Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of monitoring panels and hub genes for heterogeneous relapse outcomes in acute lymphoblastic leukemia

Xin He
Chongqing University Jiangjin Hospital
Jianhong Zhang
Chongqing University Jiangjin Hospital
Guowei He
Chongqing University Jiangjin Hospital
Jing Shi
First Affiliated Hospital of Chongqing Medical University
Bo Tian
Chongqing Tenth People's Hospital
Jing Huang
Chongqing University Jiangjin Hospital
Feng Li
Chongqing University Jiangjin Hospital
Debing Xiang
Chongqing University Jiangjin Hospital
Bin Tang
Chongqing University Jiangjin Hospital
Qian Lou
Chongqing Fifth People's Hospital
Delu Gan
Second Affiliated Hospital of Chongqing Medical University
Pu Li
lipu.cqu@cqu.edu.cn

Chongqing University Jiangjin Hospital

Research Article

Keywords: Acute Lymphoblastic Leukemia, Relapse Outcome, Monitoring Model, Serological Panel, Hub Gene

Posted Date: April 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8999580/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Relapse is the leading cause of mortality in acute lymphoblastic leukemia (ALL) patients. Minimal residual disease (MRD) detection is effective for bone marrow relapse (BMR) but less so for central nervous system relapse (CR) and combined bone marrow-central nervous system relapse (BMCR). Furthermore, there remains an unmet need for more effective therapeutic targets tailored to distinct relapse outcomes in ALL. Consequently, the identification of minimally invasive, serological biomarkers capable of monitoring diverse relapse outcomes, coupled with the exploration of hub genes underlying these outcomes, holds significant clinical relevance.

Methods

Clinical (n = 927), whole-exome mutation (n = 618), copy number variation (CNV, n = 250), and transcriptomic (n = 343) data were collected from the TARGET database for ALL patients with four outcomes: no relapse (NR), BMR, CR, and BMCR. The characteristics of each dataset were analysed. Serological panels and monitoring models for each relapse outcome were screened and constructed using optimized XGBoost and SHAP algorithms, with 50 random training iterations and top 5 indicator combinations. Hub genes were identified via the STRING database and Boruta algorithm (50 iterations). Functional impacts of representative hub genes on ALL progression were validated in cellular models.

Results

Multi-omics analysis showed higher mortality in relapse groups than in NR, with BMCR posing the highest risk. Elevated mutation burden in BMCR was linked to poor survival. Distinct mutational profiles, CNV signatures, and transcriptomic dysregulation patterns were observed across groups, contributing to heterogeneous relapse outcomes. Optimal serological panels and monitoring models were constructed for each group: NR (LRBA/SH2B1A/IL36G/ADH1A/CEP85, AUC = 0.985), BMR (CD22/CRMP1/EPHA2/KLK14, AUC = 0.997), CR (ACVRL1/FAT3/SKAP1, AUC = 0.990), and BMCR (AGRN/CP/EFEMP1/LGALS7/ST6GAL2, AUC = 0.999), all superior to existing models (AUC < 0.903). Hub gene screening yielded 18 candidates for BMR (e.g., VAMP2, EFNB2), 13 for CR (e.g., MT-ATP8, CXCR3), and 3 for BMCR (e.g., TGFB1, PLCG1). Experimental validation demonstrated that knockdown of CD74 (BMR) and TGFB1 (BMCR)-highly expressed hub genes-significantly inhibited ALL cell proliferation and induced apoptosis, whereas knockdown of MT-CO1 (CR)-a low-expressed hub gene-produced the opposite effect.

Conclusion

This study established serological panels for diverse ALL relapse outcomes and identified hub genes with therapeutic potential, providing a foundation for early relapse surveillance and targeted interventions in ALL.

Introduction

Relapse remains a pivotal challenge impacting long-term survival and quality of life in acute lymphoblastic leukemia (ALL), a malignancy originating from lymphoid progenitor cells with complex pathogenesis involving dysregulated gene expression networks. Despite therapeutic advances, 15–30% of ALL patients experience relapse, with post-relapse 5-year survival rates plummeting to 36% (1–2). Relapse manifests in diverse anatomical patterns, including isolated bone marrow relapse (BMR, 70–80% of cases), isolated central nervous system relapse (CR, ~ 10–15% of cases), and combined bone marrow-central nervous system relapse (BMCR, ~ 5% of cases) (3). Notably, adult ALL patients with central nervous system (CNS) relapse involvement exhibit dismal prognoses, with a median survival duration of less than one year (4–8). The heterogeneity of relapse patterns and persistent high relapse rates pose a significant challenge to improving ALL patients survival outcomes. Consequently, monitoring different recurrence outcomes and identifying hub genes from networks holds critical implications for early therapeutic intervention and survival improvement in high-risk ALL populations.

Minimal residual disease (MRD) detection, currently the cornerstone of ALL relapse surveillance, demonstrates high sensitivity in monitoring BMR. However, its performance declines in CR and BMCR, with reduced sensitivity and specificity for distinguishing distinct relapse outcomes (9–10). Furthermore, MRD assessment requires invasive bone marrow aspiration, posing procedural risks and patient discomfort while risking false-negative results due to sampling variability in timing and anatomical location (11–12). Serological biomarkers, as a non-invasive alternative, could circumvent these limitations (13–17), yet no validated serological monitoring tool exists for ALL relapse. Consequently, the development of minimally invasive, outcome-specific serological models with high predictive accuracy remains an unmet clinical need. Beyond diagnostic limitations, the pathophysiological mechanisms underlying ALL relapse outcomes remain poorly characterized. Although mutations in TP53, NOTCH1, and epigenetic alterations have been implicated in relapse pathogenesis (18–23), these studies predominantly adopt single-omics approaches or focus on isolated genes, lacking multidimensional integration and systemic network analysis (24). This fragmented perspective hinders comprehensive understanding of relapse heterogeneity. To address this limitation, multi-omics integration and holistic network modeling are essential to delineate relapse-specific molecular signatures and identify hub genes regulating these distinct outcomes. Such insights would enable precision intervention strategies targeting relapse-driving genes, thereby improving clinical outcomes in high-risk ALL.

The Therapeutically Applicable Research to Generate Effective Treatments (TARGET) database integrates multi-dimensional datasets across seven malignancies, including ALL, providing a critical resource for identifying predictive, therapeutic, and prognostic targets in cancer progression (25–26). This study utilized clinical, whole-exome mutation, copy number variation (CNV), and transcriptomic data from ALL patients to systematically investigate the biological foundations of four relapse outcomes: no relapse (NR), BMR, CR, and BMCR. Specifically, we firstly elucidated associations between clinical variables and relapse outcomes, alongside survival prognosis evaluation for each relapse group. Simultaneously, our research characterized molecular alterations across genomic, epigenomic, and

transcriptomic layers to delineate relapse-specific patterns. Subsequently, following the selection of the optimal parameters for the XGBoost training model using 5-fold cross-validation and grid search, the prioritisation of features through SHAP interpretability analysis, the evaluation of the model's discrimination and calibration using AUC and HL p-values, the 50 random trainings based on optimised algorithm (27–31), and the further training of the top five indicators in various combinations, the optimal serum panels for monitoring different recurrence outcomes were screened. Ultimately, we constructed protein-protein interaction networks and applied the Boruta algorithm to identify hub genes significantly associated with each relapse group, and the functional validation of candidate hub genes (e.g., CD74, TGFB1, MT-CO1) was performed using in vitro cellular assays to confirm their biological relevance in ALL pathogenesis. Our research employs a multi-omics perspective to analyze clinical issues, constructing serological models for minimally invasive relapse monitoring and seeking mechanistic insights into relapse outcomes. These findings hold significant potential to transform clinical management strategies by enabling early relapse detection and precision targeting of relapse-driving pathways in ALL.

Materials and methods

1. Clinical Data Acquisition

Data for this study were sourced from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) database (https://portal.gdc.cancer.gov/analysis_page?app=CohortBuilder&tab=general). We retrieved open-access samples with a primary diagnosis of acute lymphoblastic leukaemia (ALL) from the TARGET-ALL-P1, TARGET-ALL-P2, and TARGET-ALL-P3 datasets in this database. The samples encompassed clinical, whole-exome mutation, CNV, and transcriptomics data. Following the processes of data merging and collation, samples that satisfied the subsequent criteria were selected for further analysis. Firstly, the cancer type was primary blood-derived cancer. Secondly, the primary tumor site was bone marrow. Thirdly, the follow-up recurrence outcomes were clearly classified as one of the following four types: NR, BMR, CR, and BMCR. Ultimately, a total of 927 clinical samples met these inclusion criteria, including 618 with whole-exome mutation data, 250 with CNV profiles, and 343 with transcriptomic data. With the exception of transcriptomics data, the sample sizes for the remaining data were consistent with the number of clinical cases. The detailed clinical characteristics are summarized in supplementary materials.

2. Clinical Data Analysis

The clinical dataset comprised seven categorical variables, including *Gender* and *Race*. We tabulated the frequency of each categorical variable across the four relapse outcomes and quantified significance and strength of association using the $assocstats$ function from the R package vcd. Pearson's chi-squared test was employed for contingency tables with sample sizes ≥ 40 and all cell counts ≥ 5 to analyse the association significance. While the likelihood ratio chi-squared test was used for tables containing cell counts < 5 to analyse the association significance. The chi-square test $p < 0.05$ denoted statistical

significance. Association strength was measured using the Contingency Coefficient and Cramer's V, with values approaching 1 indicating stronger associations and values near 0 suggesting weaker associations.

3. Genomic Mutation Analysis

To investigate mutation profiles across ALL relapse outcomes, whole-exome mutation data were analyzed using the R package $maf \rightarrow ols$. After importing data via the $read.maf$ function, mutation burden was compared between ALL relapse groups and TCGA pan-cancer cohorts using $tcgaCompare$. Differential mutation burden across relapse groups was assessed with the tmb function, and mutation summary plots were generated using $plotmafSummary$. Mutational signature analysis was performed using cosine similarity scoring against 60 COSMIC mutational signatures (v3.2), incorporating the $BS \geq nome.Hsa\pi ens.UCSC.hg38$ reference genome and NMF package. Pathway enrichment analysis evaluated the contribution of mutated genes to 24 established oncogenic signaling pathways using the $pathways$ function. The mutant data are available in the supplementary materials.

4. CNV Analysis

To analyze CNV differences across ALL relapse outcomes, genes with a copy number of 2 were labeled as "neutral", while those with copy numbers other than 2 were classified as "change". Four pairwise comparisons were performed: NR vs. Other (non-NR samples), BMR vs. Other (non-BMR samples), CR vs. Other (non-CR samples), and BMCR vs. Other (non-BMCR samples). For each comparison, 2×2 contingency tables (rows: CNV status; columns: relapse groups) were constructed, and chi-squared tests were applied to identify differentially altered CNV genes. Additionally, genes were categorized as "loss" (copy number < 2), "neutral" (copy number = 2), or "gain" (copy number > 2). The frequency of gain/loss events was calculated as the proportion of affected samples within each relapse group to quantify genomic instability patterns. The CNV data are available in the supplementary materials.

5. Transcriptomic Differential Expression Analysis

RNA-seq count data were normalized and analyzed for differential expression using the R package DESeq2. Pairwise comparisons (NR vs. Other, BMR vs. Other, CR vs. Other, BMCR vs. Other) identified differentially expressed genes (DEGs) with thresholds of $p < 0.05$ and $|\log_2 FC| > 1$. The RNA-seq data are available in the supplementary materials.

6. Principal Component Analysis (PCA)

To visualize transcriptomic heterogeneity across relapse groups, DESeq2-normalized counts underwent $\log_2(x + 1)$ transformation. Dimensionality reduction and visualization were performed using PCA from

the R package factoextra, with the $fviz_pca_\in d$ function generating PCA plot to illustrate sample clustering patterns.

7. Gene Biotype Annotation

Gene biotypes for DEGs were obtained using the $a \cap oGe \neq$ function from the R package AnnoProbe, enabling functional classification of differentially expressed transcripts (e.g., protein-coding, lncRNA, pseudogene).

8. Enrichment Analysis

For differentially altered CNV genes, chromosome enrichment was assessed using the $MSigDBc1.all.v2023.2.Hs.symbols.gmt$ file and the $enricher$ function from the R package clusterProfiler. Gene Ontology(GO) enrichment was performed via the DAVID database (https://david.ncifcrf.gov/) with parameters counts = 2 and ease = 0.1. Both adopt over-representation analysis (ORA) for enrichment. For DEGs, GO and Reactome pathway enrichments were conducted using MSigDB's $c5.g \odot bp.v2024.1.Hs.symbols.gmt$ and $c2.cp.reac \rightarrow me.v2024.1.Hs.symbols.gmt$ files, respectively, with Gene Set Enrichment Analysis (GSEA). This approach enabled assessment of pathway activation or inhibition status and quantification of effect magnitude via normalized enrichment scores (NES): NES > 0 indicates pathway activation, NES < 0 indicates suppression, and absolute NES values positively correlate with effect strength. To characterize genes in optimal monitoring panels, we retrieved functional annotations from the GeneCards database (https://www.genecards.org/) and highlighted representative GO pathways from the 'Pathways' section. To explore regulatory roles of hub genes, GO ORA were conducted through DAVID database with parameters counts = 1 and ease = 1.

9. Plasma Protein Curation and Data Integration

Plasma protein-coding genes were compiled from the Human Protein Atlas (HPA, https://www.proteinatlas.org/) and the Human Plasma Proteome Project (HPPP, https://www.hupo.org/plasma-proteome-project/). The union of these datasets was used to filter plasma protein-coding candidates from DEGs. The plasma protein-coding genes are available in the supplementary materials.

10. Disease Database Mining

The MalaCards human disease annotation database (https://www.malacards.org/) was queried to assess associations between hematopoietic malignancies and genes in optimal panels/hub gene sets, utilizing gene-disease association scores.

11. Optimal Panel Selection and Monitoring Model Development

Four relapse outcomes (NR vs. Other, BMR vs. Other, CR vs. Other, BMCR vs. Other) were modeled as binary endpoints, with upregulated plasma protein-coding DEGs serving as predictive features. Pre-trained XGBoost models were optimized via 5-fold cross-validation and grid search for hyperparameter tuning. Feature importance was quantified using SHAP values from the R package SHAPforxgboost. Model discrimination (AUC) and calibration (Hosmer-Lemeshow p-value) were evaluated with the tidymodels and PredictABEL packages. To enhance robustness, pre-training was repeated across 50 random seeds and results were retained with criteria of AUC < 0.95 and HL $p < 0.05$. Genes were ranked in descending order by mean SHAP values calculated under retention criteria, and the top 5 genes from each group were combined to generate 31 candidate panels. Each panel underwent retraining under reserved pre-training seeds with optimal parameters, The optimal panel for each relapse outcome was selected based on maximum mean AUC with mean HL $p > 0.05$. The training data are available in the supplementary materials.

12. Hub Gene Identification

Interaction networks for DEGs in each relapse group were constructed using the STRINGdb package (STRING database version 12.0, Homo sapiens, minimum interaction score threshold of 700). Degree values (number of direct interactions) for each gene node were calculated using the igraph package. For BMR vs. Other, CR vs. Other, and BMCR vs. Other comparisons, DEGs with degree > 10 were selected as input features for the Boruta algorithm. To enhance stability, Boruta training was repeated across 50 random seeds, with features classified as "Confirmed" (definitively important), "Tentative" (potentially important), or "Rejected" (not important) based on variable importance. Genes that were classified as "Confirmed" in all 50 iterations and were unique to one group were retained as group-specific hub genes.The networks and Boruta results are available in the supplementary materials.

13. Quantitative real-time PCR (qPCR)

Total RNA was isolated by TRIzol® Reagent (Tiangen, China) and converted to cDNA using a reverse transcription kit (Tiangen, China). Then, the expression of target gene was detected by real-time PCR (SYBR green) with specific primers. The PCR reaction conditions were as follows: 95°C for 3 min, followed by 40 cycles of 95°C for 15s, 60°C for 15s, and 72°C for 20 s, concluding with an extension step of 72°C for 5min. The data were analyzed using $2^{-\Delta\Delta Ct}$ method. Primers were listed in the supplementary materials.

14. Cell lines and culture conditions

The human Ph + ALL cell line SUP-B15 and human embryonic kidney cell line 293FT were purchased from the American Type Culture Collection. SUP-B15 cells were cultured in Iscove's modified Dulbecco's

medium (IMDM) (Cytiva, Cat# SH30228.FS) supplemented with 1% penicillin-streptomycin, 20% FBS (Gibco; Thermo Fisher Scientific, Inc.), and 0.05 mM β-mercaptoethanol (Gibco, Cat# 21985-023). 293FT cells were cultured in DMEM (HyClone, Cat# SH30022.01) with 10% FBS. All cells were maintained in a humidified incubator at 37°C in an atmosphere of 5% CO_2.

15. Lentivirus Production and Cell Transduction

ShRNA sequences targeting CD74, MT-CO1, and TGFB1, along with a negative control sequence, were respectively cloned into the pLKO.1 lentiviral vector (addgene, #8453). Lentiviral particles were produced by co-transfecting 293FT cells with the lentiviral vector, pMD2.G, and psPAX2 viral packaging plasmids using Lipofectamine 3000 (Invitrogen, USA) according to the manufacturer's protocol. Viral supernatants were harvested at 48 h and 72 h post-transfection, and viral titers were determined before storage at -80°C. For cell infection, SUP-B15 cells were incubated with viral culture medium supplemented with 10 μg/mL polybrene. Stably transduced cells were selected using 2 μg/mL puromycin for 14 days.

16. Cell Proliferation Assay

Stable SUP-B15 cell lines were seeded at 2,000 cells/well in 96-well plates (30 replicates per sample). Cell viability was measured daily for 5 days using the Cell Counting Kit-8 (MCE, HY-K0301) according to the manufacturer's instructions, with five wells measured per day.

17. Flow Cytometry Analysis

For S-phase analysis, stable cells were pulsed with 100 μM 5-bromo-2'-deoxyuridine (BrdU, Thermo Fisher Scientific) for 1 h at 37°C before harvesting. Then cells were fixed with 70% ethanol at 4°C overnight. Next, permeabilized with 0.3% Triton X-100 for 30 min, and stained with FITC-anti-BrdU (Thermo Fisher Scientific) for 1 h at room temperature. Finally, cells were labeled with 50 ng/mL propidium iodide (PI, Thermo Fisher Scientific) and 10 μg/mL RNase A (Thermo Fisher Scientific) for 30 min at 37°C, and then analysed by flow cytometry. For apoptosis analysis, cells were stained with PI and APC-conjugated annexin V (BestBio, C1062) according to the manufacturer's protocol prior to flow cytometry. Data acquisition was performed on a CytExpert flow cytometer (Beckman Coulter, Inc), with a minimum of 20,000 events analyzed per sample.

Results

1. Clinical Characteristics of ALL Relapse Outcomes

To investigate clinical features associated with distinct relapse outcomes in ALL, we analyzed clinical data from 927 TARGET-ALL patients categorized into four groups: 708 cases in the NR group, 167 with BMR, 30 with CR, and 22 with BMCR (Fig. 1A). Assessment of the significance and strength of

associations between categorical clinical variables and various recurrence outcomes revealed no significant intergroup differences in *Race*, *MLL status*, and *Down syndrome*. However, *Gender*, *CNS status at diagnosis*, *ETV6/RUNX1 fusion status*, *TRISOMY 4/10 status*, and *TGF3 PBX1 status* differed significantly across groups (Fig. 1B), suggesting these differential variables may contribute to relapse heterogeneity. Additionally, *Vital status* showed significant difference among groups and the strongest association with relapse outcome, with BMCR patients exhibiting the highest mortality rate (77.3%, 17/22), followed by BMR (61.1%, 102/167) and CR (43.3%, 13/30) (Fig. 1B). Survival analyses also confirmed that BMCR patients had the poorest overall survival (OS) and event free survival (EFS) outcomes (Fig. 1C). The findings emphasize the significance of recurrence patterns, particularly BMCR, in determining prognosis, further highlighting the urgent need for early detection and intervention in ALL recurrence.

2. Exome-Wide Mutation Analysis in ALL Relapse Outcomes

618 ALL samples with mutation data and varying relapse outcomes from the TARGET database were analyzed to characterize mutation profiles (Fig. 1A). Pan-cancer mutation burden comparison revealed that ALL relapse groups exhibited lower overall mutation rates compared to most solid tumors but higher than acute myeloid leukemia (LAML), suggesting relatively elevated genomic instability in ALL among hematologic malignancies (Fig. 2A). Notably, the BMCR group displayed the highest differential mutation burden (Fig. 2B), aligning with its poorest survival outcomes. These suggest a potential link between elevated genomic alterations and ALL malignant progression. Mutation signature analysis presented high similarity among groups in terms of variant classification, variant type, and SNV class but divergent top 10 mutated genes (Fig. 2C, Fig. S1). Mutational etiology assessment indicated similar exposure profiles among NR, BMR, and BMCR groups, with the exception of the CR patients, implying that mutation triggers do not explain group mutation heterogeneity (Fig. 2D, Fig. S2). Pathway enrichment analysis revealed group-specific oncogenic signaling alterations: MAPK signaling (25.22% mutation frequency) in NR, genome integrity pathways (13.51%) in BMR, NOTCH signaling (35.29%) in CR, and chromatin regulation/NOTCH/ubiquitin-proteasome systems (40% combined frequency) in BMCR (Fig. 2E, Fig. S3), suggesting that distinct oncogenic pathway dysregulation may drive relapse-specific outcomes in ALL.

3. CNV Analysis in ALL Relapse Outcomes

250 ALL samples with CNV data from the TARGET database were analyzed to characterize CNV profiles across relapse outcomes (Fig. 1A). Differential CNV analysis identified 2,414, 924, 3,212, and 43 significantly altered CNV genes in the NR, BMR, CR, and BMCR groups, respectively (Fig. 3A, Fig. S4). The limited number of differential CNVs in BMCR may be influenced by its smaller sample size. Chromosomal enrichment, gene biotype annotation, and GO enrichment analyses were performed to interpret group-specific CNV patterns. Notably, NR and CR groups exhibited similar CNV profiles in chromosomal distribution, gene biotype composition, and functional enrichment, whereas BMR and BMCR groups showed distinct patterns (Fig. 3B-D). Comprehensive analysis of the top 30 differential

CNVs revealed predominant CNV loss in NR and CR groups, contrasting with CNV gain in BMR and BMCR groups. However, comparisons to control cohorts indicated increased loss frequency in NR, reduced loss frequency in CR, decreased gain frequency in BMR, and elevated gain frequency in BMCR (Fig. 3E). These findings elucidate CNV characteristics associated with distinct relapse outcomes, providing novel insights into the mechanisms driving ALL relapse heterogeneity.

4. Transcriptomic Analysis of ALL Relapse Outcomes

A comprehensive transcriptomic analysis was performed on 343 ALL samples with RNA-sequencing data and relapse outcomes from the TARGET database (Fig. 1A). PCA plots revealed group-specific expression patterns. However, the NR and CR groups exhibited high similarity, as did the BMR and BMCR groups (Fig. 4A). This transcriptional similarity aligns with previous CNV-based group characterizations, suggesting potential regulatory links between genomic alterations and gene expression profiles. Differential expression analysis identified group-specific DEGs (Fig. 4B, Fig. S5), which were subjected to biotype annotation and functional enrichment analysis using GO and Reactome databases. The biotype classification exhibited diversity, yet protein-coding genes were confirmed as the predominant DEG category across groups (Fig. 4C). GO enrichment revealed distinct biological processes. NR group showed suppressed immune response, cell differentiation, and apoptosis pathways. BMR exhibited enhanced cell activation, adhesion and oxidative phosphorylation. CR displayed activated type II interferon production and mitochondrial functions. BMCR demonstrated augmented cell activation and adhesion, leukocyte chemotaxis, and B-cell activation (Fig. 4D). Reactome pathway analysis corroborated these findings, showing predominant pathway suppression in NR versus activation in other groups (Fig. 4E). These transcriptomic disparities provide novel mechanistic insights into the molecular drivers of relapse heterogeneity in ALL.

5. Optimized XGBoost Combined with SHAP Algorithm Identifies Optimal Serum Monitoring Panels for ALL Relapse Outcomes

Current ALL relapse monitoring primarily relies on MRD detection, which is mainly applied to BMR surveillance but shows limited efficacy for CR/BMCR monitoring and requires invasive bone marrow sampling. Development of minimally invasive serological panels capable of accurately monitoring diverse relapse outcomes holds significant clinical value. To achieve this, we intersected plasma protein-coding genes from the HPA and HPPP databases with group-specific upregulated genes to identify candidate biomarkers for panel screening (Fig. 5A-B, Fig. S6A). Heatmaps revealed marked expression differences in the top 10 candidate genes across groups, suggesting their potential as relapse-specific markers (Fig. S6B). We optimized XGBoost models using 5-fold cross-validation and grid search, followed by SHAP analysis to rank feature importance. Fifty iterations of model pre-training were performed, with AUC values and Hosmer-Lemeshow (HL) test p-values used to filter suboptimal results and ensure robustness (Fig. S7). Mean SHAP values from retained iterations were calculated for final feature ranking (Fig. 5C). Combining top 5 features from each group generated 31 candidate panels, which were evaluated through iterative pre-training. Optimal panels satisfied with HL $p<0.05$ included: NR

panel (LRBA, SH2B1A, IL36G, ADH1A, CEP85, AUC=0.985); BMR panel (CD22, CRMP1, EPHA2, KLK14, AUC=0.997); CR panel (ACVRL1, FAT3, SKAP1, AUC=0.990); and BMCR panel (AGRN, CP, EFEMP1, LGALS7, ST6GAL2, AUC=0.999) (Fig. 5D).

Genomic analysis of panel genes revealed significant expression differences across groups but rare mutations (1% missense in FAT) and limited CNV alterations (statistically significant only in SH2D1A), suggesting expression regulation of panel genes caused by non-genomic factors (Fig. S8A-C). Functional characterization using the HPA protein class identified cancer/disease-associated genes (e.g., LRBA, SH2D1A) and FDA-approved drug targets (e.g., ADH1A, CD22) within optimal panels, supporting therapeutic relevance. Protein half-life predictions via ExPASy ProtParam indicated serum suitability of panel genes (>30 hours). Malacards analysis linked most panel genes to hematologic disorders, while GeneCards GO enrichment highlighted cell adhesion/migration pathways in BMR/CR/BMCR panels and CNS development processes in CR/BMCR panels, aligning with clinical relapse phenotypes (Fig. S8D). Survival analysis revealed poor prognosis associations with elevated levels of panel genes (CD22, LGALS7, CRMP1, AGRN, and KLK14), suggesting dual monitoring and prognostic potential (Fig. S9).

6. Identification and Validation of Hub Genes in ALL Relapse Outcomes

To identify hub genes driving ALL relapse progression and potential therapeutic targets, we constructed protein-protein interaction networks using the STRING database and group-specific DEGs from BMR, CR, and BMCR groups. Subnetworks were extracted with nodes exhibiting degree>10, representing 14.44%, 5.62%, and 10.56% of total nodes in BMR/CR/BMCR groups, respectively. These subnetworks governed 39.23%, 15.42%, and 50.41% of total interactions, highlighting their regulatory importance (Fig. S10). Boruta algorithm was applied across 50 iterations to identify group-specific hub genes with confirmed importance scores of 50, yielding 18, 13, and 3 hub genes for BMR, CR, and BMCR groups, respectively (Fig. 6A).

Multi-omics analysis revealed downregulated hub genes in CR and upregulated hub genes in BMCR, while BMR showed both up- and downregulated patterns. Notably, only ITPR2, ADCY5, and HLA-DRA exhibited missense mutations in NR samples, and CXCR3 was the sole hub gene with CNV differences across groups, suggesting expression regulation of hub genes caused by non-genomic factors (Fig. S11A, S11C-D). Functional characterization via HPA protein classes classified 28 hub genes (including ADCY5) as cancer/disease-related and 17 (including CD19) as FDA-approved/potential drug targets. Malacards analysis confirmed hematologic disease associations for all hub genes, underscoring their therapeutic relevance (Fig. S11B). DAVID GO enrichment of subnetwork-regulated hub genes revealed pathway convergence on signal transduction, biological function, immune responses, DNA replication, and transcription in BMR/BMCR groups, with additional CNS-related pathway enrichment in CR group (Fig. S12).

Representative hub genes CD74 (BMR), MT-CO1 (CR), and TGFB1 (BMCR) were selected for functional validation. CD74 and TGFB1 demonstrate elevated levels of expression in the BMR and BMCR groups,

respectively, while MT-CO1 exhibits reduced expression in the CR group (Fig. 6B). Consistent with expression profiles, shRNA-mediated knockdown in SUP-B15 cells (Fig. 6C) demonstrated that CD74/TGFB1 depletion enhanced proliferation and inhibited apoptosis, while MT-CO1 knockdown showed opposite effects (Fig. 6D-F). These results confirm the regulatory roles of identified hub genes in ALL pathogenesis, validating the screening approach.

Discussion

Relapse remains a critical cause of mortality in ALL, with distinct outcomes including BMR, CR, and BMCR (32). Current clinical surveillance predominantly relies on MRD monitoring, which is primarily limited to BMR detection and requires invasive bone marrow aspiration (11–12). Consequently, there is an urgent clinical need to develop minimally invasive serological biomarkers capable of monitoring diverse relapse outcomes. Furthermore, mechanistic insights into ALL relapse heterogeneity remain insufficient, necessitating the identification of hub genes driving relapse processes to enable early targeted interventions (33). Leveraging multi-omics data from the TARGET database, we systematically characterized clinical, genomic, CNV, and transcriptomic features across ALL relapse outcomes. This analysis yielded optimal serological monitoring panels and predictive models for each relapse group, alongside the identification of group-specific hub genes, thereby establishing novel pathways for early relapse detection and intervention in ALL.

Utilizing machine learning algorithms and transcriptomic data, we identified serological monitoring panels tailored to ALL relapse outcomes. Predictive models constructed using these panels achieved mean AUC values of 0.985, 0.997, 0.990, and 0.999 for distinguishing NR, BMR, CR, and BMCR groups from their respective controls. These panels outperform existing models such as the Seven-lncRNA-mRNA Signature (AUC = 0.901) (34) and clinical variable-based models incorporating age, WBC count, and hemoglobin levels (AUC = 0.902) (35), demonstrating superior discriminatory capacity. The high AUC values underscore the panels' efficacy in identifying distinct relapse outcomes, offering clinicians a more precise monitoring tool. Compared to conventional MRD detection methods, serum-based panels avoid invasive bone marrow sampling, reduce procedural costs, and facilitate broader clinical implementation, highlighting their significant translational potential. Notably, several panel genes (e.g., CD22, CRMP1, KLK14, AGRN, LGALS7) exhibit prognostic relevance in ALL patients. Targeted therapies against CD22 have demonstrated efficacy in refractory/relapsed ALL (36–37), validating the therapeutic relevance of our findings and prompting similar potential for other panel components. Additionally, through network analysis and Boruta algorithm training, we identified hub genes significantly associated with ALL relapse outcomes, including VAMP2 and EFNB2 in BMR, MT-ATP8 and CXCR3 in CR, and HLA-DRA and HLA-F in BMCR. CR-associated hub genes showed enrichment in mitochondrial functional pathways, aligning with reported alterations in mitochondrial DNA levels in cerebrospinal fluid from patients with CNS relapse (38), implicating mitochondrial dysfunction in ALL CNS relapse pathogenesis. Functional validation experiments confirmed that knockdown of representative hub genes (CD74, MT-CO1, TGFB1) significantly impacts ALL cell proliferation and apoptosis, further emphasizing their critical roles in ALL progression. The identification of these hub genes not only advances mechanistic understanding of ALL

relapse but also provides actionable therapeutic targets for early intervention, particularly for mitochondrial dysfunction-related strategies in CNS relapse management. Furthermore, our methodology incorporated algorithmic optimizations, including 50 randomized pre-trainings, 5-fold cross-validation, and grid search parameter tuning (39–41), which substantially enhanced the performance and stability of XGBoost models and Boruta algorithms compared to non-optimized approaches. This study establishes a robust framework for biomarker discovery in heterogeneous diseases, offering clinically translatable solutions for precision oncology (34–35).

This study has several limitations. First, the data were sourced from the TARGET database, which contains relatively fewer samples in the CR and BMCR groups compared to the NR and BMR cohorts. To mitigate overfitting risks associated with small sample sizes, we did not partition internal training/validation sets. Furthermore, no external validation datasets containing ALL relapse outcomes were identified in public repositories beyond the TARGET database. To partially address this limitation, we employed 5-fold cross-validation to enhance model robustness. Fusion genes have been reported to play significant roles in ALL pathogenesis and relapse (42–43). However, due to data type limitations in the TARGET database, we were unable to investigate their specific contributions across relapse outcomes. Additionally, this study did not validate the serological panels using clinical samples or establish primary ALL cell lines from relapse-specific cohorts for functional hub gene verification. Nevertheless, functional assessments in ALL cell lines partially compensated for these experimental gaps.

Conclusions

In conclusion, this study analyzed multi-omics characteristics of ALL relapse outcomes from the TARGET database and developed optimized algorithms to construct serological monitoring panels and identify therapeutically relevant hub genes. While clinical and experimental validations remain limited, multi-dimensional data analysis, literature corroboration, and preliminary functional experiments suggest strong translational potential for the identified panels and hub genes, providing novel theoretical foundations for early relapse monitoring and precision interventions in ALL. Future studies should incorporate multi-center clinical trials to validate panel external validity, elucidate hub gene regulatory mechanisms, and explore targeted therapeutic strategies based on these genes, with the ultimate goal of improving clinical outcomes for ALL patients.

Abbreviations

Acute lymphoblastic leukemia (ALL)

Minimal residual disease (MRD)

Bone marrow relapse(BMR)

Central nervous system relapse(CR)

Bone marrow relapse with concurrent central nervous system relapse/combined relapse(BMCR)

No relapse (NR)

Central nervous system (CNS)

Circulating tumor DNA (ctDNA)

Overall survival (OS)

Event free survival (EFS)

Skin cutaneous melanoma (SKCM)

Acute myeloid leukemia (LAML)

Single nucleotide polymorphisms (SNPs)

Oligonucleotide polymorphisms (ONPs)

Copy number variation (CNV)

Gene Ontology(GO)

Principal Component Analysis (PCA)

Gene Set Enrichment Analysis (GSEA)

Differentially expressed genes (DEGs)

The Human Protein Atlas (HPA)

Haemoglobin (HB)

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The clinical, whole-exome mutation, CNV and transcriptomic data utilized in our research were obtained from the TARGET database. The data generated during the analysis and training phases , the source code, and the improved algorithms are available at https://github.com/Vera8023/Article.git.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Joint Key Project of the Chongqing Science and Technology Bureau and the Chongqing Municipal Health Commission (2025ZDXM001), and Key Project of Chongqing Municipal Education Commission (KJZD-K202400103), Chongqing Technology Innovation and Application Development, Chuan-Yu (Sichuan-Chongqing) Scientific and Technological Innovation Cooperation Program (CSTB2024TIAD-CYKJCXX0031), 2024 Hospital-level Cultivation Project of Chongqing University Jiangjin Hospital (2024YCXM010), Research Startup Funding Project of Chongqing University Jiangjin Hospital (2025qdjfxm001, 2025qdjfxm002), Key Research Project for Enhancing Medical Service Capabilities of County-Level Medical Institutions in 2025 (PS202511), Guangdong Medical Science and Technology Research Fund Project (B2021181), and Chongqing Municipal Science and Technology Bureau, Natural Science Fund (Chongqing Science and Technology Development Foundation) Project (CSTB2024NSCQ-KJFZMSX0018), Chongqing Health Commision and Science and Technology Bureau (2026MSXM018), Chongqing Youth Outstanding Medical Talent Project (YXQN2025049)

Authors' contributions

Xin He and Pu Li wrote the paper and conceived the design of the experiment; Xin He,Jianhong Zhang and Guowei He analyzed the data; Jing Shi, Bo Tian, Jianhong Zhang and Qian Lou collated the data; Delu Gan collated the pictures, Guowei He collated the tables; Bin Tang, Jing Huang, Feng Li and Debing Xiang gave guidelines for revising the paper; Pu Li fixed the paper. All authors have read and approved the final submitted manuscript.

Acknowledgements

Not applicable

References

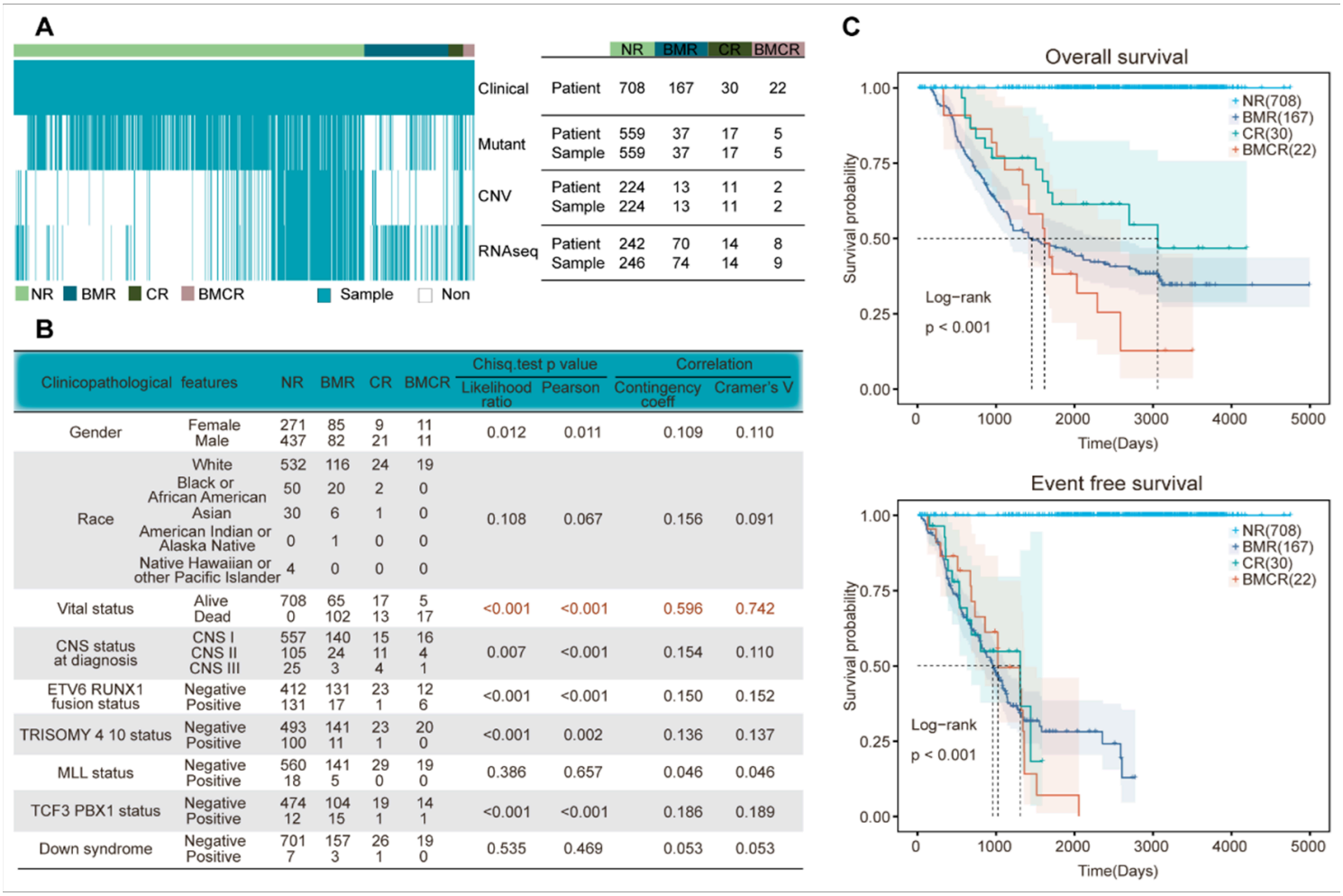
1. Hunger SP and Raetz EA. How I treat relapsed acute lymphoblastic leukemia in the pediatric population. Blood. 2020;136:1803-1812.
2. Sidhu J, Gogoi MP, Krishnan S and Saha V. Relapsed Acute Lymphoblastic Leukemia. Indian J Pediatr. 2024;91:158-167.

3. Rivera GK, Zhou Y, Hancock ML, Gajjar A, Rubnitz J, Ribeiro RC, et al. Bone marrow recurrence after initial intensive treatment for childhood acute lymphoblastic leukemia. Cancer. 2005;103:368-76.
4. Kopmar NE and Cassaday RD. How I prevent and treat central nervous system disease in adults with acute lymphoblastic leukemia. Blood. 2023;141:1379-1388.
5. Kantarjian HM, O'Brien S, Smith TL, Cortes J, Giles FJ, Beran M, et al. Results of treatment with hyper-CVAD, a dose-intensive regimen, in adult acute lymphocytic leukemia. J Clin Oncol. 2000;18:547-61.
6. Reman O, Pigneux A, Huguet F, Vey N, Delannoy A, Fegueux N, et al. Central nervous system involvement in adult acute lymphoblastic leukemia at diagnosis and/or at first relapse: results from the GET-LALA group. Leuk Res. 2008;32:1741-50.
7. Fielding AK, Richards SM, Chopra R, Lazarus HM, Litzow MR, Buck G, et al. Outcome of 609 adults after relapse of acute lymphoblastic leukemia (ALL); an MRC UKALL12/ECOG 2993 study. Blood. 2007;109:944-50.
8. Surapaneni UR, Cortes JE, Thomas D, O'Brien S, Giles FJ, Koller C, et al. Central nervous system relapse in adults with acute lymphoblastic leukemia. Cancer. 2002;94:773-9.
9. van der Velden VH, de Launaij D, de Vries JF, de Haas V, Sonneveld E, Voerman JS, et al. New cellular markers at diagnosis are associated with isolated central nervous system relapse in paediatric B-cell precursor acute lymphoblastic leukaemia. Br J Haematol. 2016;172:769-81.
10. Saygin C, Cannova J, Stock W and Muffly L. Measurable residual disease in acute lymphoblastic leukemia: methods and clinical context in adult patients. Haematologica. 2022;107:2783-2793.
11. Pierce E, Mautner B, Mort J, Blewett A, Morris A, Keng M, et al. MRD in ALL: Optimization and Innovations. Curr Hematol Malig Rep. 2022;17:69-81.
12. Bartram J, Patel B and Fielding AK. Monitoring MRD in ALL: Methodologies, technical aspects and optimal time points for measurement. Semin Hematol. 2020;57:142-148.
13. Pessoa LS, Heringer M and Ferrer VP. ctDNA as a cancer biomarker: A broad overview. Crit Rev Oncol Hematol. 2020;155:103109.
14. Cohen SA, Liu MC and Aleshin A. Practical recommendations for using ctDNA in clinical decision making. Nature. 2023;619:259-268.
15. Kamtchum-Tatuene J and Jickling GC. Blood Biomarkers for Stroke Diagnosis and Management. Neuromolecular Med. 2019;21:344-368.
16. Foster JB, Koptyra MP and Bagley SJ. Recent Developments in Blood Biomarkers in Neuro-oncology. Curr Neurol Neurosci Rep. 2023;23:857-867.
17. Sztolsztener K, Żywno H, Hodun K, Konończuk K, Muszyńska-Rosłan K and Latoch E. Apolipoproteins-New Biomarkers of Overweight and Obesity among Childhood Acute Lymphoblastic Leukemia Survivors? Int J Mol Sci. 2022;23.
18. Saiki R and Ogawa S. Adult Low-Hypodiploid Acute Lymphoblastic Leukemia Evolves from TP53-Mutated Clonal Hematopoiesis. Blood Cancer Discov. 2023;4:102-105.

19. Kim R, Bergugnat H, Larcher L, Duchmann M, Passet M, Gachet S, et al. Adult Low-Hypodiploid Acute Lymphoblastic Leukemia Emerges from Preleukemic TP53-Mutant Clonal Hematopoiesis. Blood Cancer Discov. 2023;4:134-149.
20. Baran N, Lodi A, Dhungana Y, Herbrich S, Collins M, Sweeney S, et al. Inhibition of mitochondrial complex I reverses NOTCH1-driven metabolic reprogramming in T-cell acute lymphoblastic leukemia. Nat Commun. 2022;13:2801.
21. Campagnari A and Belver L. NOTCH1-Induced T-Cell Acute Lymphoblastic Leukemia In Vivo Models. Methods Mol Biol. 2024;2773:9-24.
22. Ruan Y, Xie L and Zou A. Association of CDKN2A/B mutations, PD-1, and PD-L1 with the risk of acute lymphoblastic leukemia in children. J Cancer Res Clin Oncol. 2023;149:10841-10850.
23. Ampatzidou M, Papadhimitriou SI, Paisiou A, Paterakis G, Tzanoudaki M, Papadakis V, et al. The Prognostic Effect of CDKN2A/2B Gene Deletions in Pediatric Acute Lymphoblastic Leukemia (ALL): Independent Prognostic Significance in BFM-Based Protocols. Diagnostics (Basel). 2023;13.
24. Mengxuan S, Fen Z and Runming J. Novel Treatments for Pediatric Relapsed or Refractory Acute B-Cell Lineage Lymphoblastic Leukemia: Precision Medicine Era. Front Pediatr. 2022;10:923419.
25. Brady SW, Roberts KG, Gu Z, Shi L, Pounds S, Pei D, et al. The genomic landscape of pediatric acute lymphoblastic leukemia. Nat Genet. 2022;54:1376-1389.
26. Alexander TB, Gu Z, Iacobucci I, Dickerson K, Choi JK, Xu B, et al. The genetic basis and cell of origin of mixed phenotype acute leukaemia. Nature. 2018;562:373-379.
27. Widman AJ, Shah M, Frydendahl A, Halmos D, Khamnei CC, Øgaard N, et al. Ultrasensitive plasma-based monitoring of tumor burden using machine-learning-guided signal enrichment. Nat Med. 2024;30:1655-1666.
28. Álvez MB, Edfors F, von Feilitzen K, Zwahlen M, Mardinoglu A, Edqvist PH, et al. Next generation pan-cancer blood proteome profiling using proximity extension assay. Nat Commun. 2023;14:4308.
29. Bhat M, Rabindranath M, Chara BS and Simonetto DA. Artificial intelligence, machine learning, and deep learning in liver transplantation. J Hepatol. 2023;78:1216-1233.
30. Silva GFS, Fagundes TP, Teixeira BC and Chiavegatto Filho ADP. Machine Learning for Hypertension Prediction: a Systematic Review. Curr Hypertens Rep. 2022;24:523-533.
31. Ouyang Y, Li X, Zhou W, Hong W, Zheng W, Qi F, et al. Integration of machine learning XGBoost and SHAP models for NBA game outcome prediction and quantitative analysis methodology. PLoS One. 2024;19:e0307478.
32. Lew G, Chen Y, Lu X, Rheingold SR, Whitlock JA, Devidas M, et al. Outcomes after late bone marrow and very early central nervous system relapse of childhood B-acute lymphoblastic leukemia: a report from the Children's Oncology Group phase III study AALL0433. Haematologica. 2021;106:46-55.
33. Verma D, Kapoor S, Kumari S, Sharma D, Singh J, Benjamin M, et al. Decoding the genetic symphony: Profiling protein-coding and long noncoding RNA expression in T-acute lymphoblastic leukemia for clinical insights. PNAS Nexus. 2024;3:pgae011.

34. Qi H, Chi L, Wang X, Jin X, Wang W and Lan J. Identification of a Seven-lncRNA-mRNA Signature for Recurrence and Prognostic Prediction in Relapsed Acute Lymphoblastic Leukemia Based on WGCNA and LASSO Analyses. Anal Cell Pathol (Amst). 2021;2021:6692022.
35. Pan L, Liu G, Lin F, Zhong S, Xia H, Sun X, et al. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. Sci Rep. 2017;7:7402.
36. Hu Y, Zhou Y, Zhang M, Ge W, Li Y, Yang L, et al. CRISPR/Cas9-Engineered Universal CD19/CD22 Dual-Targeted CAR-T Cell Therapy for Relapsed/Refractory B-cell Acute Lymphoblastic Leukemia. Clin Cancer Res. 2021;27:2764-2772.
37. Kantarjian HM, DeAngelo DJ, Stelljes M, Martinelli G, Liedtke M, Stock W, et al. Inotuzumab Ozogamicin versus Standard Therapy for Acute Lymphoblastic Leukemia. N Engl J Med. 2016;375:740-53.
38. Egan K, Kusao I, Troelstrup D, Agsalda M and Shiramizu B. Mitochondrial DNA in residual leukemia cells in cerebrospinal fluid in children with acute lymphoblastic leukemia. J Clin Med Res. 2010;2:225-9.
39. Liu Q, Yang L, Shi Z, Yu J, Si H, Jin Y, et al. Development and validation of a preliminary clinical support system for measuring the probability of incident 2-year (pre)frailty among community-dwelling older adults: A prospective cohort study. Int J Med Inform. 2023;177:105138.
40. Bhagat SK, Tiyasha T, Awadh SM, Tung TM, Jawad AH and Yaseen ZM. Prediction of sediment heavy metal at the Australian Bays using newly developed hybrid artificial intelligence models. Environ Pollut. 2021;268:115663.
41. Tarwidi D, Pudjaprasetya SR, Adytia D and Apri M. An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach. MethodsX. 2023;10:102119.
42. Gale RP. Progress in Transplants for Acute Lymphoblastic Leukemia. Clin Cancer Res. 2022;28:813-815.
43. Beyermann B, Agthe AG, Adams HP, Seeger K, Linderkamp C, Goetze G, et al. Clinical features and outcome of children with first marrow relapse of acute lymphoblastic leukemia expressing BCR-ABL fusion transcripts. BFM Relapse Study Group. Blood. 1996;87:1532-8.

Figures



Clinicopathological features		NR	BMR	CR	BMCR	Chisq.test p value: Likelihood ratio	Chisq.test p value: Pearson	Correlation: Contingency coeff	Correlation: Cramer's V
Gender	Female	271	85	9	11	0.012	0.011	0.109	0.110
	Male	437	82	21	11				
Race	White	532	116	24	19	0.108	0.067	0.156	0.091
	Black or African American	50	20	2	0				
	Asian	30	6	1	0				
	American Indian or Alaska Native	0	1	0	0				
	Native Hawaiian or other Pacific Islander	4	0	0	0				
Vital status	Alive	708	65	17	5	<0.001	<0.001	0.596	0.742
	Dead	0	102	13	17				
CNS status at diagnosis	CNS I	557	140	15	16	0.007	<0.001	0.154	0.110
	CNS II	105	24	11	4				
	CNS III	25	3	4	1				
ETV6 RUNX1 fusion status	Negative	412	131	23	12	<0.001	<0.001	0.150	0.152
	Positive	131	17	1	6				
TRISOMY 4 10 status	Negative	493	141	23	20	<0.001	0.002	0.136	0.137
	Positive	100	11	1	0				
MLL status	Negative	560	141	29	19	0.386	0.657	0.046	0.046
	Positive	18	5	0	0				
TCF3 PBX1 status	Negative	474	104	19	14	<0.001	<0.001	0.186	0.189
	Positive	12	15	1	1				
Down syndrome	Negative	701	157	26	19	0.535	0.469	0.053	0.053
	Positive	7	3	1	0				

Figure 1

Clinical Characteristics of ALL Relapse Outcomes

(A) Sample distribution of ALL patients across relapse outcomes in the TARGET multidimensional data. (B) Association significance and strength between clinical features and ALL relapse outcomes. (C) OS (top) and EFS (bottom) curves for ALL patients across relapse groups.

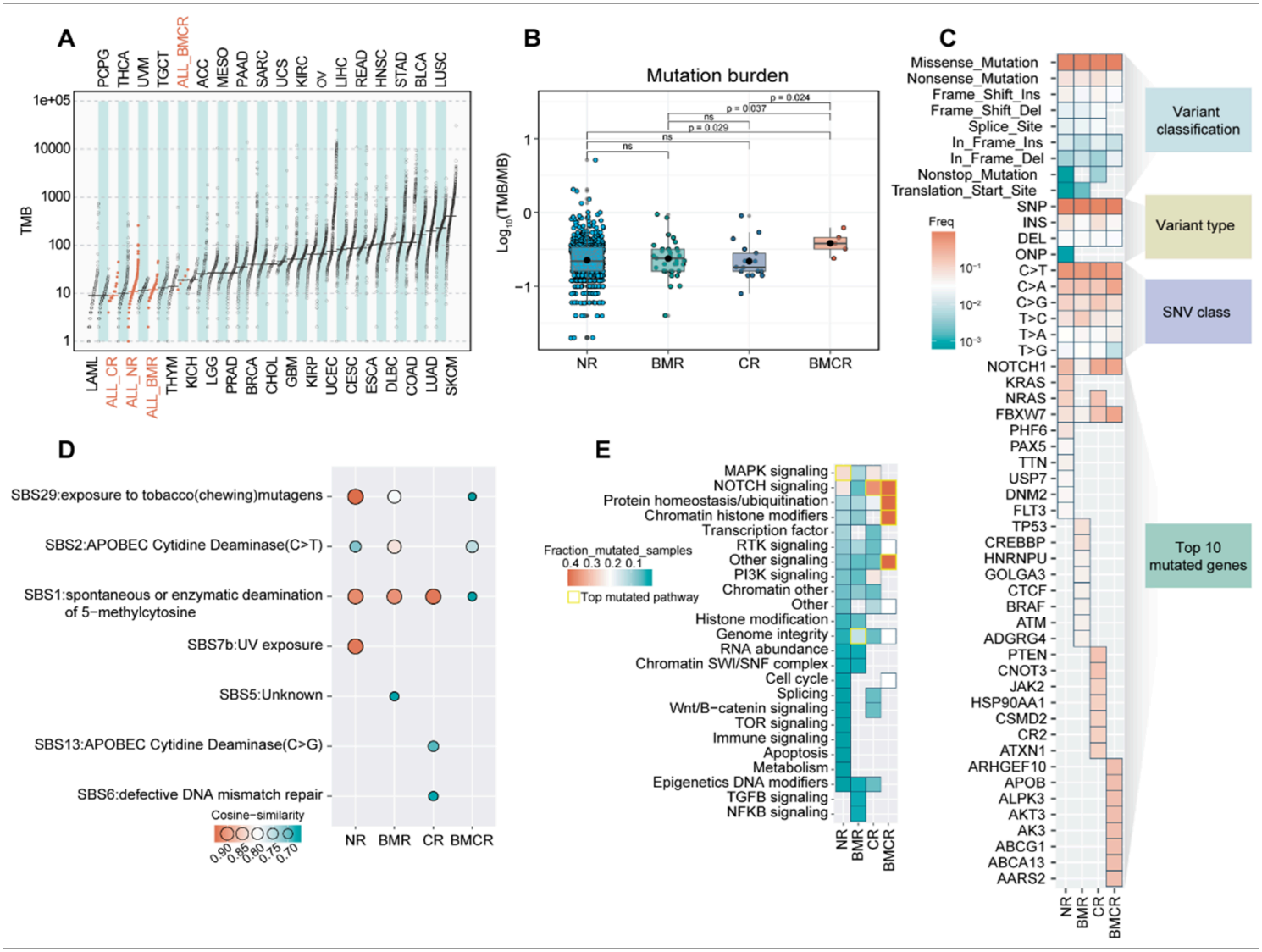


Figure 2

Exome-Wide Mutation Analysis in ALL Relapse Outcomes

(A) Comparative mutation burden profiles of ALL relapse groups and pan-cancer cohorts. (B) Mutation burden analysis of ALL samples across relapse outcomes. (C) Mutation type distribution and top 10 mutated genes in ALL relapse groups. (D) Cosine similarity analysis of mutational signatures in each relapse group. (E) Oncogenic pathway alteration profiles (mutation frequency>0) across relapse groups.

TMB: Tumor mutation burden; SKCM: Skin Cutaneous Melanoma; LAML: Acute Myeloid Leukemia; ONP: Oligonucleotide Polymorphism

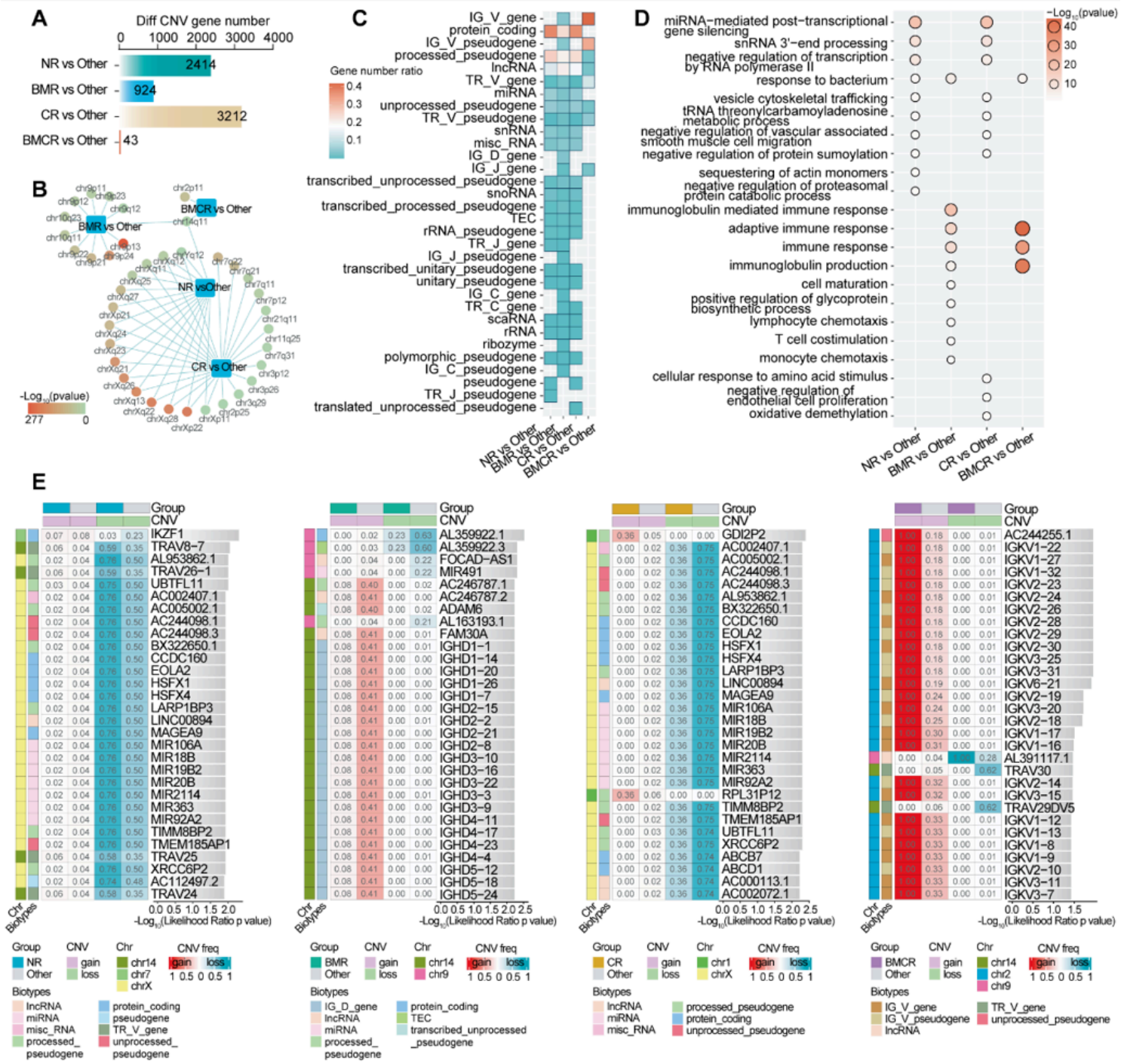


Figure 3

CNV Analysis in ALL Relapse Outcomes

(A) Bar plot displaying the number of differential CNV genes in ALL relapse outcome groups. (B) Network plot illustrating chromosomal enrichment results of differential CNV genes per group. (C) Heatmap showing biotype classification of differential CNV genes in each group. (D) Bubble plot of top 10 GO enrichment terms for differential CNV genes in each group, ranked by p-value. (E) Heatmap depicting gain/loss frequencies of top 30 differential CNV genes (ranked by p-value) across groups.

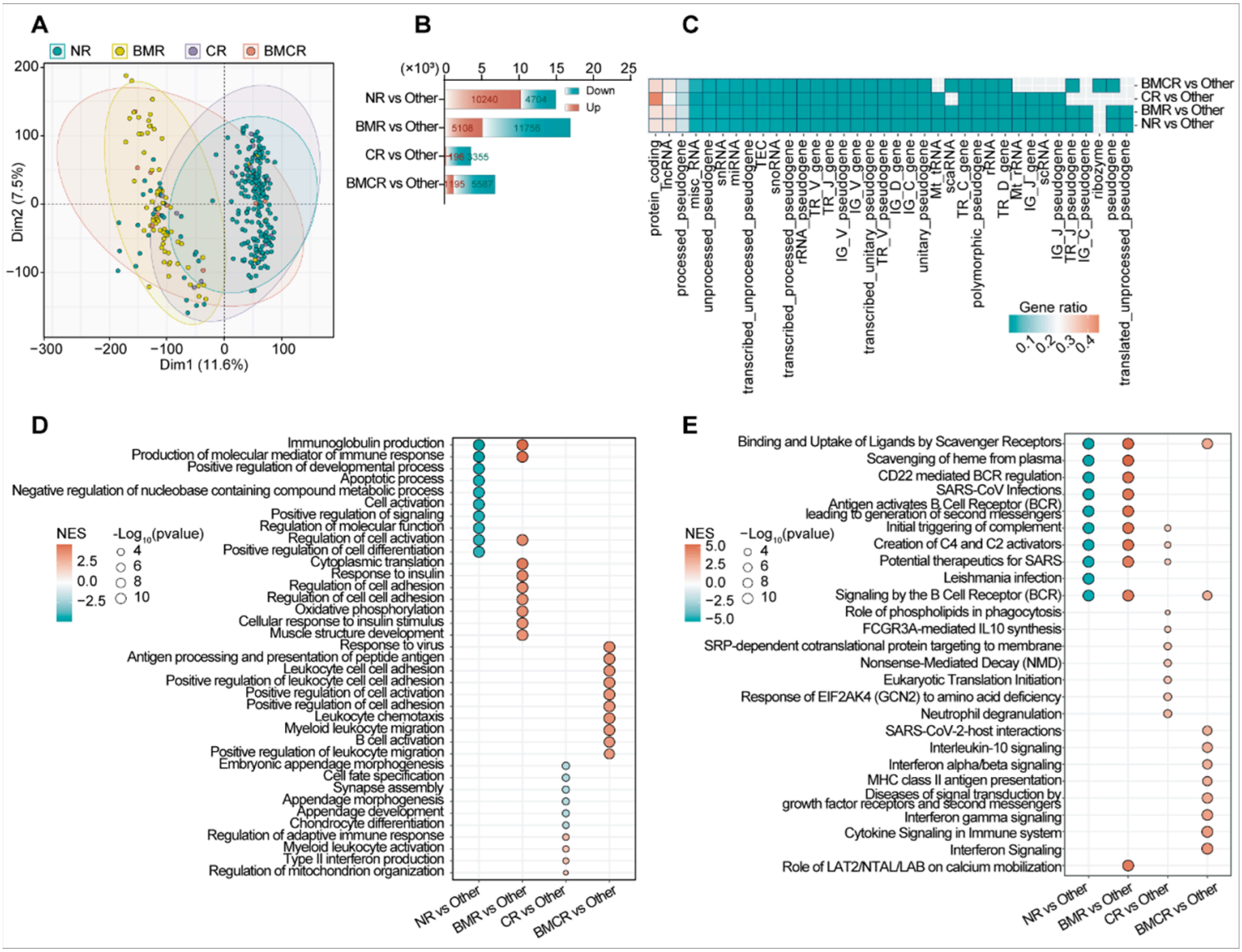


Figure 4

Transcriptomic Characterization of ALL Relapse Outcomes

(A) PCA plot displaying global expression pattern differences among ALL relapse groups at the transcriptomic level. (B) Bar plot showing the number of DEGs in each group. (C) Heatmap illustrating biotype distribution of DEGs across groups. Bubble plots of top 10 pathways from GO (D) and Reactome (E) GSEA analyses, ranked by absolute NES scores.

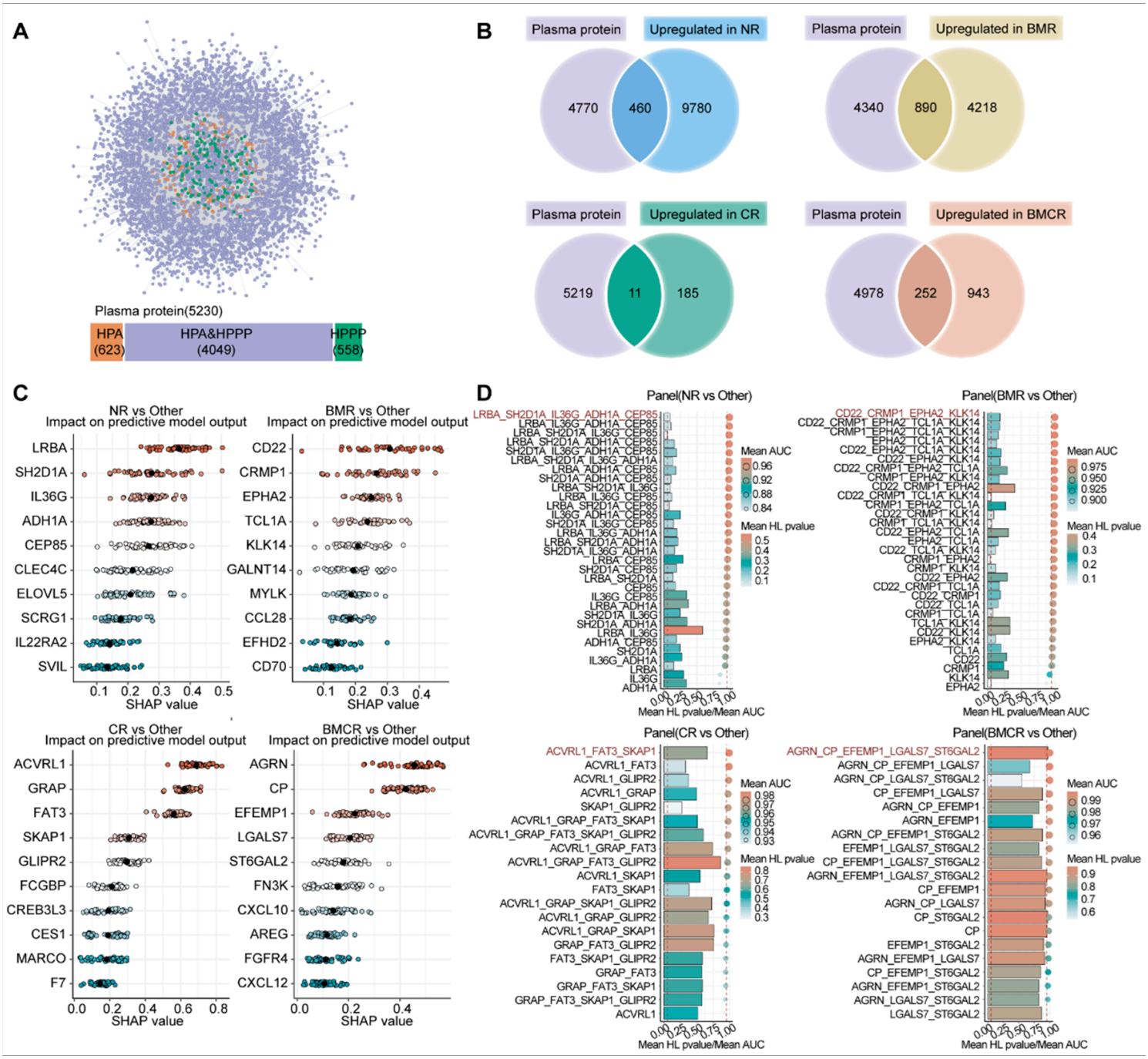


Figure 5

Optimized XGBoost Combined with SHAP Algorithm Identifies Optimal Monitoring Panels for ALL Relapse Outcomes

(A) Network plot showing union of plasma protein-coding genes from HPA and HPPP databases. (B) Venn diagram illustrating intersections between plasma protein-coding genes and group-specific upregulated genes. (C) Bubble plot displaying final SHAP values of top 10 candidate genes per group after pre-training. (D) Bar plot showing mean AUC values and HL p-values for candidate panels in each group.

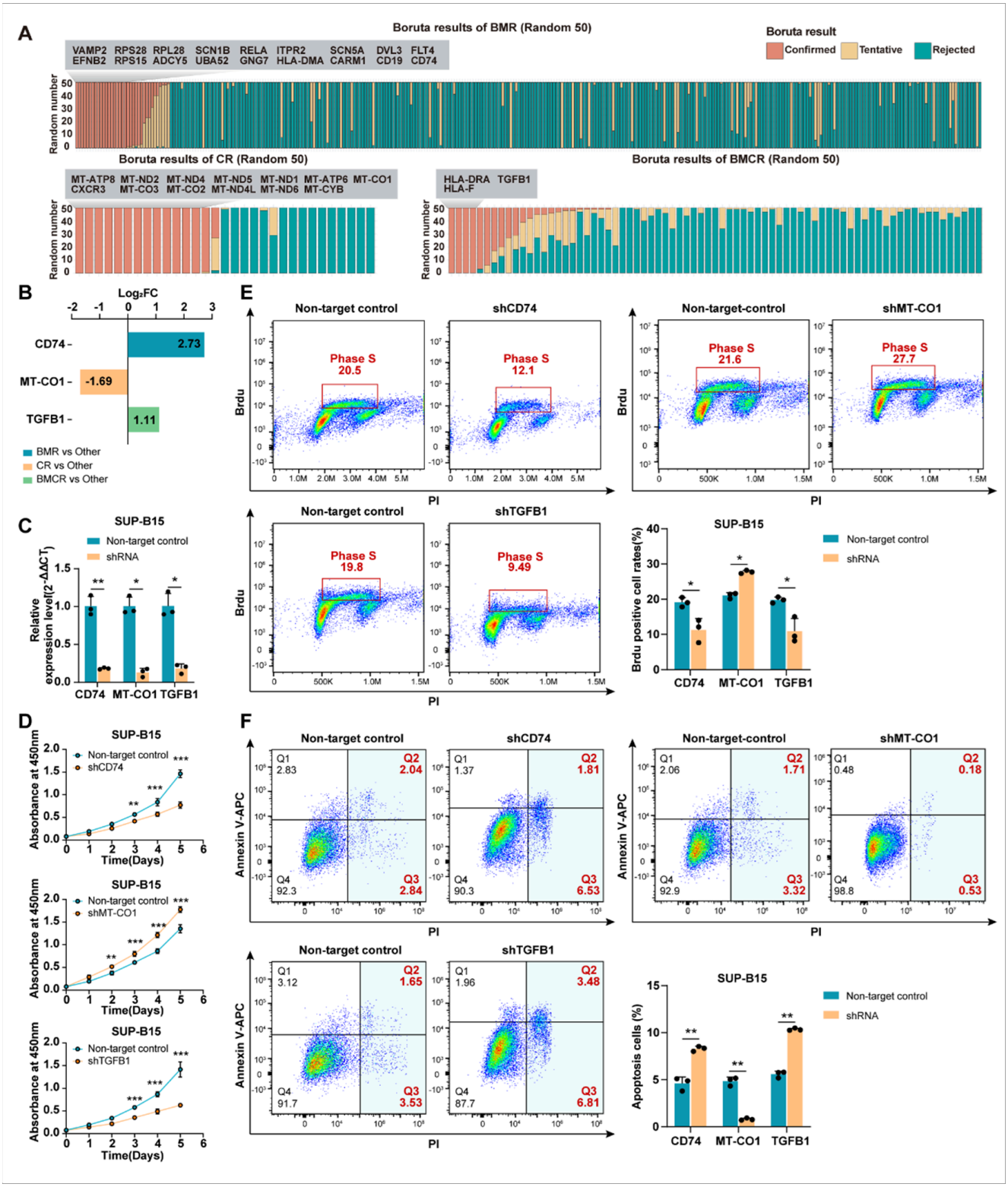


Figure 6

Identification and Validation of Hub Genes in ALL Relapse Outcomes

(A) Bar plot showing hub gene screening results using the Boruta algorithm in each group. Group-specific genes with confirmed importance scores of 50 were designated as hub genes. (B) Bar plot displaying differential expression fold changes of representative hub genes (CD74, MT-CO1, TGFB1)

across ALL relapse outcomes. (C) QPCR validation of stable CD74, MT-CO1, and TGFB1 knockdown efficiency in SUP-B15 cells. Functional validation through CCK8 proliferation assays (D), BrdU incorporation assays (E), and apoptosis assays (F) post-knockdown. All experiments were performed in triplicate with paired t-test statistics. Data are presented as mean ± SD. *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfiguresandfigurelegends.docx
- Graphicalabstract.png