ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL ARTICLE

Genomic and transcriptomic evaluation of primary and recurrent luminal B-like breast cancer

L. Li[1,2,3,4], G. Yin[1,2], X. Liu[1], J. Zhang[1,2], C. Zhou[2,5], P. Song[6], L. Zhao[2,7], W. Zhou[1,2,3], S. Huang[1,2,3,4]* & Z. Yu[1,2,3,4]*

[1]Breast Center, The Second Qilu Hospital of Shandong University, Jinan; [2]Shandong Key Laboratory of Cancer Digital Medicine, Jinan; [3]Shandong Provincial Engineering Laboratory of Translational Research on Prevention and Treatment of Breast Disease, Jinan; [4]Institute of Translational Medicine of Breast Disease Prevention and Treatment, Shandong University, Jinan; [5]Department of Pathology, The Second Qilu Hospital of Shandong University, Jinan; [6]Department of Thyroid Surgery, Binzhou Medical University Hospital, Binzhou; [7]Institute of Medical Sciences, The Second Qilu Hospital of Shandong University, Jinan, China

Available online xxx

Background: Luminal B-like breast cancer (LBBC) constitutes approximately 40% of breast cancer cases and carries a higher risk of recurrence and poorer prognosis than luminal A-like disease. This study aimed to investigate the molecular alterations between primary and recurrent LBBC tumors, elucidate the drivers of LBBC recurrence and identify potential therapeutic targets.

Materials and methods: We performed DNA sequencing using a custom-designed panel covering 769 cancer-related genes and RNA sequencing (RNA-seq) on 15 LBBC patients (49 samples: 15 primary tumors, 21 recurrent lesions and 15 blood samples; six patients had two recurrent lesions each). Of these, 13 had paired-end DNA sequencing data and five had paired-end RNA-seq data (three of whom also had DNA data). Molecular features altered in recurrent tumors were compared with their corresponding primary tumors, focusing on gene mutations, copy number alterations, tumor heterogeneity, gene expression, pathways, and immune cell composition. Molecular features associated with LBBC recurrence were further identified.

Results: The majority of gene mutations and copy number alterations are primarily established in the primary tumor and persist through recurrence. *TP53* and *PIK3CA* were the most frequently altered genes in both primary and recurrent tumors, whereas mutations in *EGFR* and *GNAS* emerged in recurrent tumors. Upregulated pathways in recurrences included cell cycle checkpoints, DNA replication, and antigen processing—cross-presentation. Compared with primary tumors, recurrent tumors exhibited an increased abundance of T cells in the tumor microenvironment. Furthermore, higher mutant-allele tumor heterogeneity (MATH) scores in primary tumors were associated with worse recurrence-free survival in exploratory unadjusted analysis ($P = 0.032$).

Conclusions: Our study reveals the genomic and transcriptomic evolution of LBBC recurrence, suggesting that most drivers are established in the primary tumor. Acquired mutations or alterations warrant further investigation as potential therapeutic targets and MATH scores may represent a candidate biomarker associated with LBBC recurrence.

Key words: genomics, immune phenotype, luminal B-like breast cancer, mutation, recurrence, transcriptomics, tumor heterogeneity

INTRODUCTION

Breast cancer is the most common malignancy globally and a leading cause of cancer-related deaths in women.[1] Despite advancements in therapeutic strategies and clinical screening that enable early detection and curative resection, survival rates for advanced breast cancer patients remain poor due to high recurrence rates. Recurrence often signals a poor prognosis, with hormone receptor (HR)-positive breast cancer being the most common subtype.[2,3] Luminal breast cancer, defined as HR-positive, can be categorized into surrogate luminal A-like and luminal B-like based on human epidermal growth factor receptor 2 (HER2) and Ki67 expression using immunochemistry (IHC) analysis, approximating the transcriptomically defined luminal A and luminal B intrinsic subtypes. The luminal A-like subtype is characterized by HER2 negativity and Ki67 <14%, while all others are classified as luminal B-like.

*Correspondence to: Dr Zhigang Yu, Breast Center, The Second Qilu Hospital of Shandong University, No. 247 Beiyuan Street, Tianqiao District, Jinan 250033, China. Tel: +86531-85875048

E-mail: yuzhigang@sdu.edu.cn (Z. Yu).

Dr Shuya Huang, Breast Center, The Second Qilu Hospital of Shandong University, No. 247 Beiyuan Street, Tianqiao District, Jinan 250033, China. Tel: +86531-85875048

E-mail: huangsya@sdu.edu.cn (S. Huang).

Luminal B-like breast cancer (LBBC) accounts for 40% of all breast cancer subtypes and is characterized by large tumor size, higher rates of lymph node involvement, low histological differentiation, and relative insensitivity to endocrine therapy.[4] These features contribute to its higher recurrence risk and worse prognosis compared with luminal A-like breast cancer. Although additional treatments are often administered to LBBC patients to reduce postoperative recurrence risk, genetic biomarkers that can reliably predict high recurrence risk in LBBC patients are lacking.

Studies in preclinical models and patient cohorts have shown significant molecular and cellular changes during tumor recurrence, including discordant HR and HER2 status between primary and metastatic tumors.[5-7] High-throughput sequencing has revealed that most oncogenic driver mutations in primary tumors are retained in recurrence, though certain genes, such as *ESR1*, *MYLK*, *PEAK1*, *SLC2A4RG*, and *EVC2*, are preferentially mutated in metastatic breast cancer.[8] However, findings vary across different subtypes. Research on triple-negative breast cancer (TNBC) has demonstrated mutational and transcriptomic shifts, as well as decreased immune-activating gene expression in recurrent tumors.[9] However, genomic and transcriptomic comparisons between primary and recurrent LBBC tumors remain limited.

To address this gap, the primary objective of the present study was to perform a comprehensive, descriptive comparison of the genomic and transcriptomic profiles of matched primary and recurrent tumors in LBBC patients using a multiomics characterization strategy. Furthermore, molecular features related to tumor recurrence were further analyzed by integrating omics and clinical characteristics using recurrence-free survival (RFS) as a standard oncologic endpoint, which may help elucidate potential drivers of recurrence and therapeutic resistance in LBBC patients.

Table 1. Demographic and clinical characteristics of the enrolled LBBC patients

Characteristics	*n* (%)
Age at diagnosis, median (range), years	43 (26-69)
Menopausal status	
Premenopausal	13 (86.7)
Postmenopausal	2 (13.3)
AJCC stage	
I	5 (33.3)
II	3 (20.0)
III	7 (46.7)
Histological grade	
Grade 1	1 (6.7)
Grade 2	10 (66.7)
Grade 3	4 (26.7)
Neoadjuvant chemotherapy	
Yes	5 (33.3)
No	10 (66.7)
Chemotherapy	
Yes	13 (86.7)
No	2 (13.3)
HER2 targeted therapy	
Yes	4 (26.7)
No	11 (73.3)
Radiation therapy	
Yes	9 (60.0)
No	6 (40.0)
Antiestrogen therapy	
Yes	10 (66.7)
No	5 (33.3)
RFS, median (range), months	30 (12-62)
<36 months	10 (66.7)
36-60 months	4 (26.7)
>60 months	1 (6.7)

AJCC, American Joint Committee on Cancer; HER2, human epidermal growth factor 2; LBBC, luminal B-like breast cancer; RFS, recurrence-free survival.

MATERIALS AND METHODS

Patients and sample collection

All procedures were conducted in accordance with the Declaration of Helsinki, and approval from the Institutional Review Board of the Second Qilu Hospital of Shandong University was obtained after the study. The inclusion criteria were as follows: (i) LBBC patients with matched formalin-fixed paraffin-embedded (FFPE) tissues from primary breast cancers and local recurrences or distant metastases, as well as blood samples for genomic DNA extraction; (ii) biospecimens containing sufficient tumor tissue for sequencing; and (iii) continuous treatment after recurrence. All primary tumor samples were treatment-naïve, comprising pretreatment biopsy specimens from five patients who subsequently received neoadjuvant chemotherapy and surgical specimens from 10 patients who underwent upfront surgery. All biopsy specimens underwent central pathology review to ensure they were derived from a viable, representative tumor area (as assessed by accompanying hematoxylin and eosin imaging), minimizing the inclusion of nondiagnostic or necrotic samples.

This study included a cohort of 15 patients with histologically confirmed LBBC based on IHC results. From these 15 patients, a total of 36 FFPE tumor samples were obtained: 15 primary tumors (one per patient) and 21 matched recurrent/metastatic tumors. This higher number reflects that six patients contributed two recurrent samples each (e.g. samples from different recurrent sites or at different times of recurrence). In addition, peripheral blood samples were collected from all 15 patients to filter out germline mutations. Demographic data and clinical characteristics of patients are summarized in Table 1. Estrogen receptor (ER), progesterone receptor (PR), HER2 and Ki67 status were evaluated using IHC or *in situ* hybridization. The pathological features of primary and recurrent tumors are summarized and compared in Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2026.107736. For each case, 10-15 (depending on tumor size) 5-μm FFPE sections adjacent to the hematoxylin and eosin-analyzed section were pooled for DNA or RNA extraction. For genomic analysis, targeted sequencing using a custom-designed panel covering 769 cancer-related genes was performed on 13 primary and 18 recurrent tumors, with five patients contributing two recurrent lesions each. RNA sequencing (RNA-seq) was performed on five primary and seven

recurrent tumors from five patients (with two patients each contributing two recurrent lesions). Among these, three patients had both DNA and RNA sequencing data available, while the remaining two patients were included exclusively for transcriptomic analysis.

DNA extraction and targeted sequencing

DNA was extracted from FFPE tumor specimens using the MagPure DNA Micro Kit (Magen Biotech, Guangdong, China) and from peripheral blood using the TGuide S32 magnetic blood genomic DNA kit (TIANGEN, Beijing, China). Fragmented DNA libraries were constructed with the KAPA HTP library preparation kit (KAPA Biosystems, Wilmington, MA) and quantified using the Equalbit 1 × dsDNA HS Assay Kit (Nuovezan, Nanjing, China). Library size was assessed on the Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA). DNA libraries were captured with a predesigned panel covering 769 cancer-associated genes (Genecast, Wuxi City, China) and sequenced on the Illumina NovaSeq 6000 in paired-end 150 bp mode.

Identification of somatic variants (SNVs)

After quality control, reads were aligned to the human reference genome (hg19, NCBI Build 37.5) using Burrows-Wheeler Aligner (version 0.7.17; http://bio-bwa.sourceforge.io).[10] Duplicates were marked with Picard toolkit (version 2.1.0; https://broadinstitute.github.io/picard),[11] and realignment was performed with the Genome Analysis ToolKit (version 3.7; https://gatk.broadinstitute.org).[12] SNV mutations were called with VarDict (version 1.5.1; https://github.com/AstraZeneca-NGS/VarDict),[13] and compound heterozygous mutations were identified using FreeBayes (version 1.2.0; https://github.com/ekg/freebayes).[14] Mutations were annotated with ANNOVAR (http://www.openbioinformatics.org/annovar/).[15] Germline mutations were filtered out, and somatic mutations were identified based on the following criteria: (i) frequency ≥1%; (ii) not located in intergenic or intronic regions and nonsynonymous SNVs; (iii) support reads ≥2 and (iv) allele frequency ≤0.2% in the Exome Aggregation Consortium (ExAC)[16] and Genome Aggregation Database (gnomAD; https://gnomad.broadinstitute.org).

Copy number analysis

Copy number variation (CNV) mutations were pair-called using the ctCNV software (v1.0.7)[17] based on DNA from FFPE tumor tissue samples and genomic DNA from peripheral blood. Copy numbers greater than 4 were defined as CNV gains and less than 1 as CNV losses.

Tumor mutation burden and MATH analysis

Tumor mutation burden (TMB) was determined by quantifying somatic nonsynonymous SNVs in the whole exome (depth >40×, allele frequency ≥0.05) detected by next-generation sequencing, excluding known oncogenic driver alterations. TMB was measured in mutations per Mb. The mutant-allele tumor heterogeneity (MATH) score was calculated using the formula 100 × median absolute deviation (MAD)/median of the variant allele frequency (VAF) for variants with a VAF between 0.05 and 1.

Copy number instability (CNI) score calculation

Read depth was transformed into $\log_2$ ratios, corrected for GC content and target length, and converted into Z-scores based on Gaussian transformations versus a normal baseline ($n = 494$). The CNI score was calculated by summing the Z-scores of target regions (Z-score >95th percentile) and twice the absolute standard deviation of the normal baseline.

RNA-seq and data quality control

For FFPE samples, 5-8 sections were used depending on tissue size and thickness, with RNA extracted using the RNeasy FFPE Kit (#73504; QIAGEN, Hilden, Germany). After rRNA removal, RNA concentration was measured using the Qubit® RNA HS Assay Kit (#Q32855; Thermo Fisher Scientific, Waltham, MA) and Qubit 3.0 fluorometer, and RNA integrity was assessed with Agilent 2100 bioanalyzer (Agilent). cDNA libraries were constructed from 50 ng of RNA using the SMARTer Stranded Total RNA-Seq Kit v2 (#634412; Takara Bio, Shiga, Japan). Library quality was assessed using the 1× dsDNA HS Assay Kit and Qubit 3.0 fluorometer, and size profiles were checked using the Agilent 2100 bioanalyzer. After PCR enrichment and purification of adapter-ligated fragments, the libraries were paired-end sequenced on the Illumina NovaSeq 6000 Sequencing System using PE150 mode.

Gene expression analyses

Raw reads were preprocessed using fastp (version 0.23.1; https://github.com/OpenGene/fastp) to remove low-quality sequences,[18] followed by rRNA removal using Bowtie2 (version 2.4.4; http://bowtie-bio.sourceforge.net/bowtie2).[19] Cutadapt (version 4.4; https://cutadapt.readthedocs.io)[20] was used to trim the last three bases from each read, and Trimmomatic (version 0.36; http://www.usadellab.org/cms/?page=trimmomatic)[21] removed adapter sequences, poly-N sequences, and low-quality reads. The trimmed reads were then aligned to the human hg19 reference transcriptome using STAR (version 2.7.9a; https://github.com/alexdobin/STAR),[22] and BAM files were sorted and indexed with SAMtools (version 1.14; http://www.htslib.org).[23] Gene expression levels were estimated using FeatureCounts (version 2.0.3; https://subread.sourceforge.net)[24] and quantified using the Fragments Per Kilobase Million Mapped Reads and Transcripts Per Million (TPM) methods. Differentially expressed genes (DEGs) were identified using the DESeq2 package (version 1.38.3; https://bioconductor.org/packages/DESeq2),[25] with thresholds set at $|\log_2(\text{fold change})| \geq 1$. Due to the limited sample size for RNA-seq, no genes reached statistical significance after Benjamini–Hochberg false discovery rate (FDR) correction. Therefore, genes with a nominal *P* value

< 0.05 were considered for DEG identification and downstream pathway analysis.

Pathway analysis. The Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway Database enrichment analysis of DEGs was performed using the ClusterProfiler package.[26] KEGG pathways with nominal P value < 0.05 were considered for exploratory analysis, as no pathways reached statistical significance after FDR correction. Gene Set Enrichment Analysis (GSEA, version 4.3.3; https://www.gsea-msigdb.org/gsea)[27] was used for pathway enrichment analysis among different groups. Enriched pathways were considered statistically significant at FDR <0.05.

Immune subtype identification. STAR-counts data and clinical information for tumors were downloaded from the TCGA database (https://portal.gdc.cancer.gov). Data were normalized using $\log_2$(TPM+1) transformation, and samples with both RNA-seq data and clinical information were selected for analysis. Single-sample GSEA (ssGSEA)[28] with the GSVA function and CIBERSORT algorithm[29] were used to estimate the abundance of immune cell subsets in the tumor microenvironment. ssGSEA inferred the abundance of 28 immune cell subtypes based on expression data, while CIBERSORT characterized 22 immune cell fractions through linear support vector regression analysis.

Statistical analysis

Categorical variables were compared using chi-square or Fisher's exact tests. RFS, defined as the time from initial diagnosis to recurrence, was analyzed as a secondary, exploratory endpoint using Kaplan—Meier methods and compared by log-rank test. Patients were divided into high and low groups based on the mean values of TMB, CNI, and MATH of the primary tumor. Each biomarker was assessed via an independent two-group comparison testing a distinct biological hypothesis. Given the exploratory nature of these analyses, no correction for multiple testing was applied, consistent with recommendations for hypothesis-generating studies.[30] Statistical significance was defined as $P < 0.05$ for each comparison. Analyses were performed using SPSS 23.0 (IBM Corporation, Armonk, NY) and GraphPad Prism 8.0 (GraphPad, San Diego, CA).

RESULTS

Clinical characteristics of LBBC patients

Fifteen LBBC patients with paired primary and recurrent or metastatic tumor tissues were enrolled to dissect clonal evolution and identify drivers of recurrence. Patient characteristics, including histopathology, stage, recurrence pattern, systemic therapy and RFS, are summarized in Table 1. The median age at diagnosis was 43 years (range 26-69). Five patients received neoadjuvant chemotherapy. The median RFS, measured from initial diagnosis to recurrence, was 30 months (range 12-62). After initial diagnosis, the majority of patients (13 of 15) received a combination of anthracycline- and/or taxane-based chemotherapy. Endocrine therapy, predominantly tamoxifen or an aromatase inhibitor, was administered to patients with ER-positive disease ($n = 10$). A subset of nine patients received adjuvant radiotherapy. HER2-targeted therapy including trastuzumab was performed on HER2-positive patients ($n = 4$).

Fifteen primary and 21 recurrent tumors were analyzed. Most recurrences were locoregional (lymph node, $n = 5$; chest wall, $n = 8$; sternum, $n = 2$; ipsilateral breast, $n = 4$); two were distant (liver, $n = 1$; bone, $n = 1$). All primary tumors were ER-positive, 86.7% were PR-positive, and 26.7% were HER2-positive. Among recurrent tumors, 65.0% retained ER/PR positivity and 20.0% were HER2-positive. ER/PR status was discordant in 26.7% (4/15) of pairs (primary-positive, recurrent-negative; Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2026.107736).

Genomic landscape of LBBC from primary to recurrent disease

Targeted sequencing of 769 cancer-related genes was performed on 13 primary and 18 matched recurrent tumors. Mutational landscapes are compared in Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2026.107736, and depicted in Figure 1A. Driver genes were largely shared between primary and recurrent lesions. *TP53* was mutated in 46.2% (6 of 13) of primary tumors and 44.4% (8 of 18) of recurrent tumors and *PIK3CA* in 38.5% (5 of 13) and 38.9% (7 of 18), respectively (Figure 1B). *SPEN* mutations were enriched in primary tumors (23.1% versus 11.1%), whereas *KMT2C*, *TSC1*, and *WHSC1L1* alterations were more frequent in recurrences (16.7% versus 7.7%). *EGFR* and *GNAS* mutations were detected exclusively in recurrent tumors (3 of 18 versus 0 of 13), although the difference did not reach statistical significance. Hot-spot analysis revealed *PIK3CA* p.H1047R (3 of 13) and p.E545K (3 of 18) as the most common single-nucleotide variants in Chinese LBBC (Supplementary Table S3, available at https://doi.org/10.1016/j.esmoop.2026.107736). The evolution of somatic mutations in matched primary and recurrent tumors was shown in Figure 1C. Comparing matched primary and recurrent pairs, an overall concordance rate of 83.3% was observed for *PIK3CA* mutation status and 100% for *TP53* mutations. The concordance of *SPEN*, *KMT2C*, *ARIDA1*, *MAP3KA*, *TSC1* and *WHSC1L1* mutations was only identified in one case.

Copy number alterations were highly concordant between primary and recurrent tumors. *MCL1* and *MYC* amplifications were the most prevalent events in both cohorts. *ABCB1*, *IRF4* and *NKX2-1* deletions (2 of 13) were restricted to primaries, whereas *FGF3*/*FGF4* deletions and *ICOSLG*, *U2AF1*, *ARID5B* and *PIK3R3* amplifications (2 of 18) emerged only in recurrences (Figure 2).

The association between tumor genetic heterogeneity and clinical outcome

Tumor genetic heterogeneity, including TMB, CNI and MATH, was further calculated according to DNA sequencing data in primary and recurrent tumors. The results showed that TMB, CNI and MATH values did not differ significantly between primary and recurrent tumors (Figure 3A).

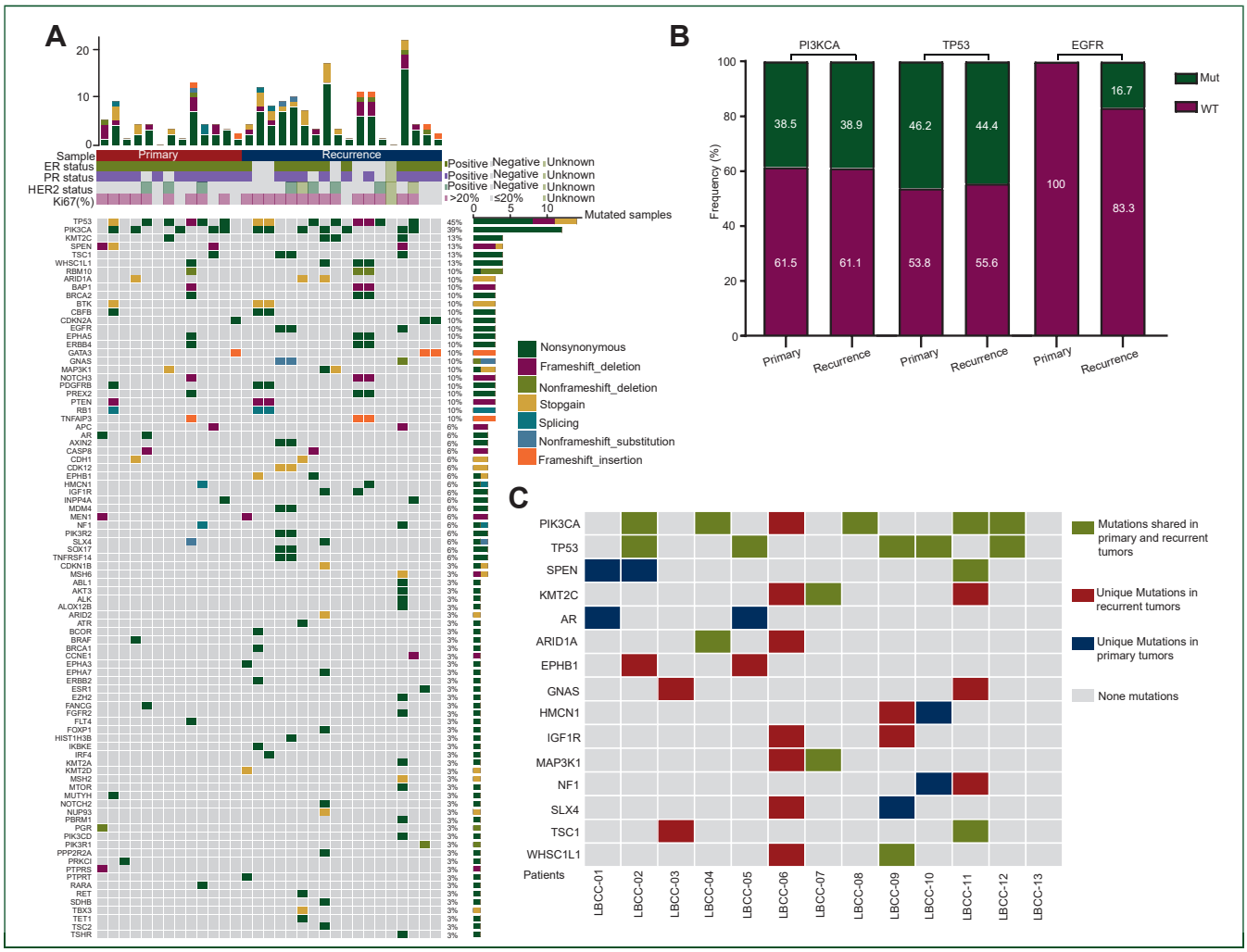

Figure 1. Genomic landscapes of paired primary and recurrent tumors in LBBC. (A) Oncoprint summarizing somatic mutations detected by targeted sequencing of 769 cancer-related genes in 13 treatment-naïve primary tumors and 18 matched recurrent lesions. Samples are annotated (top bars) for clinicopathological variables and receptor status. Mutation types are color-coded; right-side histograms indicate the total number of alterations per gene and top histograms show the mutational load per sample. (B) Bar charts comparing the mutation frequencies of TP53, PIK3CA, and EGFR between primary ($n = 13$) and recurrent ($n = 18$) tumors. (C) The somatic mutation profiles of matched primary and recurrent tumors from each patient were analyzed to identify mutations shared between the two timepoints, as well as mutations unique to either the primary or the recurrent tumor.
LBBC, luminal B-like breast cancer.

We further explored the effects of TMB, CNI and MATH in primary tumors on clinical outcomes. In exploratory unadjusted Kaplan—Meier analyses, a high MATH score in the primary tumor was associated with shorter RFS of LBBC patients ($P = 0.032$), whereas TMB and CNI showed no significant association with RFS (Figure 3B).

Transcriptomic and immune landscape of LBBC from primary to recurrent disease

RNA-seq was successfully performed on five primary and seven recurrent tumors of LBBC patients. A total of 781 DEGs were identified between primary and recurrent tumor tissues, including 134 upregulated genes and 647 downregulated genes [$|\log_2(\text{fold change})| > 2$, $P < 0.05$; Figure 4A and Supplementary Table S4, available at https://doi.org/10.1016/j.esmoop.2026.107736]. To characterize transcriptional alterations associated with recurrence, we performed both KEGG pathway enrichment analysis and GSEA. KEGG enrichment highlighted oxidative phosphorylation as the most significantly altered pathway ($P < 0.05$; Figure 4B and C and Supplementary Table S5, available at https://doi.org/10.1016/j.esmoop.2026.107736). However, because KEGG enrichment does not capture coordinated pathway-level directionality, we performed GSEA to assess whether pathways exhibit coherent upregulation or downregulation. GSEA revealed significant upregulation of cell cycle checkpoints, DNA replication, and antigen processing—cross-presentation pathways in recurrent tumors (normalized enrichment score >1, FDR <0.01; Figure 4D and Supplementary Table S6, available at https://doi.org/10.1016/j.esmoop.2026.107736). Of note, oxidative phosphorylation did not reach significance in GSEA, suggesting that although individual genes within this pathway are differentially expressed, they do not exhibit a coordinated directional shift at the pathway level.

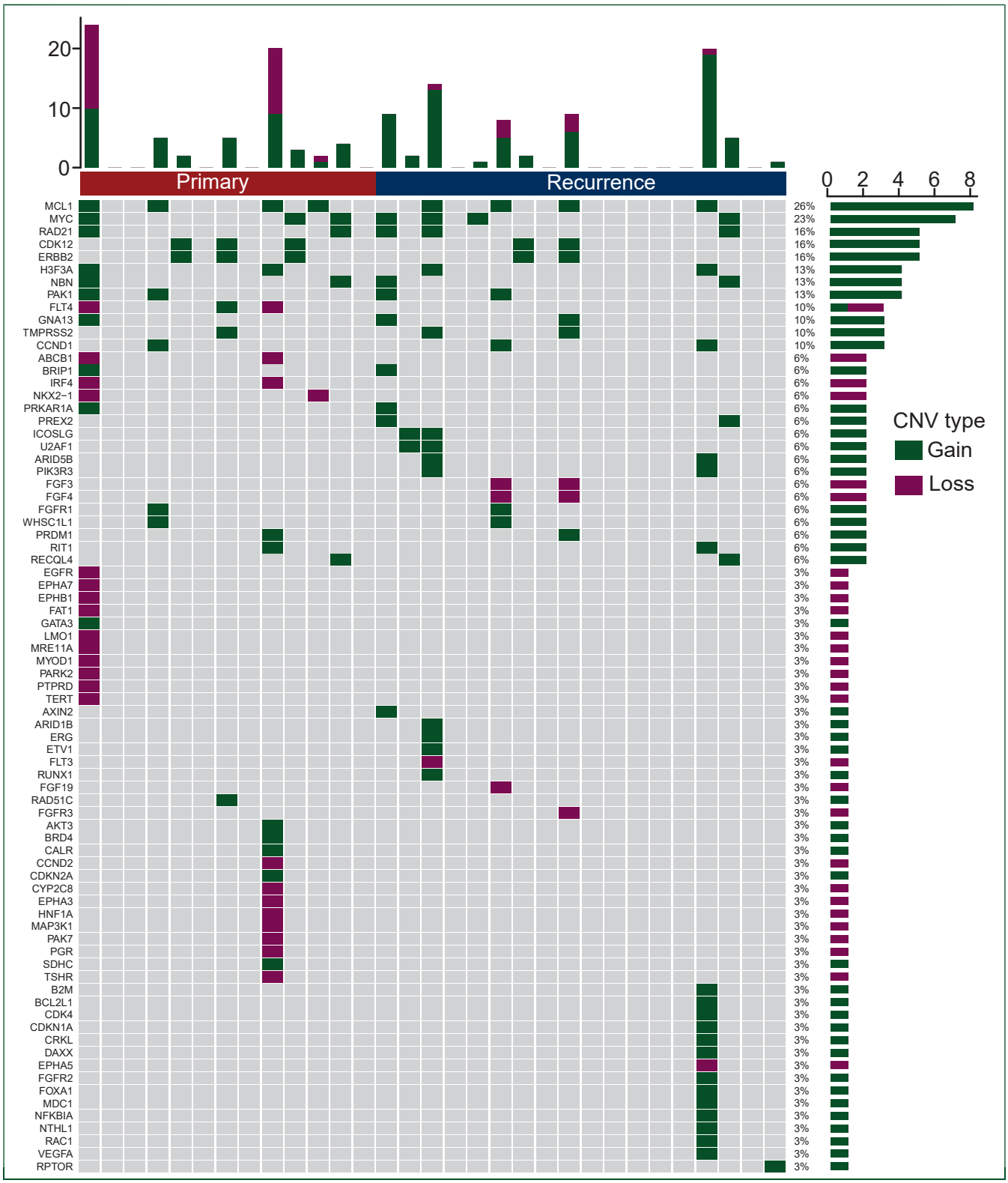

Figure 2. CNA profiles across primary and recurrent tumors. Heatmap depicts CNAs in 13 primary and 18 recurrent tumors (columns); red indicates copy number gain, blue indicates loss. Histograms on the right summarize the cumulative CNA burden among the tumors.
CNA, copy number alteration.

To further study the relationship between the immune microenvironment and the progression of LBBC, we compared the immune cell composition among normal tissues, primary tumors and recurrent tumors. CIBERSORT and ssGSEA were used to profile the immune microenvironment. Compared with normal tissue, LBBC tumors exhibited marked immune remodeling, including reduced $CD8^+$ and $CD4^+$ T-cell infiltrates (Supplementary Figure S1A, available

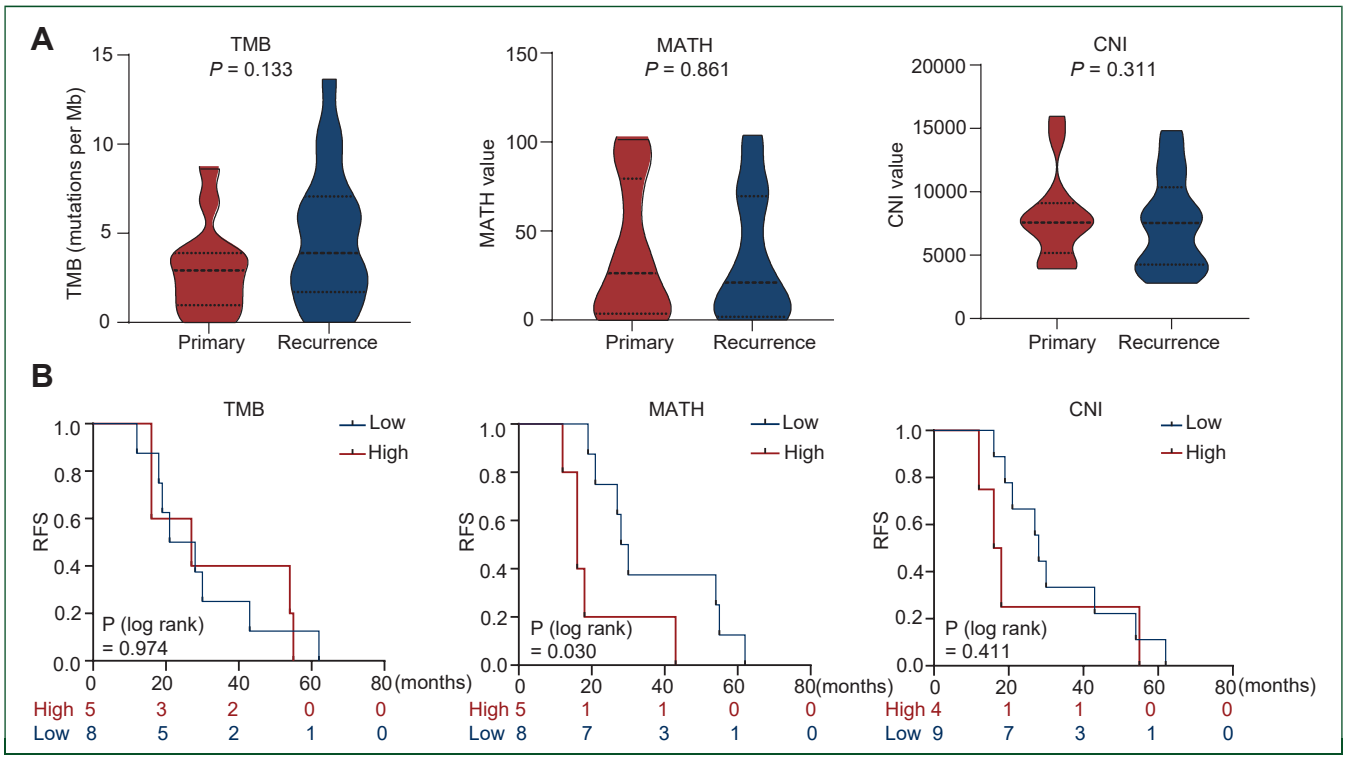

Figure 3. Tumor heterogeneity and clinical outcome. (A) Comparison of TMB, MATH, and CNI between primary and recurrent tumors of LBBC patients. *P* values are calculated based on the Wilcoxon signed-rank test. (B) Kaplan—Meier curves of the RFS rates stratified by high versus low TMB, MATH, and CNI. The *P* value represents the log-rank test assessing the difference in survival between the high and low groups in the unadjusted analysis.
CNI, copy number instability; LBBC, luminal B-like breast cancer; MATH, mutant-allele tumor heterogeneity; RFS, recurrence-free survival; TMB, tumor mutation burden.

at https://doi.org/10.1016/j.esmoop.2026.107736). However, transition from primary to recurrent disease was accompanied by only an increase in total T-cell abundance ($P < 0.05$; Figure 4E and Supplementary Figure S1B, available at https://doi.org/10.1016/j.esmoop.2026.107736).

DISCUSSION

Previous studies identified the genomic and gene expression profiles in luminal breast cancer and the changes in ER-positive tumors were also detected before and after endocrine therapy. However, combined genomic and transcriptomic studies in paired primary and recurrent tumors of LBBC patients remain limited. In this study, we provide several key advancements to the genomic and transcriptomic understanding of LBBC recurrence. First, we present a paired multiomics analysis that delineates the parallel analysis of the genome and transcriptome in matched primary and recurrent tumor samples—an approach that differs from that of prior studies, which largely focused on single-omic or unpaired designs. Second, we identified that recurrent tumors displayed markedly divergent mutational landscapes, acquired additional oncogenic alterations, and frequently exhibited discordant ER/PR/HER2 status. Furthermore, our data suggest that intratumoral genetic heterogeneity may serve as an early predictor of recurrence risk in LBBC.

The prevalence of alterations in the three most mutated genes in breast cancer is *TP53* (30%), *PIK3CA* (20%-40%), and *KMT2C* (6%-9%). Mutations in *PIK3CA* are frequent in ER-positive tumors, close to a frequency of 40%, and LBBC has a lower frequency of *PIK3CA* mutations compared with luminal A.[31] We identified that *TP53* and *PIK3CA* variants dominated the mutational spectrum of both primary and recurrent LBBC tumors, which aligns with prior large-scale analyses. In TCGA, these two genes each accounted for 29% of LBBC cases,[32] while a Taiwanese cohort of 305 HER2-negative LBBC patients reported *PIK3CA* (42%), *FGFR1* (25%), and *BRCA1/2* (10.5%) mutations as the most common alterations.[33] The third most frequently mutated gene in our primary tumors was *SPEN* (SMRT/HDAC1-associated repressor protein, SHARP), with a mutation frequency of 23.3%. *SPEN* encodes a 402-kDa nuclear scaffold containing four N-terminal RNA-recognition motifs and a conserved C-terminal SPOC domain. In previous studies, 13% of mutations in *SPEN* were identified in ER-positive/PR-negative breast cancer.[34] Somatic *SPEN* inactivation—via 23% loss of heterozygosity and 3%-4% point mutations—has been documented in diffuse large B-cell lymphoma, splenic marginal-zone lymphoma and pancreatic adenocarcinoma.[35-37] Mechanistically, SPEN orchestrates transcriptional repression within Notch, TCF/LEF, and EGFR signaling cascades during embryogenesis and adult tissue homeostasis.[38,39] As an estrogen-inducible corepressor, SPEN attenuates ERα-dependent transcription; consequently, its loss promotes endocrine resistance.[40] Given its role as a transcriptional corepressor and its relatively high mutation frequency in our cohort, *SPEN* may represent a candidate driver warranting further investigation in LBBC recurrence.

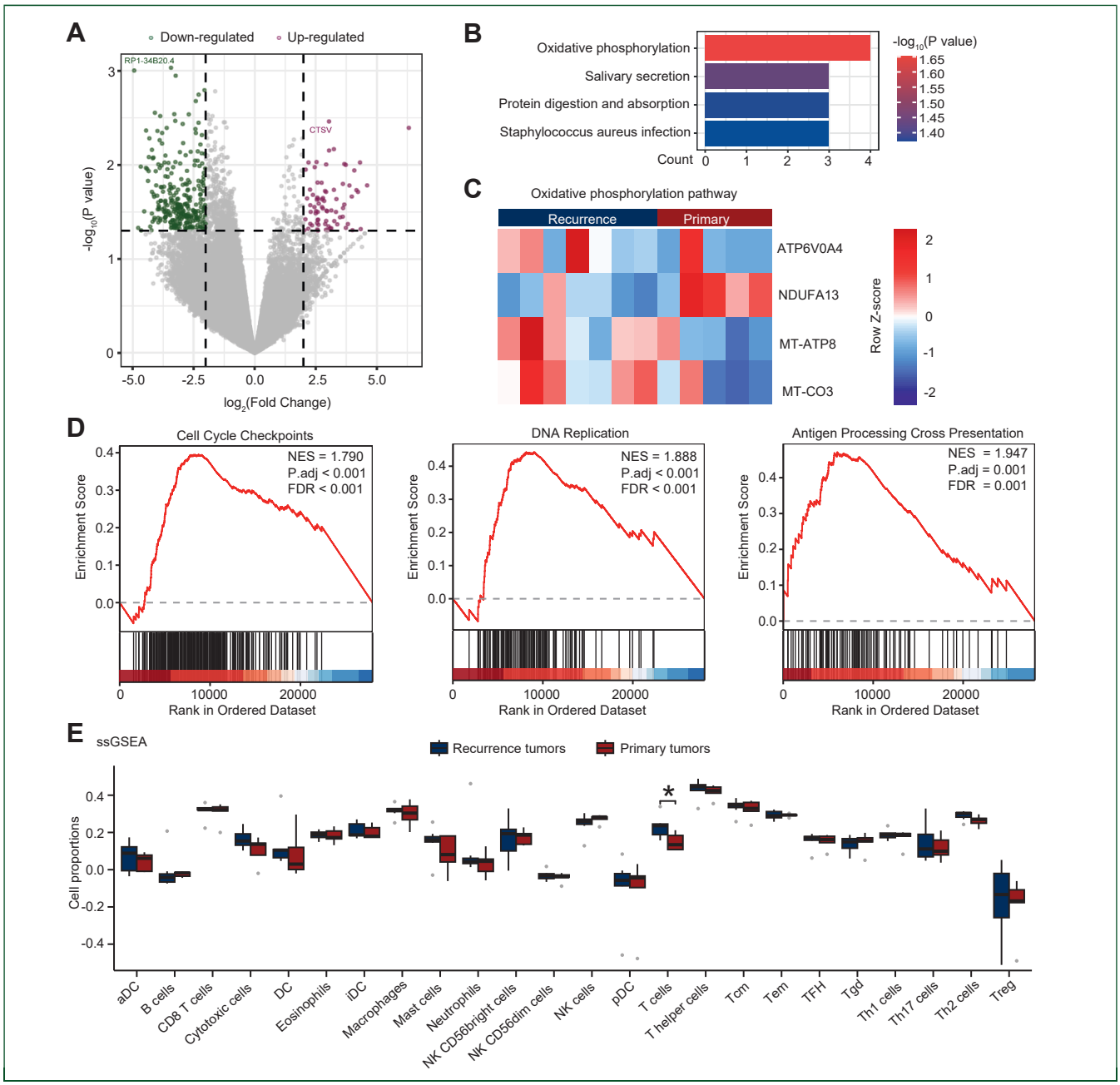

Figure 4. Transcriptomic reprogramming and immune microenvironment evolution from primary to recurrent disease. (A) Volcano plot of DEGs between five primary and seven recurrent tumors. Significant DEGs are shown as red (upregulated) and blue (downregulated) dots ($|\log_2(\text{fold change})| > 2$, $P < 0.05$). (B) Kyoto Encyclopedia of Genes and Genomes pathway enrichment of DEGs ($P < 0.05$). (C) Heatmap of core genes in the oxidative phosphorylation pathway. (D) Gene Set Enrichment Analysis plots showing significant upregulation of cell cycle checkpoints, DNA replication, and antigen processing/cross-presentation pathways in recurrent tumors (FDR <0.05). (E) Single-sample Gene Set Enrichment Analysis-derived immune cell infiltration scores comparing primary ($n = 5$) and recurrent ($n = 7$) tumors; *$P < 0.05$.

DEGs, differentially expressed genes; FDR, false discovery rate.

The exclusive identification of *EGFR* and *GNAS* mutations in recurrent tumors in our cohort, albeit at low frequency, points to a potential mechanism of acquired resistance and a therapeutic vulnerability. EGFR, a member of the ErbB family, contains an extracellular ligand-binding domain, a transmembrane segment, and an intracellular tyrosine kinase region that activates the Ras/Raf/MEK/ERK, PI3K/AKT, STAT, and PKC pathways to drive breast cancer progression. Previous studies showed that *EGFR* mutations are largely confined to TNBC, with reported frequencies of 2.6%-11.4%.[41,42] Tyrosine-kinase inhibitors (TKIs), which target activating *EGFR* mutations, have transformed the management of non-small cell lung cancer, yet have shown limited efficacy in unselected breast cancers. Our observation of recurrent-restricted *EGFR* mutations generates a hypothesis that these patients could be considered for eligibility trials with EGFR TKIs or antibody—drug conjugates, although this strategy requires further investigation in LBBC.

GNAS encodes the stimulatory G-protein α-subunit (Gαs) that couples seven-transmembrane receptors to adenylyl

cyclase, driving cAMP production. Activating mutations lock Gαs in a GTP-bound state, resulting in ligand-independent cAMP elevation and chronic activation of downstream PKA and MAPK/ERK cascades. In breast cancer, GNAS overexpression accelerates proliferation and metastasis through the PI3K/AKT/Snail1/E-cadherin axis.[43] Pan-cancer analyses reveal that oncogenic *GNAS* variants are enriched in mucinous histotypes and occur in >1% of gastrointestinal and gynecological malignancies, yet are present in only 0.2% of breast tumors.[44] Importantly, *GNAS*-mutated lesions display heightened peritoneal dissemination, diminished responsiveness to first-line systemic therapy, and inferior survival[44]; moreover, they are markedly enriched in patients with detectable minimal residual disease, a key determinant of early recurrence and poor prognosis.[45] We observed *GNAS* mutations exclusively in recurrent LBBC tumors, suggesting that selective GNAS inhibitors or combination regimens targeting cAMP/PKA and PI3K/AKT signaling may represent promising therapeutic strategies warranting further exploration in preclinical models and clinical trials. Future studies validating these alterations in independent cohorts and *in vitro* models are warranted to test their functional role and therapeutic relevance.

MATH quantifies intratumoral genomic heterogeneity as the percentage ratio of the scaled MAD to the median VAF. Previous studies showed that elevated MATH scores correlate with adverse outcomes in multiple cancer types, including bladder, head and neck and endometrial carcinomas.[46] In breast cancer, high MATH is associated with shorter overall survival following adjuvant chemotherapy and reduced infiltration of 24 immune cell subsets, suggesting that MATH may serve as both a prognostic biomarker and a predictor of immunotherapy response.[47] Gastric cancer data further link low MATH and low immune checkpoint inhibitor (ICI) scores to superior outcomes.[48] In our exploratory analysis, we observed an association between high MATH in primary LBBC and shorter RFS (log-rank $P = 0.032$). Given the exploratory nature of this finding and the fact that no correction for multiple testing was applied, this result should be interpreted as hypothesis-generating. Together with the biological rationale linking MATH to immune evasion, MATH-high tumors may represent a subgroup of interest for prospective evaluation of immunotherapy in larger, confirmatory cohorts.

Our RNA-seq profiling revealed prominent upregulation of cell cycle checkpoint and DNA replication pathways in recurrent versus primary tumors. Dysregulated CDK4/6 activity at the G_1/S transition unleashes unchecked proliferation and amplifies oncogenic signaling. Over the past three decades, selective CDK4/6 inhibitors have demonstrated robust antitumor activity in preclinical models and randomized trials, leading to their approval in combination with endocrine therapy for HR-positive/HER2-negative breast cancer.[49] Recurrent LBBC tumors displaying transcriptional activation of cell cycle pathways may therefore represent a subgroup that could benefit from the addition of CDK4/6 inhibitors, a hypothesis that warrants further clinical investigation. Additionally, the upregulation of the antigen processing—cross-presentation pathway in recurrent tumors suggests an enhanced capacity for antigen presentation. This finding, together with our observation of increased T-cell infiltration in recurrent tumors (Figure 4E), points to a potentially more immunologically active tumor microenvironment at recurrence. Whether this immune-active phenotype could render recurrent LBBC more susceptible to ICIs remains an open question that warrants further investigation.

Notably, although KEGG analysis identified oxidative phosphorylation as the most significantly enriched pathway among DEGs, GSEA did not show coordinated directional changes. This pattern may reflect heterogeneous metabolic adaptation rather than a uniform shift in mitochondrial metabolism, potentially representing context-dependent metabolic flexibility that supports tumor survival under therapeutic pressure.

Limitations

Our study had several limitations. First, the modest sample size, particularly for the paired RNA-seq analysis ($n = 5$), restricts the statistical power for subgroup analyses and limits the generalizability of our findings. Although our study provides valuable hypothesis-generating insights into the evolutionary trajectories of LBBC, the results, especially from the exploratory genomic and transcriptomic analyses, will require validation in larger, prospective cohorts. Second, patient selection was driven by tissue availability, potentially introducing selection bias. For primary tumors, analyses were performed on both biopsies and surgical specimens; thus, the allelic frequency of mutations or the presence of rare subclonal events might be underestimated in biopsy samples due to tumor heterogeneity. However, for patients considered inoperable at diagnosis, a biopsy is the only available tissue source. Combining biopsy and surgical specimens allowed us to maximize the sample size and reflect a real-world clinical scenario. Moreover, bulk-tumor analyses—rather than laser-capture microdissection—may have diluted tumor-specific signals with stromal and immune content, despite stringent quality control measures.

Conclusions

In summary, our study delineates the divergent genomic and transcriptomic trajectories that accompany LBBC recurrence, identifies candidate alterations in *SPEN*, *EGFR*, *GNAS* and cell cycle regulators that warrant further investigation as potential therapeutic targets, and generates hypotheses for integrating immunotherapy, CDK4/6 inhibitors, and EGFR-directed agents into precision strategies aimed at preventing recurrence and improving outcomes in LBBC patients. The association between MATH score and RFS is an exploratory finding that requires validation in independent cohorts.

ACKNOWLEDGEMENTS

We thank all patients and their families involved in the study for their participation.

DATA AVAILABILITY

The DNA panel and RNA-seq data of LBBC generated and analyzed during this study are in the process of being deposited to the National Genomics Data Center,[50] the Beijing Institute of Genomics (China National Center for Bioinformation). An accession number will be provided once all the data is completely uploaded. Based on the Regulation of the People's Republic of China on the Administration of Human Genetic Resources, the Human Genetic Resource Data are accessed via an application to the Data Access Committee for research purposes. Potential users need to complete and be approved for a data access request and then the data will be made available upon reasonable request according to the terms of the consent and data use limitations of the subjects.

FUNDING

The work was supported by Shandong Provincial Natural Science Foundation [grant numbers ZR2022MC047, ZR2022QH170], National Natural Science Foundation of China [grant numbers 82573562, 82473170], Key R&D Program of Shandong Province [grant number 2025CXPT144], the Fundamental Research Funds for the Central Universities [grant number 2022JC009], and the Special Foundation for Taishan Scholars (grant number ts201511098), the China Health & Medical Development Foundation (grant number JJ-2025-0118) and the China Postdoctoral Science Foundation (grant number 2021M701985).

DISCLOSURE

The authors have declared no conflicts of interest.

REFERENCES

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229-263.
2. Colleoni M, Sun Z, Price KN, et al. Annual hazard rates of recurrence for breast cancer during 24 years of follow-up: results from the International Breast Cancer Study Group Trials I to V. *J Clin Oncol*. 2016;34(9):927-935.
3. Zhang S, Wang H, Zhang H, et al. Clinicopathological and molecular features of $HR^{+}/HER2^{-}$ breast cancer patients with distinct endocrine resistance patterns. *Chin J Cancer Res*. 2025;37(1):48-65.
4. Bago-Horvath Z, Rudas M, Singer CF, et al. Predictive value of molecular subtypes in premenopausal women with hormone receptor-positive early breast cancer: results from the ABCSG Trial 5. *Clin Cancer Res*. 2020;26(21):5682-5688.
5. Zhu A, Wang N, Yun Z, et al. HER2 status results in an unstable switch from primary to recurrent breast cancer. *Neoplasma*. 2024;71(4):392-401.
6. Qi G, Zhang X, Gai X, Yan X. Retrospective analysis of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor-2 (HER2), Ki67 changes and their clinical significance between primary breast cancer and metastatic tumors. *PeerJ*. 2024;12:e17377.
7. Shiino S, Ball G, Syed BM, et al. Prognostic significance of receptor expression discordance between primary and recurrent breast cancers: a meta-analysis. *Breast Cancer Res Treat*. 2022;191(1):1-14.
8. Paul MR, Pan TC, Pant DK, et al. Genomic landscape of metastatic breast cancer identifies preferentially dysregulated pathways and targets. *J Clin Invest*. 2020;130(8):4252-4265.
9. Hutchinson KE, Yost SE, Chang CW, et al. Comprehensive profiling of poor-risk paired primary and recurrent triple-negative breast cancers reveals immune phenotype shifts. *Clin Cancer Res*. 2020;26(3):657-668.
10. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
11. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079.
12. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498.
13. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res*. 2016;44(11):e108.
14. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing 2012. arXiv preprint arXiv:1207.3907 [q-bio.GN].
15. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
16. Karczewski KJ, Weisburd B, Thomas B, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 2017;45(D1):D840-D845.
17. Yu J, Han T, Lin X, et al. A novel computational tool for copy number variation detection in targeted circulating tumor DNA. *J Clin Oncol*. 2019/05/20;37(suppl 15):e13051.
18. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890.
19. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359.
20. Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J Comput Biol*. 2017;24(11):1138-1143.
21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.
22. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
23. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930.
24. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10(2):giab008.
25. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
26. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287.
27. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.
28. Charoentong P, Finotello F, Angelova M, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep*. 2017;18(1):248-262.
29. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243-259.
30. Menyhárt O, Győrffy B. Multiplicity corrections in life sciences: challenges and consequences. *Int J Epidemiol*. 2025;54(4):dyaf098.
31. Galappaththi SPL, Smith KR, Alsatari ES, et al. The genomic and biologic landscapes of breast cancer and racial differences. *Int J Mol Sci*. 2024;25(23):13165.
32. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70.
33. Chen BF, Tsai YF, Lien PJ, et al. Prevalent landscape of tumor genomic alterations of luminal B1 breast cancers using a comprehensive genomic profiling assay in Taiwan. *Breast Cancer*. 2024;31(2):217-227.
34. Dai D, Wu H, Zhuang H, Chen R, Long C, Chen B. Genetic and clinical landscape of ER+/PR− breast cancer in China. *BMC Cancer*. 2023;23(1):1189.

35. Hartert KT, Wenzl K, Krull JE, et al. Targeting of inflammatory pathways with R2CHOP in high-risk DLBCL. *Leukemia*. 2021;35(2):522-533.
36. Rossi D, Trifonov V, Fangazio M, et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *J Exp Med*. 2012;209(9):1537-1551.
37. Ma H, Song B, Guo S, Li G, Jin G. Identification of germline and somatic mutations in pancreatic adenosquamous carcinoma using whole exome sequencing. *Cancer Biomark*. 2020;27(3):389-397.
38. Doroquez DB, Orr-Weaver TL, Rebay I. Split ends antagonizes the Notch and potentiates the EGFR signaling pathways during *Drosophila* eye development. *Mech Dev.* 2007;124(9-10):792-806.
39. Feng Y, Bommer GT, Zhai Y, et al. *Drosophila* split ends homologue *SHARP* functions as a positive regulator of Wnt/beta-catenin/T-cell factor signaling in neoplastic transformation. *Cancer Res*. 2007;67(2):482-491.
40. Légaré S, Cavallone L, Mamo A, et al. The estrogen receptor cofactor SPEN functions as a tumor suppressor and candidate biomarker of drug responsiveness in hormone-dependent breast cancers. *Cancer Res*. 2015;75(20):4351-4363.
41. Teng YHF, Tan WJ, Thike AA, et al. Mutations in the epidermal growth factor receptor (EGFR) gene in triple negative breast cancer: possible implications for targeted therapy. *Breast Cancer Res*. 2011;13(2):R35.
42. Levva S, Kotoula V, Kostopoulos I, et al. Prognostic evaluation of epidermal growth factor receptor (EGFR) genotype and phenotype parameters in triple-negative breast cancers. *Cancer Genomics Proteomics*. 2017;14(3):181-195.
43. Jin X, Zhu L, Cui Z, Tang J, Xie M, Ren G. Elevated expression of GNAS promotes breast cancer cell proliferation and migration via the PI3K/AKT/Snail1/E-cadherin axis. *Clin Transl Oncol*. 2019;21(9):1207-1219.
44. Johannet P, Abdelfattah S, Wilde C, et al. Molecular and clinicopathologic impact of GNAS variants across solid tumors. *J Clin Oncol*. 2024;42(32):3847-3857.
45. Zhang S, Jiang Y, Zhou L, et al. Genomic characteristics and its therapeutic implications in breast cancer patients with detectable molecular residual disease. *Cancer Res Treat*. 2024;56(2):538-548.
46. Mroz EA, Tward AD, Pickering CR, Myers JN, Ferris RL, Rocco JW. High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. *Cancer.* 2013;119(16):3034-3042.
47. Li Y, Wang M, Yang S, et al. Intratumoral heterogeneity contributes to the chemotherapy prognosis of breast cancer. *J Cancer Res Ther.* 2022;18(5):1268-1275.
48. Wang SY, Wang YX, Guan LS, et al. Construction of a prognostic model for gastric cancer based on immune infiltration and microenvironment, and exploration of *MEF2C* gene function. *BMC Med Genomics*. 2025;18(1):13.
49. Gianni C, Nicolò E, Cristofanilli M. Adjuvant CDK4/6 inhibitors in hormone receptor-positive early breast cancer: one fits all? *Transl Breast Cancer Res*. 2024;5:34.
50. CNCB-NGDC Members and Partners. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2021. *Nucleic Acids Res*. 2021;49(D1):D18-D28.