

For spatial transcriptomics analyses, we selected two FFPE tissues containing HCC. Selection of the 6.5 mm × 6.5 mm capture areas in hematoxylin and eosin (H&E)-stained slides for spatial transcriptomics was based on the consensus of a pathologist (COS). Tissue

sections meeting quality control standards (DV 200 $\geq$ 30%) underwent spatial transcriptomics using the Visium CytAssist Spatial Gene Expression for FFPE platform from 10x Genomics Inc. The initial steps included deparaffinization, H&E staining, imaging, and subsequent decrosslinking. Subsequent decrosslinking, tissue permeabilization, probe hybridization, and library preparation were performed meticulously according to the manufacturer's protocol. Sample libraries were sequenced on an Illumina NovaSeq 6000 platform using specified read lengths as recommended by the manufacturer.

Preprocessing of spatial transcriptomics data

Sequences and histology images were processed using 10x Genomics Space Ranger v2.0.1. Illumina basecall files generated by the sequencing instrument were converted to FASTQ format for each sample using the mkfastq command. Visium spatial expression libraries were analyzed with the count command. Image processing proceeded in two stages: first, the fiducial alignment grid of the tissue image was used to determine the orientation and position of the input image; then, the region of the tissue covered on the slide was identified. Sequencing reads were aligned to the Visium Human Transcriptome Probe Set v2.0 GRCh38-2020-A reference using the STAR (v2.7.2a)⁸⁸ aligner.

Processing and visualization of spatial transcriptomics data

Data processing and visualization were executed using the Seurat (v4.4.0)⁸⁶ R package. Initially, we filtered and normalized the data for each sample as follows: genes were removed if they were present in fewer than three spots, and only genes observed in all sequenced data were used for downstream analysis. All spots containing fewer than 100 UMI counts were also removed. After quality control and spot filtering, ST samples were normalized using the NormalizeData function in the Seurat package with default parameters as described in the vignette. A total of 2500 variable genes were identified using the FindVariableFeatures function. Next, these normalized objects were integrated using the anchor-based integration method to adjust the batch effect across the Visium slides. The integrated object was scaled using the ScaleData function, and principal components were calculated using the RunPCA function. The top 30 principal components with resolutions of 0.25 were used to identify spatial clusters by running Seurat's ST pipeline.

Cell type decomposition

For the ST data, we used a public scRNA-seq dataset as a reference⁸⁹ to obtain the distribution of cells in the spatial regions. To map the cell types found in the reference dataset to the spatial data, we performed robust cell type decomposition (RCTD) using the spacexr R package (v2.2.1)⁹⁰, with parameter doublet_mode = "full".

WES analysis of the validation cohort

Sequenced reads were aligned to the human genome reference (GRCh37/Hg19) using BWA-MEM (v0.7.18)⁸¹ with default parameters. The aligned reads were processed with the GATK package (v3.7.0)⁸². Raw alignments were sorted using SortSam, and then duplicate reads were marked using MarkDuplicates. Base quality scores were recalibrated using BaseRecalibrator and ApplyBQSR. Somatic SNVs and short insertions or deletions (indels) were identified using Mutect2 with matched normal samples. Putative somatic mutations were annotated using Variant Effect Predictor (v110.1)⁸³ and converted to MAF format using vcf2maf (v1.6.21)⁸⁴. Somatic copy numbers were estimated using CNVkit (v0.9.12)⁸⁵ with matched normal samples. Copy number deletions were classified into 2 categories: 1) homozygous deletion (copy number log2 ratio $\leq$ -0.65) and 2) heterozygous deletion (one copy loss; -0.65 < copy number log2 ratio $\leq$ -0.25). Pathogenic RB1 mutations were defined as described above in Defining pathogenic variants of the RB1 gene.

RB1 status definition using public multiregion WGS data

Publicly available processed multiregion WGS data from a previously published PLANet study³³ were used for this study. Details of the original WGS data processing performed in the original study, including read alignment, variant calling, and copy number analysis, were described previously. For copy number analysis, thresholded GISTIC2 data provided by the original study were used directly. RB1 copy number alteration was classified as homozygous deletion (-2) or heterozygous deletion (-1) based on the GISTIC2 output. Somatic mutation annotations were obtained from the processed dataset, and only pathogenic RB1 mutations were considered. Initially, the RB1 status was determined at the regional level and categorized as RB1-Bi, RB1-Mono, or RB1-WT based on integrated copy number and mutation data. Patients were classified as RB1-Bi if at least one tumor region demonstrated RB1-Bi. Patients without RB1-Bi region but with monoallelic alteration were classified as RB1-Mono, and patients without RB1 alteration were classified as RB1-WT.

Whole slide image (WSI) preprocessing

Segmentation and patching. In our pipeline, we first automatically segment tissue regions for each digitized slide at a downscaled resolution ($\times 32$) in the HSV color space to obtain a binary tissue mask. This is achieved by thresholding the saturation channel after blurring and applying morphological operations to remove artifacts such as ink markings and holes. Following tissue segmentation, we densely cropped nonoverlapping 256 $\times$ 256 pixel patches from the segmented tissue areas at 20 $\times$ magnification (0.5 microns per pixel). The number of patches extracted varies per slide, ranging from hundreds to hundreds of thousands depending on the tissue area and scan magnification. We retained only patches containing > 15% tissue content to ensure comprehensive tissue coverage and exclude background regions.

Feature extraction. After segmentation and patching, we first applied stain normalization to mitigate color variability arising from differences in tissue preparation and staining protocols across slides. Following stain normalization, we employ Virchow2⁹¹ to extract feature representations from each image patch. Virchow2 is a Vision Transformer (ViT-H)-based pathology foundation model pretrained on over 3 million whole slide images using DINOv2⁹² self-supervised learning, which enables it to learn robust and generalizable histopathological representations without requiring manual annotations. We extract features from the penultimate layer of Virchow2, converting each 256 $\times$ 256 patch into a 2560-dimensional feature vector. This approach drastically reduces processing costs and enables faster training compared to online patch processing. Additionally, using these pre-extracted features allows all slide patches to fit simultaneously in memory on a single consumer-grade GPU, further improving computational efficiency.

Cohort splits. In this study, we utilized the TCGA-LIHC dataset, which comprises a multi-ethnic population (Supplementary Fig. 11), for model development and three independent datasets for external validation. After filtering out unusable slides with severe artifacts, the TCGA-LIHC dataset consisted of 347 patient cases with 628 diagnostic slides. To ensure robust model generalization to new patients, we implemented strict patient-level stratification where all slides from a given patient were assigned exclusively to one subset. This approach prevents data leakage that could occur with slide-level splitting and provides a more realistic estimate of clinical performance. The dataset was split into training (50%, 173 patients with 307 slides), test (40%, 139 patients with 259 slides), and validation (10%, 35 patients with 62 slides) subsets (Supplementary Table 13). The inherent class imbalance (RB1-nonBi to RB1-Bi ratio of approximately 4.5 to 1) reflects clinical prevalence alterations in HCC and was maintained across all splits.

During training, we employed balanced mini-batch sampling to ensure equal representation of both classes, preventing the model from defaulting to majority-class predictions. For external validation, we used the AMC discovery dataset of 206 slides with 206 patients and the GSE164121 cohort of 80 slides with 80 patients after exclusion of one unreadable case, along with the Ajou dataset of 79 slides with 79 patients from Ajou University Hospital (Suwon, Republic of Korea), to assess model generalizability across diverse and representative data samples.

Model framework. To solve whole slide image analysis tasks with only slide-level labels, we employ a weakly supervised learning approach based on multiple instance learning (MIL).^{93–96} The method treats each WSI as a bag containing numerous patches (instances), ranging from hundreds to hundreds of thousands per slide, eliminating the need for detailed patch-level annotations. Our FR-MIL method introduces a novel feature recalibration mechanism that dynamically adjusts patch-level feature statistics across different WSIs.⁹³ The approach integrates transformer-based modules for spatial and morphological modeling with a top-k instance classifier that identifies the most relevant patches. Feature recalibration modules then enhance discriminative power by learning to normalize and adjust features based on these top-k scoring instances per slide. This recalibration process improves slide-level classification performance by producing more robust and consistent feature representations. Our approach enhances feature separability between different WSI bags through recalibration guided by a feature distance-based learning objective. Unlike prior works^{97–99} that rely on standard aggregation functions such as max, mean, and log-sum pooling, our method provides adaptive feature adjustment capabilities and greater flexibility for data-specific optimization.

Model selection. To assess model robustness and select the optimal configuration, we repeated the training and validation procedure three times using different model initializations on patient-stratified cohort divisions, ensuring no patient overlap among splits. Each model was trained for a minimum of 100 epochs with a learning rate of 0.00001 and a batch size of 1 (single slide). The framework optimizes two loss functions: a metric loss to minimize intraslide feature representation distances and a binary cross-entropy loss using diagnostic labels. To prevent bias toward the majority class, we employed balanced sampling to ensure equal representation of slide features across both classes. The model achieving the lowest average validation loss across the three repetitions was selected as the final diagnostic model.

Cell culture, reagents, and antibodies

Human HCC cell lines, including Huh7, PLC/PRF/5, and HepG2, were purchased from the Korean Cell Line Bank (KCLB; Seoul, Republic of Korea) and cultured in RPMI-1640 medium (Welgene, Seoul, Republic of Korea) supplemented with 10% FBS (Thermo Fisher Scientific, Waltham, MA, USA), 1% sodium pyruvate (Thermo Fisher Scientific), and 1% penicillin/streptomycin (Thermo Fisher Scientific) at 37 °C with 5% CO₂. All cell lines were authenticated by short tandem repeat (STR) profiling using an AmpFLSTR Identifier kit (Applied Biosystems, Foster City, CA, USA). Before use, all cells were screened for the presence of mycoplasma by PCR using a mycoplasma detection kit (Mycro-Read™, BioMAX).

The following chemical reagents were used in this study: talazoparib (BMN-673) tosylate (DC8453), SB-715992 (ispinesib) (DC5107), volasertib (BI6727) (DC7187), and MLN8237 (alisertib) (DC2016), which were purchased from DC Chemicals (Shanghai, China). Thymidine (sc-296542) was obtained from Santa Cruz Biotechnology, and Hoechst 33342 (H3570) was acquired from Thermo Fisher Scientific. All primary and secondary antibodies employed in this study are detailed in Supplementary Table 15.

Establishment of RB1-knockout cell lines by CRISPR/Cas9

To generate stable RB1-knockout (RB1^{−/−}) cells, all-in-one CRISPR plasmids with an mCherry reporter were purchased from GeneCopoeia (#HCP216131-CG01). Cells were transfected with CRISPR plasmids, selected with puromycin, and sorted for mCherry positivity by single-cell clone isolation. Knockout clones were confirmed by PCR amplification of genomic DNA and sequencing using the following primers: forward, 5'-GTTTCTCAGGGGACGTTG-3'; reverse, 5'-GTCAAGTTGAAGCCGAGACC-3'. The RB1 and control gRNA sequences used were 5'-CGGAGGACCTGCCTCTCGTC-3' and 5'-GGCTTCGCGCCGTAGTCTTA-3', respectively.

Western blot analysis

Whole cells were harvested and lysed with RIPA buffer (Cell Signaling Technology, Danvers, MA, USA) containing phosphatase (Sigma-Aldrich, St. Louis, MO, USA) and protease inhibitors (GenDEPOT, Barker, TX, USA). The protein concentration in the cell lysates was determined using the bicinchoninic acid (BCA) method (Thermo Fisher Scientific). Laemmli sample buffer (Bio-Rad, Hercules, CA, USA) was added to the samples, which were then heated at 95 °C for 10 min to denature the proteins. Protein samples (20–30 µg per lane) were separated by SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) on 10% or 15% gels, depending on the molecular weight of the target proteins. Following electrophoresis, proteins were transferred onto polyvinylidene fluoride (PVDF) membranes using a wet transfer system. The membranes were blocked with 5% nonfat dried milk in PBST (phosphate-buffered saline with 0.1% Tween-20) for 1 h at room temperature to prevent nonspecific binding. Subsequently, the membranes were incubated with primary antibodies diluted in PBST at 4 °C overnight. The membranes were then washed three times with PBST (10 min per wash) to remove unbound antibodies. Next, the membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies (1:2000 dilution in PBST) for 1 h at room temperature, followed by three additional washes with PBST (10 min per wash). Protein bands were visualized using enhanced chemiluminescence (ECL) substrate (Thermo Fisher Scientific) and detected with a ChemiDoc MP imaging system (Bio-Rad). Image acquisition and analysis were performed using Image Lab software (versions 5.1 and 5.2.1, Bio-Rad). Protein expression levels were normalized to the loading control GAPDH to ensure equal protein loading across samples.

Drug library screening

To identify therapeutic vulnerabilities associated with RB1-Bi HCC, we performed comprehensive drug screening using isogenic RB1 wild-type (RB1^{+/+}) and knockout (RB1^{−/−}) Huh7 cell lines generated via CRISPR/Cas9 genome editing. The cells were systematically screened against three annotated compound libraries purchased from Selleck Chemicals (Houston, TX, USA): the Epigenetics Compound Library (L1900, 128 compounds), the Highly Selective Inhibitor Library (L3500, 318 compounds), and the Kinase Inhibitor Library (L1200, 430 compounds). Huh7 RB1^{+/+} and Huh7 RB1^{−/−} cells (400 cells/well) were seeded in parallel in 384-well plates (Corning, #3656) and screened using an eight-dose, interplate titration format with a Liquidator-96 multiwell pipettor (Mettler Toledo, Columbus, OH). After 5 days of incubation, cell viability was assessed using AlamarBlue (Thermo Fisher Scientific). Half-maximal inhibitory concentration (IC50) values were calculated using GraphPad Prism 6.0 (GraphPad Software, La Jolla, CA). Synthetic lethality hits selective for Huh7 RB1^{−/−} cells were identified by calculating the selectivity index (SI) as follows: SI = IC50 (Huh7 RB1^{+/+})/IC50 (Huh7 RB1^{−/−}). Compounds with a selectivity index (fold IC50 difference between RB1^{+/+} and RB1^{−/−}) greater than 2 were considered potential synthetic lethal candidates and selected for further validation.

Cell viability assay

AlamarBlue reagent was prepared by dissolving 0.025% (w/v) resazurin sodium salt (Sigma–Aldrich) in sterile PBS. The solution was filtered through a 0.22 µm filter and stored in a light-protected container at 4 °C. For the assay, AlamarBlue® reagent was added directly to the cell culture medium at a volume ratio of 1:10 (v/v) and incubated at 37 °C for 4 h until significant color changes were observed. Fluorescence was measured using a SpectraMax-M5 multiwell plate reader (Molecular Devices, Sunnyvale, CA, USA) at excitation/emission wavelengths of 560/590 nm. Percentage growth was determined relative to untreated controls. Each experiment was performed at least three times, each with triplicate samples.

Cell cycle assays

Cells cultured in six-well plates were synchronized at the G1/S boundary using the double thymidine block method, as described by Chen et al.¹⁰⁰ Briefly, cells were cultured in standard growth medium until they reached approximately 30% confluence. For the first block, cells were incubated in medium containing 2 mM thymidine for 18 h, followed by washing twice with PBS. To release the cells from arrest, they were incubated in thymidine-free medium for 9 h. For the second block, cells were reincubated in medium containing 2 mM thymidine for 18 h, followed by washing twice with PBS. Finally, the synchronized cells were released into thymidine-free medium and collected at the indicated time points for cell cycle analysis.

For cell cycle analysis, cells were harvested, gently washed with cold PBS containing 2% FBS, and fixed in 70% cold ethanol overnight at −20 °C. They were then pelleted, washed, and resuspended in PBS containing 50 µg/mL propidium iodide (Sigma–Aldrich) and 0.1 mg/mL RNase A, followed by incubation at room temperature for 30 min in the dark. Flow cytometric analysis was performed using a CytoFLEX flow cytometer (Beckman Coulter, Inc., Brea, CA, USA), and data were analyzed using FlowJo software (v10.8.1).

RB1 overexpression and rescue study

RB1 overexpression and rescue experiments were performed via lentiviral transduction. Briefly, 293T cells were seeded in 6-well plates at ~50% confluence one day prior to transfection. Lentiviral particles were generated by cotransfecting 293T cells with the following plasmids using Lipofectamine 3000 (Thermo Fisher Scientific): (i) pML-CDH-CMV-RB1(human)-EF1a-CopGFP-T2A-Puro-WPRE (RB1 overexpression construct) or pCDH-CMV-MCS-EF1-CopGFP-T2A-Puro (empty vector control); (ii) packaging plasmid pCMV-dR8.2 dvpr; and (iii) envelope plasmid pCMV-VSV-G. The plasmid ratio was maintained at 3:2:1 (transfer vector: packaging plasmid: envelope plasmid). After 48 h, viral supernatants were collected, clarified by 0.45 µm filtration, and subsequently used to transduce target cells. Transduced cells were selected with puromycin, and GFP expression was monitored to confirm transduction efficiency. RB1 overexpression was validated by Western blotting, after which drug sensitivity assays were conducted to assess cell viability.

Measurement of cell viability by ATP-based luminescence assay
Cell viability was assessed using the CellTiter-Glo Luminescence Assay (Promega, USA) according to the manufacturer's instructions. Following treatment with lenvatinib or immune checkpoint inhibitors (atezolizumab, durvalumab, or pembrolizumab), cells seeded in 96-well plates were cultured for 2 or 3 days. Subsequently, an equal volume of CellTiter-Glo reagent was added to each well. The plates were shaken for 2 min to induce cell lysis and then incubated for 10 min at room temperature in the dark to stabilize the luminescent signal. Luminescence was measured using a Victor multilabel plate reader (PerkinElmer, USA).

PD-L1 expression by flow cytometry

To assess the cell surface expression of PD-L1, MDA-MB-231 and Huh7 cells were harvested using Accutase (Thermo Fisher Scientific) to preserve surface epitopes. Cells were washed with PBS and incubated with PE-conjugated monoclonal anti-PD-L1 (557924, BD Pharmingen, USA) or the corresponding PE-conjugated isotype control (555749, BD Pharmingen) for 30 minutes on ice. After washing twice with PBS, the cells were analyzed using a BD FACSCanto II flow cytometer (BD Biosciences, USA). Flow cytometry histograms were analyzed by FlowJo software, and the mean fluorescence intensity (MFI) was calculated from gated single-cell populations. Experiments were performed in triplicate.

Isolation of peripheral blood mononuclear cells (PBMCs)

Whole blood from patients with HCC was collected at Asan Medical Center under approval of the Institutional Review Board (IRB) of Asan Medical Center (IRB No. 2020-0982). PBMCs were isolated by density gradient centrifugation using Ficoll-Paque Premium (Cytiva, Sweden). Briefly, fresh blood was diluted 1:1 with PBS and layered onto Ficoll-Paque Premium at a 3:4 (v/v) Ficoll-to-diluted blood ratio. Samples were centrifuged at 400 × g for 30 min with the brake off. The mononuclear cell layer at the plasma-Ficoll interface was collected, washed with PBS, and resuspended in complete RPMI-1640 medium. PBMCs were stimulated prior to use.

Evaluation of immune cell-mediated tumor cell killing using PBMC coculture assays

PBMCs were activated with 1 µg/mL anti-CD3 (555329, BD Pharmingen) and 1 µg/mL anti-CD28 (555725, BD Pharmingen) antibodies in the presence of IL-2 (10 ng/mL; PeproTech, USA). After activation, PBMCs were collected, counted, and used as effector cells. Target tumor cells, including the positive control cell line MDA-MB-231, were seeded in microplates and allowed to adhere overnight. Activated PBMCs were added to tumor cells at an effector-to-target (E:T) ratio of 8:1 in the presence of vehicle or immune checkpoint inhibitors and cocultured. After 48 h, non-adherent PBMCs were gently removed by washing the wells twice with complete RPMI-1640 medium to ensure that only adherent tumor cells remained. Tumor cell viability was assessed using a CellTiter-Glo (CTG) luminescence assay according to the manufacturer's instructions, and luminescence was measured using a plate reader.

Immunofluorescence analysis

Cells were seeded in 48-well plates (for p-histone H3 and γ-H2AX staining) or Nunc Lab-Tek II 8-Chamber Slides (Thermo Fisher Scientific) (for microtubule and spindle staining) and treated with synthetic lethal (SL) drugs as described in the figure legends. Following treatment, they were fixed with 4% paraformaldehyde in PBS at 37 °C for 30 min. The cells were permeabilized and blocked simultaneously with a solution containing 3% bovine serum albumin (BSA) and 0.1% Triton X-100 in PBS for 30 min at room temperature. They were then incubated with primary antibodies (as listed in Supplementary Table 15) diluted in 3% BSA in PBST at 4 °C overnight. Following primary antibody incubation, the cells were washed three times with PBST (5 min per wash) and incubated with Alexa Fluor-488-conjugated secondary antibodies (as listed in Supplementary Table 15) in 3% BSA in PBST at room temperature for 1 h. They were washed three times again with PBST (5 min per wash). For p-histone H3 and γ-H2AX staining, the cells were incubated with 1 µg/mL Hoechst 33342 at room temperature for 5 min and then washed three times with PBST (5 min per wash). They were imaged using an ImageXpress Confocal HT.ai High-Content Imaging System (Molecular Devices), and MetaXpress software was used to acquire and analyze images. The mitotic index was quantified as the

percentage of phospho-histone H3 (Ser10)-positive cells relative to the total number of Hoechst 33342-stained cells. Similarly, γ -H2AX-positive cells were quantified as the percentage of phospho-histone H2AX (Ser139)-positive cells relative to the total number of Hoechst 33342-stained cells.

For microtubule and spindle staining, cells were mounted with Fluoromount-G Mounting Medium containing DAPI (Thermo Fisher Scientific) and imaged using a Carl Zeiss LSM 880 confocal microscope equipped with a 63 $\times$ oil-immersion objective. ZEN 2.3 (black version) software was used to acquire and analyze confocal images. Abnormal spindles were defined as those that did not display normal bipolar spindle formation, as seen by the absence of a clearly visible metaphase plate or disrupted radial arrays of microtubules emanating from opposite poles. For each experiment, a total of 100 mitotic Huh7 *RB1*^{-/-} cells from each treatment condition were analyzed.

DNA repair reporter assay

The DNA DSB repair assay was performed as previously described.¹⁰¹ Briefly, cells were seeded in six-well plates at approximately 50% confluence one day prior to transfection. Cotransfection was carried out using Lipofectamine 3000 (Thermo Fisher Scientific) with 1 μ g of each of the following plasmids: pLCN-DSB Repair Reporter (DDR; Addgene #98895), pCAGGS DRR mCherry Donor EF1a BFP (Addgene #98896), and pCBAScel (Addgene #26477). After 72 h, cells were harvested and analyzed for GFP and mCherry expression using a CytoFLEX flow cytometer (Beckman Coulter, Brea, CA). Flow cytometry data were processed and quantified with FlowJo software (version 10.8.1). Control transfections included pLCN-DSB Repair Reporter and pCAGGS DRR mCherry Donor EF1a BFP without pCBAScel.

Comet assay

Cells cultured in 48-well plates were treated with alisertib, ispinesib, talazoparib, or volasertib for 48 h. Lysis buffer (2.5 M NaCl, 100 mM EDTA, 10 mM Tris, and 1% Triton X-100, pH 10) and electrophoresis buffer (300 mM NaOH, 1 mM EDTA, pH 10) were freshly prepared and chilled at 4 °C for at least 60 min before use. Low melting point agarose (1%) was melted in a microwave and maintained at 37 °C in a water bath. Microscope slides were precoated with 1.5% normal melting point agarose and allowed to solidify. Cells were trypsinized and resuspended at a density of approximately 1×10^5 cells/mL in 100 μ L of 0.5% low melting point agarose at 37 °C. The cell-agarose mixture was immediately pipetted onto precoated slides and covered with a coverslip. The slides were placed at 4 °C for 10 minutes to allow the agarose layer to solidify. After gently removing the coverslip, the slides were immersed in lysis buffer and incubated in the dark at 4 °C for at least 1 hour. Following lysis, the slides were transferred to an electrophoresis tank and equilibrated with precooled electrophoresis buffer for 20 min. Electrophoresis was performed at 300 mA (approximately 25 V) for 30 min on ice. After electrophoresis, the slides were neutralized with 0.4 M Tris-HCl (pH 7.5) three times (5 min each) and air-dried at room temperature. The samples were then stained with 10 μ g/mL propidium iodide (PI) for 10 min. Images were acquired using an EVOS fluorescence microscope (Thermo Fisher Scientific). DNA damage was quantified by measuring the percentage of DNA in the tail using ImageJ software with the OpenComet plugin.

Drug combination test

Cells were seeded in 96-well plates at a density of 2000 cells per well and treated with alisertib, ispinesib, talazoparib, or volasertib, either alone or in combination, for 4 days. Drug combinations were tested using a constant ratio design based on the IC50 values of individual compounds. Cell viability was measured using the AlamarBlue assay as previously described. The combination index (CI) values were calculated using the Chou-Talalay method with

CompuSyn software (ComboSyn, Inc.). A CI value < 1.0 indicates synergism, CI = 1.0 indicates additive effects, and CI > 1.0 indicates antagonism, with values closer to 0 representing stronger synergism. To analyze synergy with nonfixed concentration ratios, we employed SynergyFinder.¹⁰² Huh7 *RB1*^{-/-} cells were seeded in 96-well plates and treated with serial dilutions of drug combinations at varying concentration ratios for 72 h. The resulting dose-response matrices were analyzed using the SynergyFinder web application (https://synergyfinder.aittokallio.group/synfin_docs/) under the HSA model to generate 2D synergy maps and calculate synergy scores.

Tumor xenograft mouse models

All animal procedures were conducted in strict compliance with the guidelines approved by the Animal Research Ethics Committee of the University of Macau (Approval number: UMARE-0153-2024). Four- to six-week-old female athymic nude mice (The Jackson Laboratory, Bar Harbor, ME) were maintained in the specific pathogen-free (SPF) Animal Facility at the University of Macau under controlled environmental conditions (temperature: 22 ± 1 °C; humidity: $50 \pm 5\%$; 12-h light/dark cycle).

For tumor implantation, viable *RB1*-isogenic cell pairs were prepared in a 1:1 mixture of phosphate-buffered saline (PBS) and Matrigel (Corning, Corning, NY). Cell suspensions containing PLC/PRF/5 *RB1*^{+/+} cells (5×10^6 cells/mouse) and PLC/PRF/5 *RB1*^{-/-} cells (5×10^6 cells/mouse) were subcutaneously injected into the right and left flanks of nude mice, respectively. Tumor growth was monitored daily until tumors became palpable. Drug treatments were initiated when tumors reached approximately 100 mm³ in volume. Mice were randomly assigned to treatment groups ($n = 8$ per group) and administered the following treatments via intraperitoneal injection: (1) Vehicle control: sterile saline containing 5% DMSO, 5% Tween-80, and 5% polyethylene glycol-400, administered every four days; (2) alisertib: 10 mg/kg administered daily; (3) ispinesib: 0.5 mg/kg administered every four days; (4) volasertib: 10 mg/kg administered twice weekly; (5) talazoparib: 0.25 mg/kg administered daily. Tumor dimensions and mouse body weights were recorded every three days to evaluate drug efficacy and potential toxicity. Tumor volume (mm³) was calculated using the ellipsoid formula: volume = length $\times$ width² $\times$ $\pi/6$, where length represents the longest diameter and width represents the perpendicular diameter. For the drug combination study, animals were randomized into seven treatment groups: two vehicle control groups, three single-agent groups (alisertib, ispinesib, or talazoparib), and two combination groups (alisertib plus talazoparib and ispinesib plus talazoparib). Because talazoparib was included in both combination regimens, the talazoparib single-agent group served as the shared control for both corresponding analyses. In the combination arms, each compound was administered at half of the dose used in the individual synthetic lethal testing, following the same dosing schedule: 5 mg/kg alisertib, 0.25 mg/kg ispinesib, and 0.125 mg/kg talazoparib. For the toxicological assessment in mice treated with drug combinations, serum alanine aminotransferase (ALT) activity was measured at the study endpoint using a commercial kit (Jiancheng, China) following the IFCC-recommended UV kinetic method, in accordance with the manufacturer's instructions. ALT activity is reported in units per liter (U/L).

IHC analysis of mouse tumor tissues

For IHC analysis, excised tumor tissues were embedded in optimal cutting temperature (OCT) compound (Sakura Finetek, Netherlands) and sectioned at 10 μ m using a Leica CM3050 S cryostat (Leica Biosystems, Germany). The tissue sections were fixed in 4% paraformaldehyde (PFA) at room temperature for 30 min. Sections were treated with 3% hydrogen peroxide (H₂O₂) at room temperature for 10 min to block endogenous peroxidase activity. Non-specific binding sites were blocked with 3% bovine serum albumin (BSA) in phosphate-buffered saline (PBS) containing 0.1% Tween 20 for 1 h at room temperature.

Following blocking, the sections were incubated overnight at 4°C with primary antibodies targeting either p-HH3 (Ser10) or γ -H2AX (Ser139) (antibody details provided in Supplementary Table 15). Sections were then washed three times with PBS (5 minutes each) and incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies for 1 hour at room temperature. After three additional washes with PBS (5 min each), antibody binding was visualized using 3,3'-diaminobenzidine (DAB) substrate solution. Sections were counterstained with hematoxylin, dehydrated through graded alcohols, cleared in xylene, and mounted with mounting medium. Digital images of the stained sections were acquired using a Hamamatsu Digital Slide Scanner NanoZoomer (Hamamatsu Photonics K.K., Japan) and analyzed using NDP.view2 software.

Transcriptomic data for the $RB1^{-/-}$ cell lines and xenograft Raw FASTQ files from RNA-sequencing data of $RB1^{-/-}$ PLC/PRF/5 xenograft samples (wild-type, $n=3$; vehicle control, $n=3$; alisertib-treated, $n=3$; ispinesib-treated, $n=3$; volasertib-treated, $n=3$) were processed using a multistep alignment approach. Initially, reads were aligned using STAR aligner v2.7.9a⁸⁸ against a combined hybrid genome of human (hg19) and mouse (mm10) references to separate human tumor-derived reads from mouse stroma-derived reads.

Using SAMtools v1.9, reads that specifically mapped to either the human or mouse genome were extracted and assigned to corresponding BAM files. Human-specific paired reads were then extracted using the SAMtools parameter -f 1 from BAM files that had been preprocessed by subtracting reads mapped to the mouse genome. Duplicate reads were subsequently identified and marked using Sambamba v0.8.2. The SamToFastq utility of Picard v2.26.6 was used to convert the processed BAM files back to paired-end FASTQ format.¹⁰³ The resulting paired-end FASTQ files were realigned to the human genome (hg19) using the TCGA RNASeqV2 pipeline with upper quartile normalization. RNA-seq data generated in this study from xenograft tumors and human HCC cell lines have been deposited in GEO under accession number GSE306330.

Statistical analysis

Statistical comparisons of continuous and categorical variables were performed using the Wilcoxon rank sum test, Student's *t* test, or one-way ANOVA, and Fisher's exact test, respectively. Survival analyses employed the Kaplan–Meier method with log-rank testing using the R package 'survival' version 3.2. Cox proportional hazards and logistic regression analyses were conducted to assess independent associations. Multivariable Cox regression analyses were primarily performed in 324 patients with complete clinical data. To evaluate the robustness of the findings and account for missing data, multivariate imputation by chained equations (MICE) was implemented using the R package mice,²⁴ and the analyses were repeated in the full cohort ($n=560$), excluding one TCGA-LIHC patient without available survival data. All analyses were performed using R version 4.0.2, with statistical significance set at two-sided $p < 0.05$.

DATA AVAILABILITY

Publicly available datasets analyzed during this study include TCGA-LIHC data from the GDC data portal (<https://portal.gdc.cancer.gov/>) under approved project access (#Project 7043), with processed data retrieved from the Broad GDAC Firehose. Genomic and transcriptomic data from the AMC discovery cohort (GSE124751) are publicly available through the GEO and cBioPortal for Cancer Genomics (<https://www.cbioportal.org>). Normalized gene expression data from the AMC validation cohort (GSE164121) and single-cell RNA sequencing data (GSE156625) are available in the GEO database. Publicly available genomics, transcriptomics, and clinical datasets from the MSKCC-ICI and MSKCC-sorafenib cohorts were obtained from cBioPortal for Cancer Genomics. Data from the G030140 clinical trial are available from the EGA under accession number EGAD00001008129. Whole-genome

sequencing data from the PLANet multiregion cohort were obtained from previously published studies, with processed data available at Zenodo (DOI: 10.5281/zenodo.12548542). RNA sequencing data generated in this study from $RB1$ -Bi mouse models and HCC cell lines are publicly available in the GEO under accession number GSE306330.

ACKNOWLEDGEMENTS

This study was supported by grants from the Basic Science Research Program through the National Research Foundation of Korea (grants 2022R1A2C3008956, RS-2022-00166674, RS-2024-00338555, RS-2025-00557349, and RS-2025-00516124); a grant from the Asan Institute for Life Sciences of Asan Medical Center (2022IP0046); a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR21C0198); Multi-Year Research Grant of University of Macau (MYRG-GRG2024-00056-FHS); Science and Technology Development Fund of Macao S.A.R. (FDCT/0139/2024/AFJ); and Shenzhen Science and Technology Innovation Commission (SGDX20220530111004031).

AUTHOR CONTRIBUTIONS

J.A. and J.H.S. conceived and designed the study, interpreted the data, prepared the initial manuscript draft and revised it based on feedback and suggestions from all authors. W.K., J.-H.O. and C.O.S. performed statistical and bioinformatics analyses, participated in the interpretation of results, and drafted and reviewed the article. X.Z., E.J.Y., Y.P., L.-J.C. and J.S.S. performed experiments, participated in the interpretation of results and drafted and reviewed the article. P.C., H.N. and S.H.P. performed AI analyses and participated in the interpretation of results and drafted and reviewed the article. J.-S.J., D.L., and B.O. collected data, participated in the interpretation of results and drafted and reviewed the article. All authors have read and approved the article.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41392-026-02815-0>.

Competing interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- Villanueva, A. Hepatocellular carcinoma. *N. Engl. J. Med.* **380**, 1450–1462 (2019).
- Singal, A. G. et al. AASLD Practice Guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. *Hepatology* **78**, 1922–1965 (2023).
- Cheng, A. L. et al. Updated efficacy and safety data from IMbrave150: Atezolizumab plus bevacizumab vs. sorafenib for unresectable hepatocellular carcinoma. *J. Hepatol.* **76**, 862–873 (2022).
- Harding, J. J. et al. Prospective genotyping of hepatocellular carcinoma: clinical implications of next-generation sequencing for matching patients to targeted and immune therapies. *Clin. Cancer Res.* **25**, 2116–2126 (2019).
- Knudsen, E. S., Pruitt, S. C., Hershberger, P. A., Witkiewicz, A. K. & Goodrich, D. W. Cell cycle and beyond: exploiting new RB1 controlled mechanisms for cancer therapy. *Trends Cancer* **5**, 308–324 (2019).
- Dyson, N. The regulation of E2F by pRB-family proteins. *Genes Dev.* **12**, 2245–2262 (1998).
- Ahn, S. M. et al. Genomic portrait of resectable hepatocellular carcinomas: implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology* **60**, 1972–1982 (2014).
- Cancer Genome Atlas Research Network. Electronic address, w.b.e. & Cancer Genome Atlas Research, Network Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327–1341.e23 (2017).
- Xu, G. et al. Landscape of RB1 alterations in 22,432 Chinese solid tumor patients. *Ann. Transl. Med.* **10**, 885 (2022).
- Chen, L. et al. Deep whole-genome analysis of 494 hepatocellular carcinomas. *Nature* **627**, 586–593 (2024).
- Chaudhary, K. et al. Multimodal meta-analysis of 1,494 hepatocellular carcinoma samples reveals significant impact of consensus driver genes on phenotypes. *Clin. Cancer Res.* **25**, 463–472 (2019).
- Choi, S. et al. Genomic alterations in the RB pathway indicate prognostic outcomes of early-stage lung adenocarcinoma. *Clin. Cancer Res.* **21**, 2613–2623 (2015).

13. Chen, W. S. et al. Novel RB1-loss transcriptomic signature is associated with poor clinical outcomes across cancer types. *Clin. Cancer Res.* **25**, 4290–4299 (2019).
14. Feugeas, O. et al. Loss of heterozygosity of the RB gene is a poor prognostic factor in patients with osteosarcoma. *J. Clin. Oncol.* **14**, 467–472 (1996).
15. Condorelli, R. et al. Polyclonal RB1 mutations and acquired resistance to CDK 4/6 inhibitors in patients with metastatic breast cancer. *Ann. Oncol.* **29**, 640–645 (2018).
16. Bollard, J. et al. Palbociclib (PD-0332991), a selective CDK4/6 inhibitor, restricts tumour growth in preclinical models of hepatocellular carcinoma. *Gut* **66**, 1286–1296 (2017).
17. Knudson, A. G. Jr Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* **68**, 820–823 (1971).
18. Totoki, Y. et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* **46**, 1267–1273 (2014).
19. Bielski, C. M. et al. Widespread selection for oncogenic mutant allele imbalance in cancer. *Cancer Cell* **34**, 852–862.e4 (2018).
20. Bernard, E. et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nat. Med.* **26**, 1549–1556 (2020).
21. Hwang, H. S. et al. Prognostic molecular indices of resectable hepatocellular carcinoma: implications of S100P for early recurrence. *Ann. Surg. Oncol.* **28**, 6466–6478 (2021).
22. Hassler, M. et al. Crystal structure of the retinoblastoma protein N domain provides insight into tumor suppression, ligand interaction, and holoprotein architecture. *Mol. Cell* **28**, 371–385 (2007).
23. Xiao, B. et al. Crystal structure of the retinoblastoma tumor suppressor protein bound to E2F and the molecular basis of its regulation. *Proc. Natl. Acad. Sci. USA* **100**, 2363–2368 (2003).
24. Li, P., Stuart, E. A. & Allison, D. B. Multiple imputation: a flexible tool for handling missing data. *JAMA* **314**, 1966–1967 (2015).
25. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3**, 246–259 (2013).
26. Allera-Moreau, C. et al. DNA replication stress response involving PLK1, CDC6, POLQ, RAD51 and CLASPIN upregulation prognoses the outcome of early/mid-stage non-small cell lung cancer patients. *Oncogenesis* **1**, e30 (2012).
27. Teoh, N. C. et al. Defective DNA strand break repair causes chromosomal instability and accelerates liver carcinogenesis in mice. *Hepatology* **47**, 2078–2088 (2008).
28. Moynahan, M. E. & Jasin, M. Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat. Rev. Mol. Cell Biol.* **11**, 196–207 (2010).
29. Calderaro, J. et al. Histological subtypes of hepatocellular carcinoma are related to gene mutations and molecular tumour classification. *J. Hepatol.* **67**, 727–738 (2017).
30. Sia, D. et al. Identification of an immune-specific class of hepatocellular carcinoma, based on molecular features. *Gastroenterology* **153**, 812–826 (2017).
31. Sharma, A. et al. Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma. *Cell* **183**, 377–394.e21 (2020).
32. Bielski, C. M. et al. Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat. Genet.* **50**, 1189–1195 (2018).
33. Chen, J. et al. A multimodal atlas of hepatocellular carcinoma reveals convergent evolutionary paths and ‘bad apple’ effect on clinical trajectory. *J. Hepatol.* **81**, 667–678 (2024).
34. Zandarashvili, L. et al. Structural basis for allosteric PARP-1 retention on DNA breaks. *Science* **368** (2020).
35. Inoue, K. & Fry, E. A. Haploinsufficient tumor suppressor genes. *Adv. Med. Biol.* **118**, 83–122 (2017).
36. Woo, H. G. et al. Association of TP53 mutations with stem cell-like gene expression and survival of patients with hepatocellular carcinoma. *Gastroenterology* **140**, 1063–1070 (2011).
37. Dantzer, C. et al. Specific features of ss-catenin-mutated hepatocellular carcinomas. *Br. J. Cancer* **131**, 1871–1880 (2024).
38. Iovino, F., Lentini, L., Amato, A. & Di Leonardo, A. RB acute loss induces centrosome amplification and aneuploidy in murine primary fibroblasts. *Mol. Cancer* **5**, 38 (2006).
39. Alderton, G. K. Genome instability: RB moonlights in mitosis. *Nat. Rev. Mol. Cell Biol.* **11**, 541 (2010).
40. Vélez-Cruz, R. et al. RB localizes to DNA double-strand breaks and promotes DNA end resection and homologous recombination through the recruitment of BRG1. *Genes Dev.* **30**, 2500–2512 (2016).
41. Cook, R. et al. Direct involvement of retinoblastoma family proteins in DNA repair by non-homologous end-joining. *Cell Rep.* **10**, 2006–2018 (2015).
42. Pilie, P. G., Gay, C. M., Byers, L. A., O'Connor, M. J. & Yap, T. A. PARP inhibitors: extending benefit beyond BRCA-mutant cancers. *Clin. Cancer Res.* **25**, 3759–3771 (2019).
43. O'Connor, O. A. et al. Randomized Phase III Study of Alisertib or Investigator's Choice (Selected Single Agent) in Patients With Relapsed or Refractory Peripheral T-Cell Lymphoma. *J. Clin. Oncol.* **37**, 613–623 (2019).
44. Burris, H. A. III et al. A phase I study of ispinesib, a kinesin spindle protein inhibitor, administered weekly for three consecutive weeks of a 28-day cycle in patients with solid tumors. *Investig. N. Drugs* **29**, 467–472 (2011).
45. Platzbecker, U. et al. Volasertib as a monotherapy or in combination with azacitidine in patients with myelodysplastic syndrome, chronic myelomonocytic leukemia, or acute myeloid leukemia: summary of three phase I studies. *BMC Cancer* **22**, 569 (2022).
46. Lheureux, S. et al. A clinical and molecular phase II trial of oral ENMD-2076 in Ovarian Clear Cell Carcinoma (OCCC): a study of the princess margaret phase II consortium. *Clin. Cancer Res.* **24**, 6168–6174 (2018).
47. Wu, C. et al. Targeting AURKA-CDC25C axis to induce synthetic lethality in ARID1A-deficient colorectal cancer cells. *Nat. Commun.* **9**, 3212 (2018).
48. Lee, M. S. et al. Atezolizumab with or without bevacizumab in unresectable hepatocellular carcinoma (GO30140): an open-label, multicentre, phase 1b study. *Lancet Oncol.* **21**, 808–820 (2020).
49. Oh, J. H. et al. Spontaneous mutations in the single TTN gene represent high tumor mutation burden. *NPJ Genom. Med.* **5**, 33 (2020).
50. Lawrence, M. S. et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
51. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
52. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
53. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
54. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
55. Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).
56. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–d894 (2019).
57. Jung, K.S. et al. KRGDB: the large-scale variant database of 1722 Koreans based on whole genome sequencing. *Database* **2020** (2020).
58. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–d1067 (2018).
59. Sim, N. L. et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452–W457 (2012).
60. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
61. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192 (2013).
62. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
63. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
64. Jia, P. et al. MSIsensor-pro: Fast, Accurate, and Matched-normal-sample-free Detection of Microsatellite Instability. *Genom. Proteom. Bioinforma.* **18**, 65–71 (2020).
65. Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128 (2020).
66. Shinde, J. et al. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics* **34**, 3380–3381 (2018).
67. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
68. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* **14**, 7 (2013).
69. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).
70. Agrawal, A. et al. WikiPathways 2024: next-generation pathway database. *Nucleic Acids Res.* **52**, D679–D689 (2024).
71. Korotkevich, G. S.V., Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv* (2019).

72. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
73. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
74. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
75. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
76. Finotello, F. et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* **11**, 34 (2019).
77. Racle, J. & Gfeller, D. EPIC: a tool to estimate the proportions of different cell types from bulk gene expression data. *Methods Mol. Biol.* **2120**, 233–248 (2020).
78. Ho, D. W. et al. Single-cell RNA sequencing shows the immunosuppressive landscape and tumor heterogeneity of HBV-associated hepatocellular carcinoma. *Nat. Commun.* **12**, 3684 (2021).
79. Hartig, S.M. Basic image analysis and manipulation in ImageJ. *Curr. Protoc. Mol. Biol.* **14.15** (2013).
80. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).
81. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
82. Auwera, G.v.d. & O'Connor, B.D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*, (O'Reilly Media, 2020).
83. McLaren, W. et al. The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
84. Kandoth, C. et al. mskcc/vcf2maf: vcf2maf v1.6.16. v1.6.16 edn (Zenodo, 2018).
85. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873 (2016).
86. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
87. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
88. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
89. Liu, Y. et al. Identification of a tumour immune barrier in the HCC microenvironment that determines the efficacy of immunotherapy. *J. Hepatol.* **78**, 770–782 (2023).
90. Cable, D. M. et al. Cell type-specific inference of differential expression in spatial transcriptomics. *Nat. Methods* **19**, 1076–1087 (2022).
91. Zimmermann, E., Vorontsov, E. & Viret, J. Virchow2: scaling self-supervised mixed magnification models in pathology. Preprint at: <https://doi.org/10.48550/arXiv.2408.00738> (2024).
92. Oquab, M.E.A. DINOv2: Learning robust visual features without supervision. Preprint at: <https://doi.org/10.48550/arXiv.2304.07193> (2023).
93. Chikontwe, P. et al. FR-MIL: distribution re-calibration-based multiple instance learning with transformer for whole slide image classification. *IEEE Trans. Med. Imaging* **44**, 409–421 (2025).
94. Chikontwe, P., Kang, M., Luna, M., Nam, S. & Park, S.H. Low-shot prompt tuning for multiple instance learning based histology classification. In *Proc. Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, Vol. 15004 (ed. Linguraru, M.G.) (Springer, 2024).
95. Chikontwe, P. et al. Feature Re-calibration Based Multiple Instance Learning for Whole Slide Image Classification. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, Vol. 13432 (eds. Wang, L., Dou, Q., Fletcher, P.T., Speidel, S. & Li, S.) (Springer, 2022).
96. Ilse, M., Tomczak, J. & Welling, M. Attention-based deep multiple instance learning. In *Proceedings of the International Conference on Machine Learning* 2127–2136 (2018).
97. Carbonneau, M. A., Granger, E., Raymond, A. J. & Gagnon, G. Robust multiple-instance learning ensembles using random subspace instance selection. *Pattern Recognit.* **58**, 83–99 (2016).
98. Couture, H.D., Marron, J.S., Perou, C.M., Troester, M.A. & Niethammer, M. Multiple Instance Learning for Heterogeneous Images: training a CNN for Histopathology. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Vol. 11071 (eds. Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C. & Fichtinger, G.) (Springer, 2018).
99. He, C., Shao, J., Zhang, J. & Zhou, X. Clustering-based multiple instance learning with multi-view feature. *Expert Systems with Applications* **162**, 113027 (2019).
100. Chen, G. & Deng, X. Cell synchronization by double thymidine block. *Bio Protoc.* **8** (2018).
101. Arnoult, N. et al. Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. *Nature* **549**, 548–552 (2017).
102. Ianevski, A., Giri, A. K. & Aittokallio, T. SynergyFinder 2.0: visual analytics of multi-drug combination synergies. *Nucleic Acids Res.* **48**, W488–W493 (2020).
103. Rokita, J. L. et al. Genomic profiling of childhood tumor patient-derived xenograft models to enable rational clinical trial design. *Cell Rep.* **29**, 1675–1689.e9 (2019).



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026