British Journal of Cancer www.nature.com/bjc

ARTICLE

The heterogeneity of tumour immune microenvironment revealing the CRABP2/CD69 signature discriminates distinct clinical outcomes in breast cancer

Jie Mei[1,2,7], Yun Cai[1,7], Lingyan Chen[1,7], Youqing Wu[3,7], Jiayu Liu[4], Zhiwen Qian[1], Ying Jiang[4], Ping Zhang[5], Tiansong Xia[6]✉, Xiang Pan[3]✉ and Yan Zhang[1,4]✉

BACKGROUND: It has been acknowledged that the tumour immune microenvironment (TIME) plays a critical role in determining therapeutic responses and clinical outcomes in breast cancer (BrCa). Thus, the identification of the TIME features is essential for guiding therapy and prognostic assessment for BrCa.

METHODS: The heterogeneous cellular composition of the TIME in BrCa by single-cell RNA sequencing (scRNA-seq). Two subtype-special genes upregulated in the tumour-rich subtype and the immune-infiltrating subtype were extracted, respectively. The CRABP2/CD69 signature was established based on CRABP2 and CD69 expression, and its predictive values for the clinical outcome and the neoadjuvant chemotherapy (NAT) responses were validated in multiple cohorts. Moreover, the oncogenic role of CRABP2 was explored in BrCa cells.

RESULTS: Based on the heterogeneous cellular composition of the TIME in BrCa, the BrCa samples could be divided into the tumour-rich subtype and the immune-infiltrating subtype, which exhibited distinct prognosis and chemotherapeutic responses. Next, we extracted CRABP2 as the biomarker for the tumour-rich subtype and CD69 as the biomarker for the immune-infiltrating subtype. Based on the CRABP2/CD69 signature, BrCa samples were re-divided into three subtypes, and the $CRABP2^{high}CD69^{low}$ subtype exhibited the worst prognosis and the lowest chemotherapeutic response, while the $CRABP2^{low}CD69^{high}$ subtype showed the opposite results. Furthermore, CARBP2 functioned as a novel oncogene in BrCa, which promoted tumour cell proliferation, migration, and invasion, and CRABP2 inhibition triggered the activation of cytotoxic T lymphocytes (CTLs).

CONCLUSION: The CRABP2/CD69 signature is significantly associated with the TIME features and could effectively predict the clinical outcome. Also, CRABP2 is determined to be a novel oncogene, which could be a therapeutic target in BrCa.

British Journal of Cancer (2023) 129:1645–1657; https://doi.org/10.1038/s41416-023-02432-6

BACKGROUND

Breast cancer (BrCa) is a widespread malignant tumour with the highest morbidity and fatal mortality among all tumour types in the world [1]. Based on the latest data issued by the American Cancer Society, there will be nearly 300,000 estimated new cases and more than 43,000 estimated cancer-related deaths in 2022 [1]. The clinical outcomes of BrCa patients have been constantly optimised due to the rapid progression of emerging and personalised therapeutic strategies, particularly neoadjuvant therapy (NAT), which allows a great deal of inoperable cases to regain the opportunity for surgical resection [2]. Nowadays, NAT has become part of the standard-of-care therapeutic strategy of patients with locally advanced tumours. Given the complexity and heterogeneity of BrCa tumours, it is difficult to precisely predict the therapeutic response to NAT. Gene-expression-based assessment is significant for therapeutic and prognostic prediction [3], but the forecasting ability of a single gene is usually insufficient, and a multiple gene signature is not convenient.

In the past decade, it has been acknowledged that the tumour immune microenvironment (TIME) participates in regulating cancer progression and determining NAT efficacy [4, 5]. Tumours, including BrCa, are composed of the mixtures of tumour cells and microenvironment cells consisting of stromal cells, tumour-infiltrating lymphocytes (TILs), cancer-associated fibroblasts, etc., with the roles of these cells being complicated and unclear [6, 7]. Tumours can be divided into "immuno-cold" or "immuno-hot" depending on the properties of the TIME. "Immuno-cold" tumours are featured as immunosuppressive TIME and resistant to most treatments, such as immunotherapy and cytotoxic chemotherapy, but "immuno-hot" tumours always exhibit well therapeutic efficacy to various therapies, which are characterised by effective T-cell infiltration and the immuno-supportive TIME [8, 9]. It has

[1]Wuxi Maternal and Child Health Hospital, Wuxi Medical Center of Nanjing Medical University, 214023 Wuxi, China. [2]The First Clinical Medical College, Nanjing Medical University, 211166 Nanjing, China. [3]School of Artificial Intelligence and Computer Science, Jiangnan University, 214122 Wuxi, China. [4]Department of Oncology, The Women's Hospital of Jiangnan University, 214023 Wuxi, China. [5]Department of Breast Surgery, The Women's Hospital of Jiangnan University, 214023 Wuxi, China. [6]Jiangsu Breast Disease Center, The First Affiliated Hospital of Nanjing Medical University, 210029 Nanjing, China. [7]These authors contributed equally: Jie Mei, Yun Cai, Lingyan Chen, Youqing Wu.
✉email: xiats@njmu.edu.cn; xiangpan@jiangnan.edu.cn; fuyou2007@126.com

Received: 16 February 2023 Revised: 31 August 2023 Accepted: 6 September 2023
Published online: 15 September 2023

been reported that TILs abundance in TIME is associated with pathologically complete response (pCR) in BrCa patients receiving NAT [10–12]. However, the features of TIME are difficult to be comprehensively described using a minor panel of biomarkers, shrinking the clinical application prospects of TIME.

In this research, we first described the heterogeneous cellular composition of the TIME in BrCa by single-cell RNA sequencing (scRNA-seq), and subtyping BrCa samples into two subtypes, including the tumour-rich and the immune-infiltrating subtypes. Next, we extracted two subtype-special genes, namely CRABP2 and CD69, which was upregulated in the tumour-rich and the immune-infiltrating subtypes, respectively. Based on CRABP2 and CD69 expression, BrCa samples were re-divided into three subtypes, and BrCa patients with CRABP2^{high}CD69^{low} showed the worst prognosis and the lowest chemotherapeutic response. In addition, CRABP2 promoted BrCa progression and cytotoxic T lymphocytes (CTLs) exhaustion in vitro. Overall, we presented a novel tumour subtyping strategy and determined a novel oncogene CRABP2 in BrCa.

METHODS

Study design

The current study was designed to describe the global immune landscape in BrCa and identify novel immune subtypes. Then, the subtype-specific genes were extracted, and the CRABP2/CD69 signature was established. Multiple in-house and public clinical cohorts were involved in this research, which were used for scRNA-seq analysis, subtyping development and clinical measurement (Supplementary Table S1). The functional role of CRABP2 in BrCa was also explored. The overview of the current study design is exhibited in Fig. 1a.

Clinical cohorts

A total of eight clinical cohorts were involved in our current research. The collection of the cohort 1 and the cohort 6 was approved by the institutional review board at Wuxi Maternal and Child Health Hospital, and the collection of the cohort 8 was approved by the Clinical Research Ethics Committee at Outdo Biotech (Shanghai, China). The cohort 2, the cohort 3, the cohort 4, the cohort 5, and the cohort 7 were public, and no ethical approval was needed.

Cohort 1 included ten female BrCa samples, which were collected from Wuxi Maternal and Child Health Hospital and submitted for scRNA-sequencing. To exclude the effect of anti-tumour therapies on the TIME, none of these ten patients received any preoperative anti-tumour therapies. The detailed clinic-pathological features are shown in Supplementary Table S2.

Cohort 2 ($n = 1069$) was the Cancer Genome Atlas (TCGA) BrCa cohort, which was used as a validated cohort for the tumour-rich and the immune-infiltrating subtypes and a discovery cohort for the CRABP2/CD69 signature, as well as to define the expression of CRABP2 in tumour and para-tumour tissues. The normalised RNA-seq data and clinical information were downloaded from UCSC Xena data portal (https://xenabrowser.net/datapages/).

Cohort 3 was the METABRIC cohort ($n = 1903$), used as the validated cohort for the CRABP2/CD69 signature in predicting prognosis. The normalised RNA-seq data and clinical information in the METABRIC cohort were obtained from the cBioPortal data portal (http://www.cbioportal.org/datasets) [13].

Cohort 4 was the GSE34138 cohort ($n = 178$) [14], and cohort 5 was the GSE163882 cohort ($n = 222$) [15], which were used as validated cohorts for the CRABP2/CD69 signature in predicting pCR in neoadjuvant chemotherapy in BrCa. The normalised RNA-seq data and patient information in the two cohorts were collected from the Gene Expression Omnibus (GEO) database.

Cohort 6 included 92 female BrCa patients recruited from Wuxi Maternal and Child Health Hospital, and none of these patients received any preoperative anti-tumour therapies, which was named the WXMCCH cohort. The tumour FFPE samples were used to construct a tumour microarray (TMA), which was then submitted for multiplexed immunohistochemistry (mIHC) analysis to validate the CRABP2/CD69 signature in predicting clinical outcomes. The recurrent risk of each patient was assessed according to the guideline for diagnosis and treatment of BrCa of the Chinese Anti-cancer Association (version 2021). The detailed clinic-pathological features are shown in Supplementary Table S3.

Cohort 7 was the CPTAC BrCa cohort, which was applied to define the protein levels of CRABP2 in tumour and para-tumour tissues. Proteome data of tumour and para-tumour samples was collected from the CPTAC dataset (http://cptac-data-portal.georgetown.edu/).

Cohort 8 contained 45 paired tumour and para-tumour samples, which was purchased from Outdo BioTech (Cat. HBreD090CS01). The TMA was submitted for the immunohistochemistry (IHC) assay to define the protein expression of CRABP2 in tumour and para-tumour tissues.

ScRNA-seq and bioinformatics analysis

ScRNA-seq was submitted for Shanghai Genechem Co., Ltd. Single-cell library sequencing was conducted on the Illumina HiSeq XTen, with 150 nt paired-end sequencing. The Cell Ranger 3.0.2 was applied to perform sample demultiplexing, barcode processing and generating gene count data for each cell. The cDNA inset was aligned to the hg38/GRCh38 reference genome. The feature-barcode matrices were generated for each individual sample by counting the valid barcodes and unique molecular identifier. Further analysis, such as quality control and unsupervised clustering, was performed using the Seurat (4.1.0, http://satijalab.org/seurat/) R toolkit [16].

To avoid the influence of abnormal cells and experimental noise on downstream analysis, we retained the high-quality cells based on the following criteria: (1) the number of detected genes was between 200 and 5000; (2) the percentage of reads mapped to the mitochondrial genome was less than 20%. The cell number, the distribution of gene number, unique molecular identifier (UMI) number and the mitochondrial percentage per cell after filtering were shown in Supplementary Table S4. Finally, a total of 36,634 cells were reserved for downstream analysis. To account for multiple biological and experimental characteristics, the Harmony algorithm [17] was used to integrated the 36,634 cells from BrCa patients. The principal component analysis (PCA) was performed on the top 4000 genes with the highest variability. Subsequently, the first 20 principal components (PCs) were used to reduce the dimensionality of the scaled integrated dataset to two-dimensional space. The cell clusters were recognised by a shared nearest neighbour (SNN) modularity optimisation-based clustering algorithm with a resolution of 1. According to the expression levels of some well-known markers, the 36,634 cells were annotated as ten cell types, including luminal cells, basal cells, fibroblasts, endothelial cells, pericytes, mast cells, macrophages, B cells, plasma cells and T/NK cells. The inferCNV package was applied to evaluate the copy number variants (CNVs) of each cell based on scRNA-seq raw counts. Initial CNVs for each region of cells were estimated by inferCNV by using luminal and basal cells as the test group and the rest cell types as the control group. The CNV level of each cell was calculated as a quadratic sum of CNV for each region.

To further deconstruct the subpopulations of immune cells, the Harmony algorithm was performed to integrate 25,817 immune cells from BrCa patients. Then, these immune cells were divided into ten clusters unsupervised depending on the top 20 PCs and the 0.5 resolution. The "FindAllMarkers" function was used to identify the cluster-specific genes. Based on cluster-specific genes and conventional markers, these immune cells were recognised as eight subpopulations, including macrophages, mast cells, B cells, plasma cells, naive T cells, cytotoxic T cells, exhausted T cells and NK cells.

To explore the clinical relevance of the composition of TIME, we assessed the cell-type abundances of bulk transcriptomic data from BrCa patients by using the BayesPrism algorithm [18]. Single cells in this study were used as the reference, and labelled as different cell types. For recognising the subtypes of patients based on the composition of tumours, consensus clustering [19], NbClust testing, and Silhouette analysis [20, 21], were applied to determine the optimal number of stable subtypes. Then, for subtyping patients at the single-cell or bulk transcriptomic levels, the number supported by most testing methods was chosen as the most appropriate number of clusters at both single-cell and bulk transcriptomic levels (Spearman distance, hierarchical clustering, $n = 2$).

Cell–cell communications mediated by ligand-receptor complexes were critical to diverse biological processes, such as inflammation and tumorigenesis. To investigate the molecular interaction networks between different cell types, we used the "CellPhoneDB" tool [22], a software to infer cell–cell communication from the combined expression of multi-subunit ligand-receptor complexes, to analyse the interactions between tumour cells and microenvironment cell subpopulations. The ligand-receptor pairs

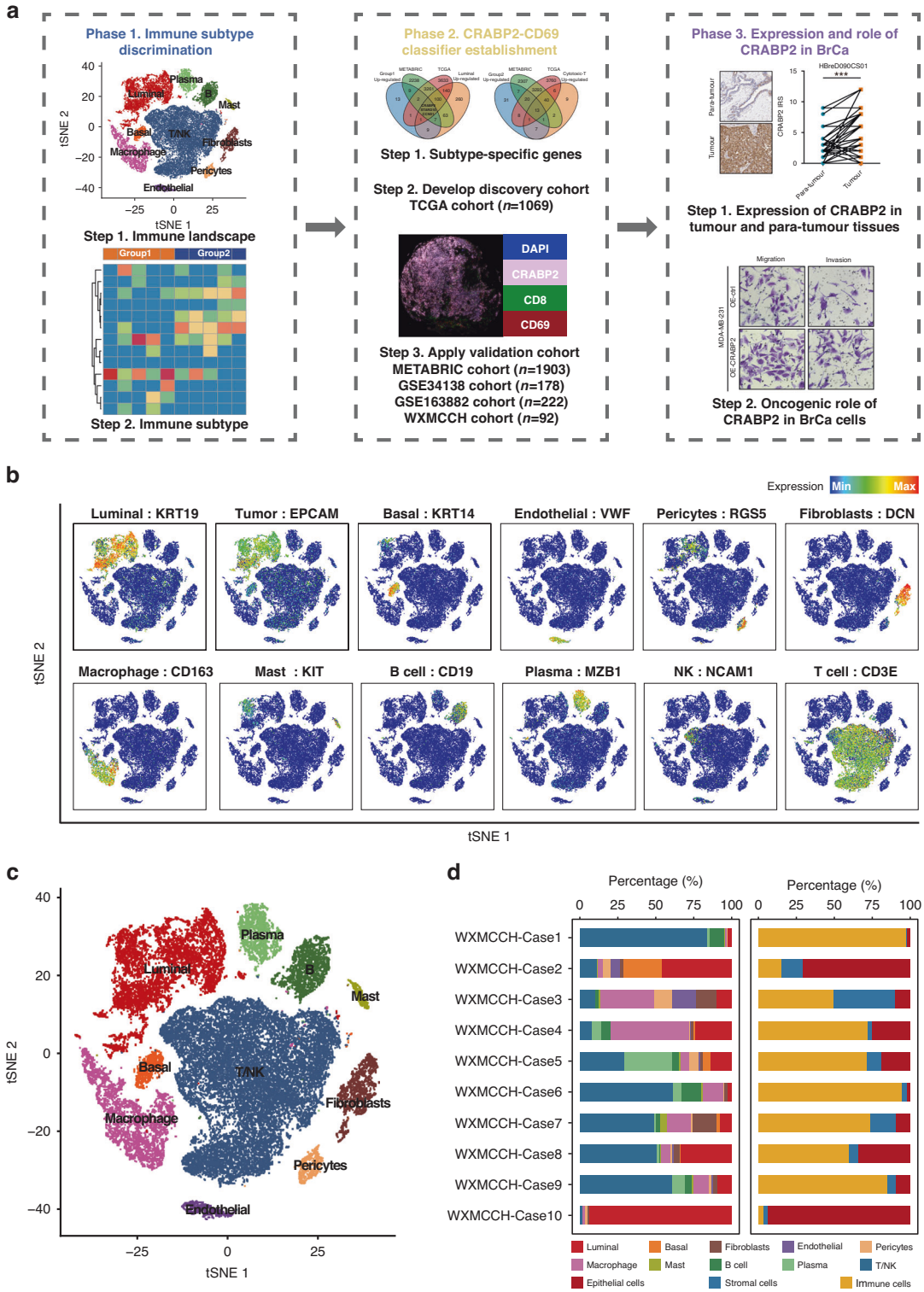

Fig. 1 Integrated scRNA-seq analysis of tumour samples from BrCa patients. a Schematic overview in this study. The cellular heterogeneity of BrCa patients was deconstructed by integrating scRNA-seq data of ten BrCa patients (in-house). Based on the cellular composition, BrCa patients showed two stable subtypes with distinct molecular characteristics. Results derived from scRNA-seq data were further validated by bulk transcriptomic data and mIHC. **b** Expression levels of cell-type-specific genes overlaid on the t-distributed Stochastic Neighbour Embedding (t-SNE) representation. **c** t-SNE visualisation of cell types annotated by classical gene markers. **d** Stacked histogram showing the percentage of minor (left) and major (right) cell types from various cell types in each sample.

with a P value < 0.05 remained for the assessment of relationships among different cell clusters.

Evaluation of the TIME characteristics in the TCGA-BrCa cohort

The immunological characteristics of TIME in the TCGA-BrCa cohort contained immunomodulators, tumour purity, tumour-infiltrating lymphocytes (TILs), and immune checkpoints. The lists of 122 immunomodulators, including MHC, receptors, chemokines, and immuno-stimulating factors, as well as immune checkpoints referred to our previous publication [23]. The ESTIMATE algorithm was conducted to assess Tumour Purity, ESTIMATE Score, Immune Score, and Stromal Score [24]. The associations between the CRABP2/CD69 signature and TIME features were assessed.

Screening the group-specific signatures

For screening the group-specific genes, the "FindAllMarkers" function (avg_log2FC = 0.25 and min.pct = 0.25) was used to measure the difference at the single-cell levels, while the "pROC" package was applied to evaluate the accuracy of gene transcriptional levels in distinguishing the tumour-rich group and the immune-infiltrating phenotypes in the transcriptomic bulk datasets.

For identifying the tumour-rich group-specific signatures, the screening criteria were followed. Firstly, we identified the genes which were significantly upregulated both in luminal cells and tumour-rich phenotypes at the single-cell levels. Fold changes (FC) of gene expression and adjusted P value were measured. In addition, we also calculated the FC of expression percentage of each gene between luminal and other cells (pct_FC1). Meanwhile, the FC of the percentage of expressed cells of each gene in the tumour-rich group of the immune-infiltrating phenotypes were also calculated (pct_FC2). Then, genes with pct_FC1 ≥ 1.2, pct.1 ≥ 0.5 and adjusted P value < 0.05 were recognised as Luminal-upregulated signatures. Genes with pct_FC2 ≥ 1.2, pct.1 ≥ 0.5 and adjusted P value < 0.05 were identified as tumour-rich group-upregulated signatures. Genes with the ROC ≥ 0.6 in the TCGA and METABRIC datasets were identified as subtype-related genes. Subsequently, genes belonging to the Luminal-upregulated signatures, tumour-rich group-upregulated markers and subtype-related genes were recognised as tumour-rich group-specific biomarkers. For identifying the immune-infiltrating group-specific signatures, the screening criteria were followed. Firstly, we identified the genes that were significantly upregulated both in cytotoxic T cells and the immune-infiltrating phenotype at the single-cell levels. FC of gene expression and adjusted P value were measured. In addition, we also calculated the FC of expression percentage of each gene between cytotoxic T cells and other cells (pct_FC1). Meanwhile, the FC of percentage of expressed cells of each gene in the tumour-rich group the immune-infiltrating groups were also calculated (pct_FC2). Then, genes with pct_FC1 ≥ 1.2, pct.1 ≥ 0.5 and adjusted P value < 0.05 were identified as cytotoxic T-upregulated signatures. Genes with pct_FC2 ≥ 1.2, pct.1 ≥ 0.5 and adjusted P value < 0.05 were identified as immune-infiltrating group-upregulated signatures. Genes with the ROC ≥ 0.6 in the TCGA and the METABRIC datasets were identified as subtype-related genes. Subsequently, genes belonging to the cytotoxic T-upregulated signatures, immune-infiltrating group-upregulated markers and subtype-related genes were recognised as immune-infiltrating group-specific biomarkers.

Multiplexed immunohistochemistry

The expression patterns of CRABP2 and CD69, as well as CD8, the marker for CTLs, were detected using the multiplexed immunohistochemistry (mIHC) according to the standard protocol with simultaneous detection of DAPI [25, 26]. The primary antibodies used for mIHC were shown below: anti-CRABP2 (1:5000 dilution, Cat. 10225-1-AP, ProteinTech, Wuhan China), anti-CD69 (1:500 dilution, Cat. ab233396, Abcam, Cambridge, UK), and anti-CD8 (Ready-to-use, Cat. PA067, Abcarta, Suzhou, China). All stained sections were independently assessed by two senior pathologists. CRABP2 expression was evaluated according to the immunoreactivity score criterion [27]. Briefly, the percentage of positively stained cells was scored as 0-4: 0 (<5%), 1 (6–25%), 2 (26–50%), 3 (51–75%) and 4 (> 75%). The staining intensity was scored as 0–3: 0 (negative), 1 (weak), 2 (moderate) and 3 (strong). The immunoreactivity score equals the percentage of positive cells multiplied with staining intensity. In addition, CD69 and CD8 expression were evaluated according to previous research [28]. Taking CRABP2 and CD69 expression together, BrCa samples were divided into four subtypes, namely the $CRABP2^{high}CD69^{high}$ subtype, the $CRABP2^{high}CD69^{low}$ subtype, the $CRABP2^{low}CD69^{high}$ subtype, and the $CRABP2^{low}CD69^{low}$ subtype.

Immunohistochemistry

The HBreD090CS01 TMA was used to perform IHC staining. The primary antibody utilised in the study was anti-CRABP2 (1:5000 dilution, Cat. 10225-1-AP, ProteinTech, Wuhan China). Antibody staining was visualised with DAB and hematoxylin counterstain. Stained TMA was evaluated by two independent senior pathologists according to the immunoreactivity score [27].

Cell culture, transfection and function detection

BrCa cell lines SK-BR-3 (Cat. KG197), MDA-MB-231 (Cat. KG033), MCF-7 (Cat. KG031), BT-549 (Cat. KG413), and MDA-MB-468 (Cat. CX0256) authenticated using short tandem repeat profiling were obtained from KeyGEN (Nanjing, China) and BOSTER (Wuhan, China). SK-BR-3 cells were cultured in McCoy's 5 A media added with 10% foetal bovine serum (FBS) at 37 °C with 5% CO_2, MCF-7 cells were cultured in RPMI-1640 media added with 10% FBS at 37 °C with 5% CO_2, BT-549 cells were cultured in RPMI-1640 media added with 10% FBS and 0.023U/ml insulin at 37 °C with 5% CO_2, and MDA-MB-231 and MDA-MB-468 cells were cultured in L-15 media added with 10% FBS at 37 °C with 5% CO_2. All assays were conducted with mycoplasma-free. For CRABP2 inhibition, BrCa cells were transfected with siRNA (5′-AGGAGGGAGACACUUUCUACATT-3′) for CRABP2 synthesised by KeyGEN (Nanjing, China) using Lipofectamine 3000 (Cat. L3000015, Invitrogen, CA). For CRABP2 overexpression, BrCa cells were transfected with overexpression plasmid synthesised by KeyGEN (Nanjing, China) using Lipofectamine 3000 (Cat. L3000015, Invitrogen, CA). The transfection efficiency was validated by quantitative real-time PCR (qRT-PCR) and western blotting analysis as previously described [29]. The primers for *CRABP2* and *GAPDH* mRNA reverse transcription were synthesised in KeyGEN (Nanjing, China). The detailed information of primers used for gene amplification was shown as follows: CRABP2: (forward) 5′-ATCGGAAAACTTCGAGGAATTGC-3′, (reverse) 5′-AGGCTCTTACAGGGC CTCC-3′; GAPDH: (forward) 5′-AGATCATCAGCAATGCCTCCT-3′, (reverse) 5′-TGAGTCCTTCCACGATACCAA-3′. The primary antibodies used as follows: CRABP2 (1:2000 dilution, Cat. 10225-1-AP, ProteinTech, Wuhan China) and GAPDH (1:2000 dilution, Cat. 60004-1-Ig, ProteinTech, Wuhan China). Protein levels were standardised to GAPDH.

The functions of CRABP2-knockdown BrCa cells were checked. For cell proliferation detection, the CCK-8 assay was applied. For cell migration and invasion detection, the Boyden chamber assay was applied. The detailed protocol was previously described [29].

In vitro cytotoxicity assay

Peripheral blood monouclear cells (PBMC) were collected from healthy control. The $CD8^+$ T cells were isolated using Dynabeads™ human CD8 selection Kit (Cat. 11333D, Invitrogen) and cultured in RPMI-1640 complete medium (10% FBS). Human T-cell activation CD3/CD28 beads (Cat. MBS-C001, Acrobiosystems) at the ratio of 1:1 beads-to-cells was used to activate cytotoxic T cells, and then activated T cells were expanded in medium supplemented with 4 ng/mL of recombinant human IL-2 protein (Cat. IL2-H4113, Acrobiosystems). Then, the T cells were transferred into a 96-well plate and co-cultured with BrCa cells at an effector-to-target ratio of 10:1 at 37 °C for 48 h. For tumour cell survival detection, tumour cells were washed to remove lymphocytes, and live cells were measured using the CCK-8 assay.

Statistical analysis

R 4.0.4 and GraphPad Prism 6 were applied for statistical analysis and figure exhibition. The parametric Student's t test or non-parametric Mann–Whitney test was used to compare the difference in quantitative data between the two groups. The parametric one-way ANOVA with Tukey's multiple comparisons test or the non-parametric Kruskal–Wallis with Dunn's multiple-comparison test was performed to measure the difference among multiple groups. Pearson correlation analysis was utilised to assess the correlation between two variables. The difference of categorical data was assessed using the Fisher exact probability test. Survival analysis was conducted by log-rank test or Cox regression analysis. For all analyses, P value < 0.05 was deemed to be statistically significant and labelled with $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$.

RESULTS

Study design and global immune landscape in BrCa

To evaluate the heterogeneous cellular composition of the TIME in BrCa, we first submitted ten BrCa samples to scRNA-seq, and then

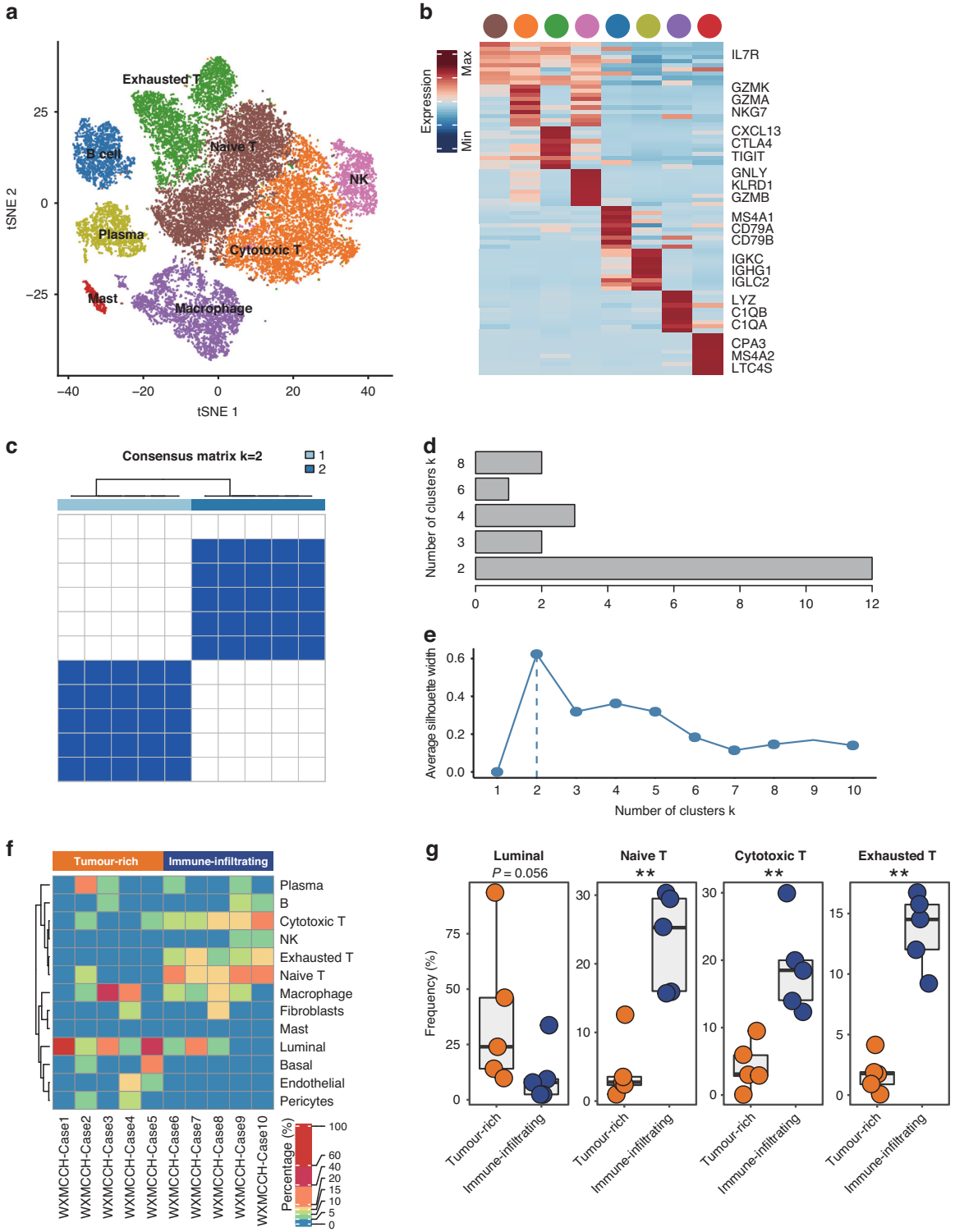

Fig. 2 Recognition of molecular subtypes of BrCa patients at the single-cell levels. a t-SNE visualisation of eight subpopulations of immune cells. **b** Heatmap of the expression of subpopulation-specific genes. **c** Consensus clustering matrix of the ten BrCa patients using the cellular composition matrix for $k = 2$. **d** NbClust analysis of BrCa patients based on the cellular composition at the single-cell level. **e** Silhouette analysis of clustering results. **f** Heatmap showing the proportions of cell types of each patient. **g** Boxplot showing the proportion of cell types between the tumour-rich group and the immune-infiltrating group. Significance was calculated with the Mann–Whitney test.

subtyping BrCa samples into two subtypes based on the cellular composition, including the tumour-rich and the immune-infiltrating subtypes. Subsequently, we extracted two subtype-special genes, namely CRABP2 and CD69, which were upregulated in tumour-rich subtype and immune-infiltrating subtype, respectively. Based on the CRABP2/CD69 signature, BrCa samples were re-divided into three subtypes, namely the CRABP2^{high}CD69^{low} subtype, the CRABP2^{low}CD69^{high} subtype, and other subgroups. Moreover, the oncogenic role of CRABP2 in BrCa was investigated using in vitro assays (Fig. 1a).

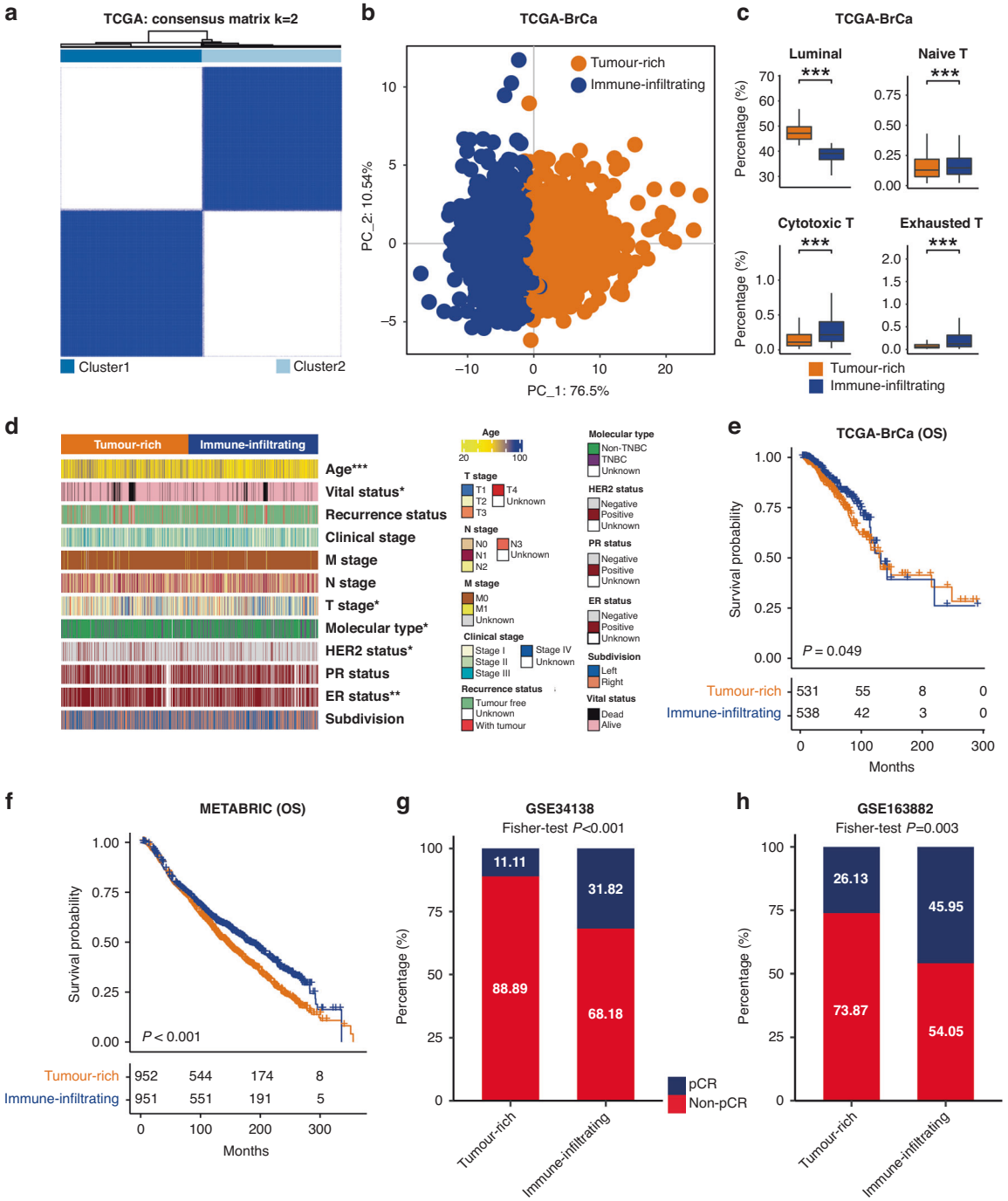

Fig. 3 Recognition of molecular subtypes of BrCa patients in the RNA-seq datasets. a Consensus clustering matrix of BrCa patients in the TCGA cohort using the cellular composition matrix estimated by BayesPrism algorithm ($k = 2$). **b** Principal component analysis of TCGA-BrCa samples based on the cellular composition matrix estimated by BayesPrism algorithm. **c** Boxplot showing the proportion of cell types between the tumour-rich group and the immune-infiltrating group. Significance was calculated with the Student *t* test. **d** Association between subgroups and clinic-pathological features in the TCGA cohort. Significance was calculated with the Fisher exact probability test. **e**, **f** Difference in the prognosis of BrCa patients in the tumour-rich group and the immune-infiltrating group in the TCGA and the METABIRC cohorts. Significance was calculated with log-rank test. **g**, **h** Difference in chemotherapeutic response of BrCa patients in the tumour-rich group and the immune-infiltrating group in the GSE34138 and the GSE163882 cohorts. Significance was calculated with the Fisher exact probability test.

To describe the cellular heterogeneity at the single-cell level, we integrated the scRNA-seq datasets of ten BrCa patients and divided the high-quality cells into 24 clusters unsupervised (Supplementary Fig. S1A, B). Based on the transcriptional levels of the conventional signatures, we annotated the various cell types in BrCa ecosystem, including endothelial cells, luminal cells, basal cells, pericytes, fibroblasts, mast cells, B cells, plasma cells, macrophages, and T/NK cells (Fig. 1b, c). By summarising the cellular components of each patient, we found that BrCa patients showed a high inter-tumoral heterogeneity (Fig. 1d and Supplementary Fig. S1C, D). To be specific, approximately five BrCa patients had higher infiltration of T/NK cells, while other patients

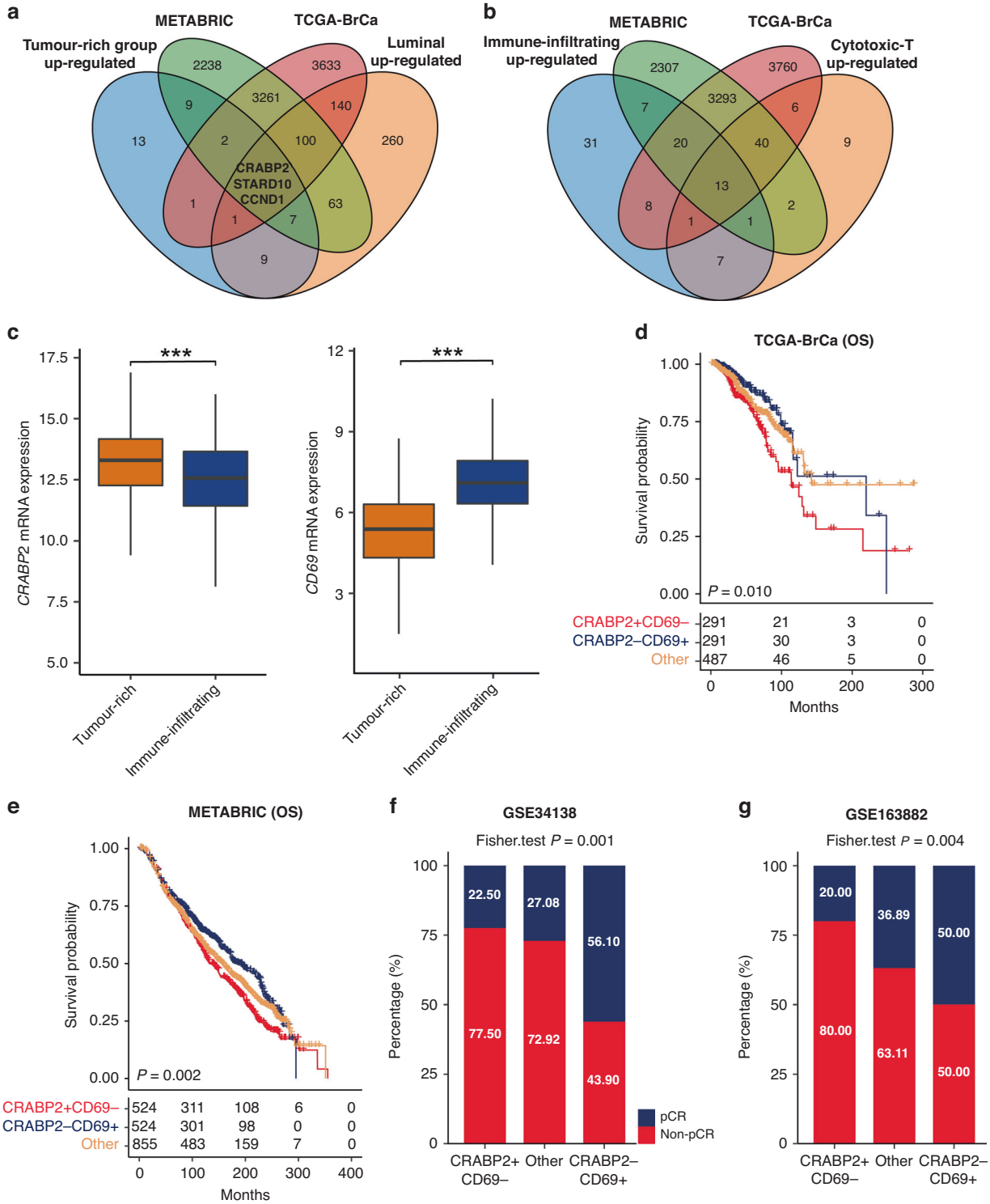

Fig. 4 Establishment of the CRABP2/CD69 signature. a, **b** Screening of special genes in the tumour-rich group and the immune-infiltrating group, respectively. **c** Expression of CRABP2 and CD69 in the tumour-rich group and the immune-infiltrating group. Significance was calculated with Student *t* test. **d**, **e** Difference in the prognosis of BrCa patients in the CRABP2^{high}CD69^{low}, the CRABP2^{low}CD69^{high} and the other subgroups in the TCGA and the METABIRC cohorts. Significance was calculated with log-rank test. **f**, **g** Difference in chemotherapeutic response of BrCa patients in the CRABP2^{high}CD69^{low}, the CRABP2^{low}CD69^{high} and the other subgroups in the GSE34138 and the GSE163882 cohorts. Significance was calculated with Fisher exact probability test.

showed higher fractions of epithelial and stromal cells. To distinguish malignant epithelial cells from non-malignant microenvironment cells, we further inferred the stemness and CNV status of each single cell according to the raw count matrix. Results showed that luminal cells had significantly higher stemness and CNV levels than those cell types (Supplementary Fig. S1E, F), suggesting that the luminal cells contribute to the malignancy of BrCa.

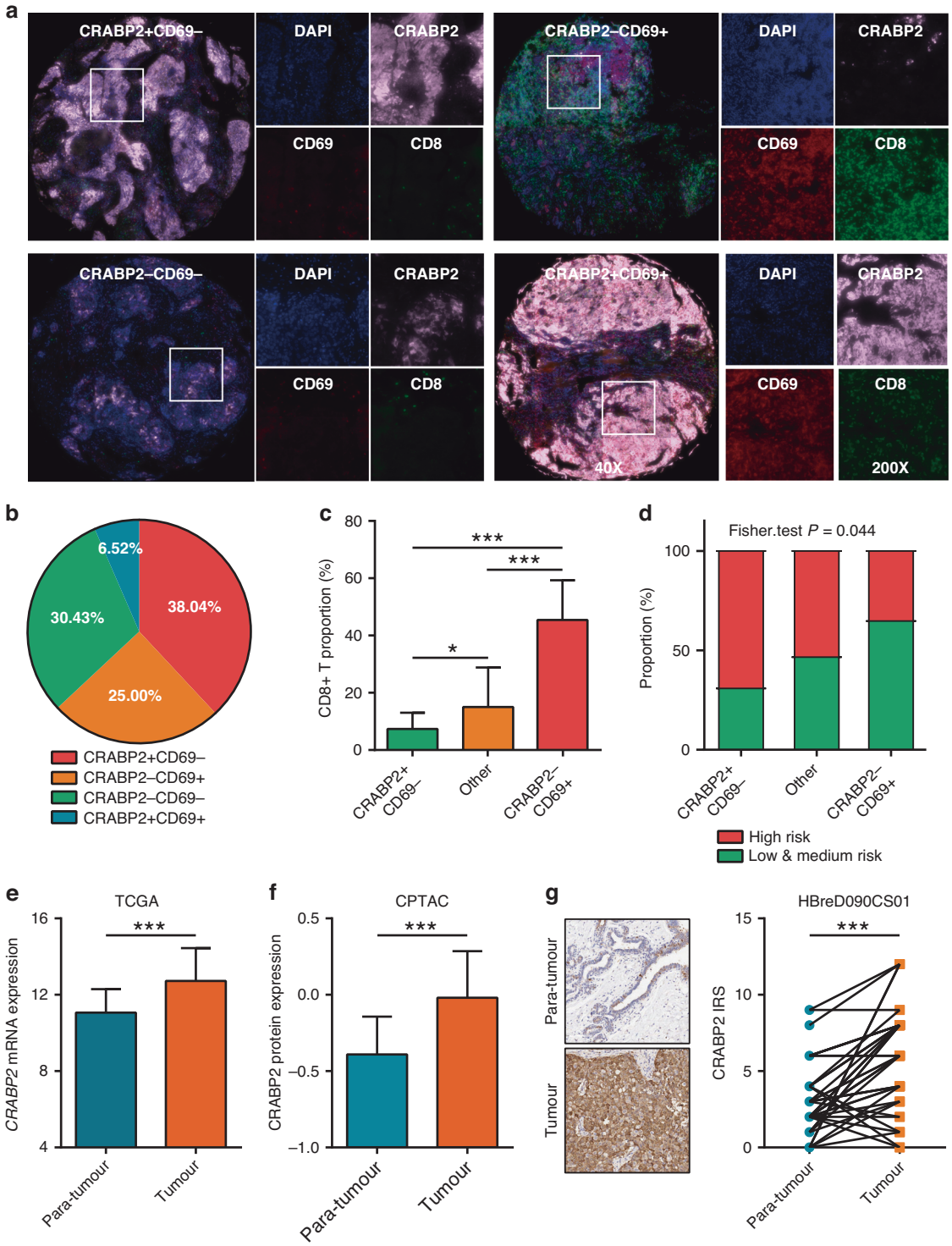

Fig. 5 Validation of the CRABP2/CD69 signature. a Representative images uncovering CRABP2, CD69 and CD8 expression in tumour tissues using mIHC staining. All tumours could be divided into the CRABP2^{high}CD69^{low}, the CRABP2^{low}CD69^{high}, the CRABP2^{low}CD69^{low} and the CRABP2^{high}CD69^{high} subgroups. **b** The proportion of the CRABP2^{high}CD69^{low}, the CRABP2^{low}CD69^{high}, the CRABP2^{low}CD69^{low} and the CRABP2^{high}CD69^{high} subgroups in the total 92 samples. **c** Difference of CD8^{+} T-cell proportion in the CRABP2^{high}CD69^{low}, the CRABP2^{low}CD69^{high} and the other subgroups. Significance was calculated with Kruskal–Wallis with Dunn's multiple-comparison test. **d** Difference of recurrent risk in the CRABP2^{high}CD69^{low}, the CRABP2^{low}CD69^{high} and the other subgroups. Significance was calculated with the Fisher exact probability test. **e** Transcriptional expression of CRABP2 in tumour and para-tumour tissues in the TCGA dataset. Significance was calculated with Student *t* test. **f** Protein expression of CRABP2 in tumour and para-tumour tissues in the CPTAC dataset. Significance was calculated with Student *t* test. **g** Representative images uncovering CRABP2 expression in para-tumour and tumour tissues using IHC staining and semi-quantitative analysis. Significance was calculated with paired Student *t* test.

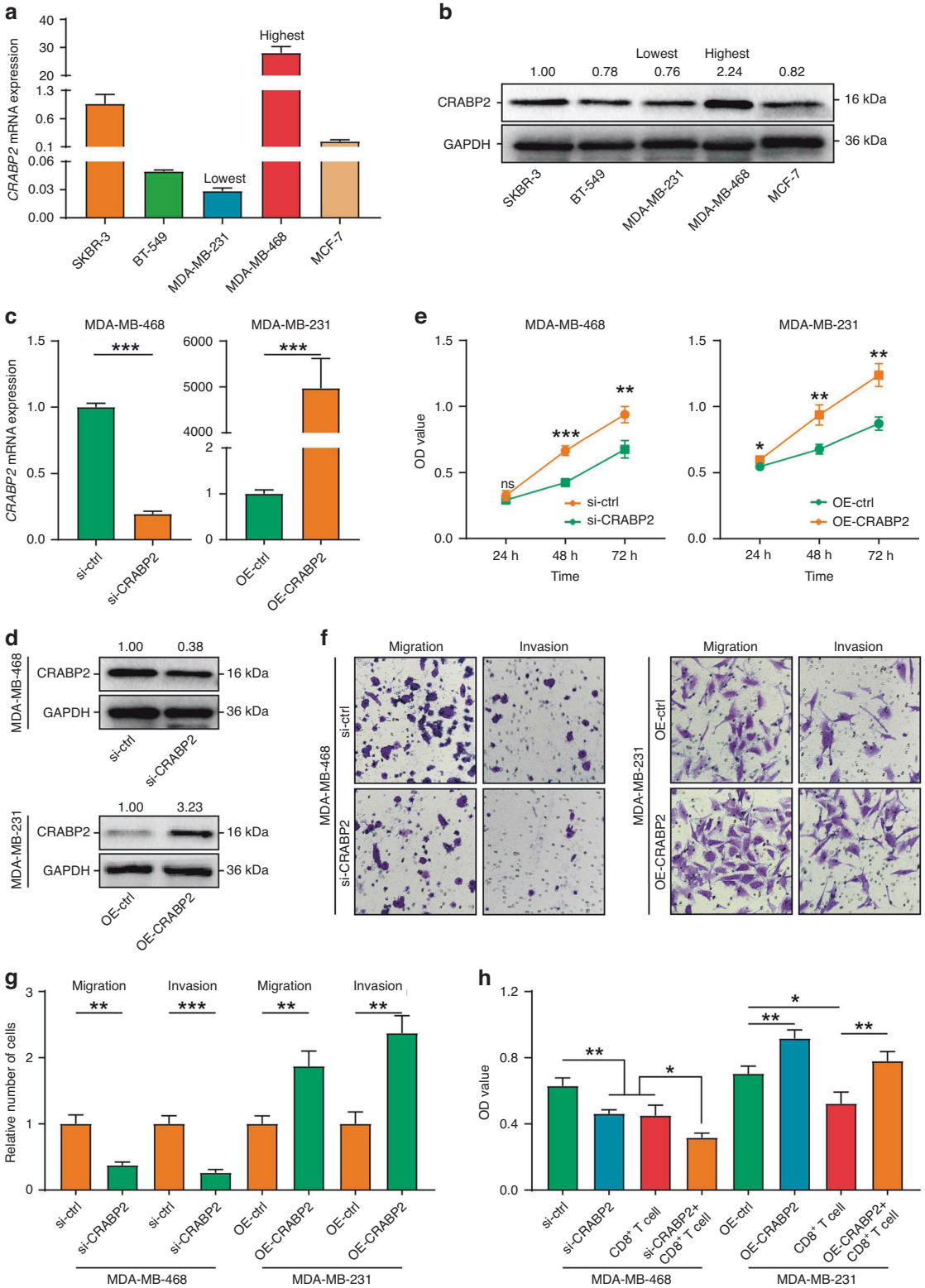

Fig. 6 The cellular role of CRABP2 in BrCa. a, **b** The expression of CRABP2 in several BrCa cells was assessed by qRT-PCR and western blotting. **c**, **d** The silencing and overexpression efficiency of CRABP2 in BrCa cells was assessed by qRT-PCR and western blotting. Significance was calculated with Student *t* test. **e** The proliferative capacity of control, CRABP2-silencing and CRABP2-overexpressed BrCa cells was investigated by the CCK-8 assay. Significance was calculated with Student *t* test. **f**, **g** The migratory and invasive capacity of control, CRABP2-silencing and CRABP2-overexpressed BrCa cells were investigated by the Boyden chamber assay. Significance was calculated with Student *t* test. **h** The cytotoxicity of $CD8^+$ T cell for control, CRABP2-silencing and CRABP2-overexpressed BrCa cells were investigated by the CCK-8 assay. Significance was calculated with one-way ANOVA with Tukey's multiple comparisons test.

Identification and validation of scRNA-seq-dependent immune subtype

Having observed that a considerable number of immune cells existed in the TIME, we wondered whether these cells distributed differently among BrCa patients. Thus, we next re-clustered the myeloid and lymphocyte cells unsupervisedly to further recognise the subpopulations of these cells. According to the cluster-specific genes and the conventional signatures, we defined them into macrophages, mast cells, B cells, plasma cells, naive T cells, exhausted T cells, cytotoxic T cells and NK (Fig. 2a, b). Given of the differentially distribution of the immune subset, especially the T/NK subpopulations among BrCa patients (Supplementary Fig. S2), we next subtype the BrCa patients into several stable groups according to the cellular components at the single-cell level. Consensus clustering, NbClust testing and silhouette analysis suggested that two was the optimal cluster number (Fig. 2c–e and Supplementary Fig. S3). Subsequently, all ten patients were classified into two heterogeneous subtypes by performing consensus clustering (see "Methods") (Fig. 2f). The tumour-rich group was featured by the accumulation of numerous luminal cells, the tumour cells which contribute to the malignancy of BrCa (Fig. 2f, g). The immune-infiltrating group was featured by the high infiltration of immune cells, especially the T-cell subpopulations (Fig. 2f, g and Supplementary Fig. S4).

To further validate the cellular composition-based grouping, we used a consensus cluster to divide the BrCa samples into two groups (Fig. 3a, b), and the cell markers between two groups were compared to ensure the accuracy of the grouping (Fig. 3c and Supplementary Fig. S5). Next, we assessed the association between the cellular composition-based subgroups and clinicopathological parameters. Obviously, subgroups were associated with age, vital status, T stage, molecular type, HER2 status and ER status (Fig. 3d). Moreover, the prognosis of BrCa patients in the tumour-rich group was worse than that in the immune-infiltrating subtypes (Fig. 3e). The prognostic value of the cellular composition-based subgroups was also validated in the METABRIC cohort, and the similar result was observed (Fig. 3f). Furthermore, we also found that the cellular composition-based subgroups were associated with the chemotherapeutic responses, and BrCa patients in the tumour-rich group exhibited poor response to chemotherapy in both the GSE34138 cohort and the GSE163882 cohort (Fig. 3g, h). Taken together, these results suggest that cellular composition-based subgroups were associated with both prognosis and chemotherapeutic response in BrCa.

Development of the CRABP2/CD69 classifier based on immune subtype

Given that the cellular composition-based subtyping strategy could not be applied in clinical practice unless scRNA-seq and/or RNA-seq data is available, we tried to extract specific genes associated with cellular composition-based subgroups. After evaluating the distinguishing ability of genes both at the single-cell and transcriptomic omics datasets (see "Methods") (Fig. 4a, b), we found that CRABP2 was specifically expressed in the tumour-rich phenotype, and CD69 was highly expressed in the immune-infiltrating group (Fig. 4c). We re-divided BrCa patients according to the median expression levels of CRABP2 and CD69. The results exhibited that BrCa patients with the $CRABP2^{high}CD69^{low}$ feature showed the worst clinical outcome, and patients with the $CRABP2^{low}CD69^{high}$ feature exhibited the most favourable prognosis in both the TCGA and the METABIRC cohorts (Fig. 4d, e). Moreover, in two independent NAT cohorts, BrCa patients with the $CRABP2^{high}CD69^{low}$ feature showed the lowest NAT response and patients with the $CRABP2^{low}CD69^{high}$ feature showed the highest NAT response (Fig. 4f, g). In addition, we also conducted Cox regression analysis and found that the CRABP2/CD69 signature was the independent prognostic factor in BrCa in the TCGA and the METABIRC cohorts (Supplementary Fig. S6A–D). Furthermore, BrCa patients with the $CRABP2^{high}CD69^{low}$ feature exhibited the lowest immune infiltration, and patients with the $CRABP2^{low}CD69^{high}$ feature exhibited the highest immune infiltration in the TCGA, the METABIRC, the GSE34138 and the GSE163882 cohorts (Supplementary Fig. S7A–E).

To further verify the predictive value of the CRABP2/CD69 signature in BrCa, we collected 92 BrCa samples from Wuxi Maternal and Child Health Hospital and submitted these samples for mIHC analysis. These 92 BrCa patients were grouped according to the median expression levels of CRABP2 and CD69 (Fig. 5a). Notably, the patients with the $CRABP2^{high}CD69^{high}$ feature were limited (Fig. 5b). To verify the above findings, we compared immune cell infiltration and recurrence risk in the current cohort. The results showed that BrCa patients with the $CRABP2^{high}CD69^{low}$ feature exhibited the highest recurrent risk and lowest $CD8^+$ T cells infiltration, and patients with the $CRABP2^{low}CD69^{high}$ exhibited the opposite findings (Fig. 5c, d). Overall, the CRABP2/CD69 signature is significantly associated with immune cell infiltration and recurrence risk in BrCa.

CRABP2 promotes BrCa progression and CTLs exhaustion

We next examined the functional role of CRABP2 in BrCa. First, the IHC assay revealed that the expression of CRABP2 was notably enhanced in BrCa tumour tissues compared with para-tumour tissues in three independent cohorts, including the TCGA cohort, the CPTAC cohort and the HBreD090CS01 cohort (Fig. 5e–g). Then, we compared CRABP2 expression in several BrCa cell lines, including MDA-MB-231, BT-549, MCF-7, SK-BR-3 and MDA-MB-468, and the results show that MDA-MB-231 expressed the lowest CRABP2 and MDA-MB-468 expressed the highest CRABP2 (Fig. 6a, b). Thus, we conducted knockdown experiments in MDA-MB-468 cells and overexpression experiments in MDA-MB-231 cells. The efficiency of CRABP2 knockdown and overexpression in these two BrCa cells was evaluated by qRT-PCR and western blotting assays (Fig. 6c, d). Compared with cells in the control group, CRABP2-silencing BrCa cells showed inhibited proliferative capacity, and CRABP2-overexpressed BrCa cells showed enhanced proliferative capacity (Fig. 6e). Moreover, CRABP2 knockdown notably suppressed the migratory and invasive capacities of BrCa cells, and CRABP2 overexpression exhibited the opposite effects (Fig. 6f, g). Moreover, CRABP2 knockdown in tumour cells enhanced the cytotoxic effects of $CD8^+$ T cells, and CRABP2 overexpression in tumour cells contributed to immune escape (Fig. 6h). Overall, CRABP2 functions as an oncogene in BrCa, which could be used as a novel target for BrCa therapy.

To investigate the interaction between luminal cells that expressed or non-expressed CRABP2 and immune cells, we perform a high-resolution dissection of interactions among various cell types based on the combined expression of multi-subunit ligand-receptor complexes. Based on the expression levels of CRABP2, the luminal cells were divided into $CRABP2^+$ and $CRABP2^-$ groups. Results showed that the $CRABP2^+$ luminal cells presented significantly more interactions with immune cells than the $CRABP2^-$ luminal cells (Supplementary Fig. S8A). Notably, we calculated the difference in interaction numbers among various cell types. The $CRABP2^+$ luminal cells, showed higher communication strength with immune cells, especially the $CD8^+$ T/NK cells (Supplementary Fig. S8B). These results collectively suggested that compared with $CRABP2^-$ luminal cells, $CARBP2^+$ luminal cells had activated cell–cell communications, especially interactions with immune cells, which potentially take part in the formation of an immunosuppressive TIME. We further identified the significant ligand-receptor interactions between luminal and $CD8^+$ T/NK using the CellphoneDB tool. Results showed that $CRABP2^+$ luminal cells communicated with $CD8^+$ T/NK cells via some inhibitory interaction, such as CD47-SIRPG, CD74-MIF and TGFB1-TGFβ receptor [30, 31] (Supplementary Fig. S8C). Besides, lines of evidence showed that the CXCR6-CXCL16 interaction between

CRABP2$^+$ luminal cells and CD8$^+$ T/NK cells played important roles in the progression of BrCa [32, 33]. Summing up, at the single-cell level, we found that compared with CRABP2$^-$ luminal cells, CRABP2$^+$ luminal cells had stronger interactions with immune cells, especially the CD8$^+$ T/NK cells. Also, CRABP2$^+$ luminal cells can communicate with CD8$^+$ T/NK cells via ligand-receptor interactions which were associated with the shaping of immunosuppressive TIME and the progression of BrCa.

DISCUSSION

A great deal of evidence support the opinion that TIME plays a critical role in regulating tumour progression and determining therapeutic responses in not only BrCa but in most solid tumours [34–36]. BrCa is the tumour type with the highest incidence worldwide, and not all patients have a satisfactory prognosis [1]. The rapid expansion of scRNA-seq technology greatly promotes the resolution of TIME and the discrimination of tumour heterogeneity in BrCa [37]. Due to the continuous progress in typing strategies, it seems that the barriers created by classical molecular typing could be broken [38–40]. In this research, we collected ten BrCa samples with various molecular types and submitted these samples for scRNA-seq. We divided these BrCa samples into two subtypes based on the heterogeneous cellular composition of the TIME, including the tumour-rich and the immune-infiltrating subtypes. Next, this subtyping strategy was also applied in several public cohorts, which indeed exhibited distinct prognosis and chemotherapeutic responses in the TCGA, the METABIRC, the GSE34138 and the GSE163882 cohorts.

However, the cellular composition-based subtyping strategy could not be applied in clinical practice unless scRNA-seq and/or RNA-seq data is available. Thus, we tried to establish a novel translational subtyping strategy based on the existing subtyping. Previous research revealed that classification based on specific cell markers could contribute to risk discrimination and TIME feature identification [41, 42]. We extracted CRABP2 as the biomarker for the tumour-rich subtype and CD69 as the biomarker for the immune-infiltrating subtype, and further used these two biomarkers to establish a novel subtyping strategy. Based on the CRABP2/CD69 signature, BrCa samples were re-divided into three subtypes, and the CRABP2^{high}CD69^{low} subtype exhibited the most unfavourable prognosis and the lowest chemotherapeutic response, and the CRABP2^{low}CD69^{high} subtype showed the opposite results.

CD69 is expressed in most leucocytes and notably enhanced upon activation. Thus, CD69 is always used as a marker for activated natural killer cells and lymphocytes [43]. CD69 is also a marker for tissue-resident memory T (Trm) cells [44]. However, the functional role of CD69 has been mostly revealed in autoimmune diseases [45], its function in tumours is largely missing. It has been uncovered that CD69 expression is positively associated with most immune checkpoints and TILs, and acts as a promising biomarker to predict the response to anti-PD-1/PD-L1 immunotherapy in lung cancer and melanoma [46]. In cholangiocarcinoma, CD69$^+$CD103$^+$ Trm-like CD8$^+$ TILs mediate momentous cancer-specific immune responses and could be used as a candidate potential therapeutic target [47]. In the current research, CD69 was used as a marker for the CRABP2/CD69 signature in BrCa. However, the functional role of CD69 was not uncovered, which should be further studied.

CRABP2 encodes a member of the retinoic acid binding protein family and lipocalin/cytosolic fatty acid-binding protein family [48]. CRABP2 has been revealed to be upregulated and function as an oncogene in various cancers, including gastric cancer [49], thyroid cancer [50], ovarian cancer [51] and lung cancer [52]. CRABP2 is a key molecule in oxaliplatin resistance by attenuating mitochondrial apoptosis in gastric cancer [49], which provides a potential correlation between CRABP2 and resistance to chemotherapy. In BrCa, Zhao et al. reveal that CRABP2 is upregulated in BrCa tissues and functions as a biomarker that is sensitive to goserelin [53], but the functional role of CRABP2 has not been well studied. In this research, we used CRABP2 as a component of the CRABP2/CD69 signature to discriminate BrCa patients into different subtypes. Moreover, we also found that CRABP2 was highly expressed in BrCa tissues and acted as a tumour-promoting gene by accelerating cell proliferation, migration and invasion. In addition, inhibition of CRABP2 also suppressed the immune evasion of tumour cells and restored the activity of CTLs.

It must be admitted that the current study still has several limitations. First, due to the fact that the BrCa patients in the WXMCCH cohort were all diagnosed in the last 3 years, the prognostic value of the CRABP2/CD69 signature could not be validated in the in-house cohort. In addition, given that the CRABP2/CD69 signature was significantly associated with TIME features, whether it could predict the immunotherapeutic responses should be further explored. Moreover, the CRABP2/CD69 signature was only tested in a small-scale multiple-centre cohort, so the universality of the model remains to be externally validated.

CONCLUSIONS

Collectively, the current research established a novel practical subtyping strategy based on CRABP2 and CD69 expression. The established signature was significantly associated with TIME features and the response to NAT in BrCa. Moreover, we also revealed CRABP2 as a novel oncogene in BrCa, which could be a notable target to control tumour progression and immune evasion.

DATA AVAILABILITY

Available of public BrCa datasets are described in "Methods". Data can be provided upon reasonable request to the corresponding author.

MATERIALS AVAILABILITY

Materials can be provided upon reasonable request to the corresponding author.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72:7–33.
2. Byrd DR, Brierley JD, Baker TP, Sullivan DC, Gress DM. Current and future cancer staging after neoadjuvant treatment for solid tumors. CA Cancer J Clin. 2021;71:140–8.
3. Kwa M, Makris A, Esteva FJ. Clinical utility of gene-expression signatures in early stage breast cancer. Nat Rev Clin Oncol. 2017;14:595–610.
4. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat Med. 2018;24:541–50.
5. Ge R, Wang Z, Cheng L. Tumor microenvironment heterogeneity an important mediator of prostate cancer progression and therapeutic resistance. NPJ Precis Oncol. 2022;6:31.
6. Giraldo NA, Sanchez-Salas R, Peske JD, Vano Y, Becht E, Petitprez F, et al. The clinical role of the TME in solid cancer. Br J Cancer. 2019;120:45–53.
7. Mao X, Xu J, Wang W, Liang C, Hua J, Liu J, et al. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. Mol Cancer. 2021;20:131.
8. Gajewski TF. The next hurdle in cancer immunotherapy: overcoming the non-T-cell-inflamed tumor microenvironment. Semin Oncol. 2015;42:663–71.
9. Mao W, Cai Y, Chen D, Jiang G, Xu Y, Chen R, et al. Statin shapes inflamed tumor microenvironment and enhances immune checkpoint blockade in non-small cell lung cancer. JCI Insight. 2022;7:e161940.
10. Ladoire S, Arnould L, Apetoh L, Coudert B, Martin F, Chauffert B, et al. Pathologic complete response to neoadjuvant chemotherapy of breast carcinoma is associated with the disappearance of tumor-infiltrating foxp3+ regulatory T cells. Clin Cancer Res. 2008;14:2413–20.
11. Ueno T, Kitano S, Masuda N, Ikarashi D, Yamashita M, Chiba T, et al. Immune microenvironment, homologous recombination deficiency, and therapeutic

response to neoadjuvant chemotherapy in triple-negative breast cancer: Japan Breast Cancer Research Group (JBCRG)22 TR. BMC Med. 2022;20:136.
12. Shepherd JH, Ballman K, Polley MC, Campbell JD, Fan C, Selitsky S, et al. CALGB 40603 (alliance): long-term outcomes and genomic correlates of response and survival after neoadjuvant chemotherapy with or without carboplatin and bevacizumab in triple-negative breast cancer. J Clin Oncol. 2022;40:1323–34.
13. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–4.
14. de Ronde JJ, Lips EH, Mulder L, Vincent AD, Wesseling J, Nieuwland M, et al. SERPINA6, BEX1, AGTR1, SLC26A3, and LAPTM4B are markers of resistance to neoadjuvant chemotherapy in HER2-negative breast cancer. Breast Cancer Res Treat. 2013;137:213–23.
15. Chen J, Hao L, Qian X, Lin L, Pan Y, Han X. Machine learning models based on immunological genes to predict the response to neoadjuvant therapy in breast cancer patients. Front Immunol. 2022;13:948601.
16. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36:411–20.
17. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96.
18. Chu T, Wang Z, Pe'er D, Danko CG. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat Cancer. 2022;3:505–17.
19. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3.
20. Seiler M, Huang CC, Szalma S, Bhanot G. ConsensusCluster: a software tool for unsupervised cluster discovery in numerical data. OMICS. 2010;14:109–13.
21. Lovmar L, Ahlford A, Jonsson M, Syvanen AC. Silhouette scores for assessment of SNP genotype clusters. BMC Genomics. 2005;6:35.
22. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020;15:1484–506.
23. Cai Y, Ji W, Sun C, Xu R, Chen X, Deng Y, et al. Interferon-induced transmembrane protein 3 shapes an inflamed tumor microenvironment and identifies immune-hot tumors. Front Immunol. 2021;12:704965.
24. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
25. Wang Y, Deng J, Wang L, Zhou T, Yang J, Tian Z, et al. Expression and clinical significance of PD-L1, B7-H3, B7-H4 and VISTA in craniopharyngioma. J Immunother Cancer. 2019;7:1–9.
26. Chen L, Dong J, Li Z, Chen Y, Zhang Y. The B7H4-PDL1 classifier stratifies immuno-phenotype in cervical cancer. Cancer cell Int. 2022;22:3.
27. Mei J, Liu Y, Yu X, Hao L, Ma T, Zhan Q, et al. YWHAZ interacts with DAAM1 to promote cell migration in breast cancer. Cell Death Discov. 2021;7:221.
28. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Ann Oncol. 2015;26:259–71.
29. Mei J, Cai Y, Wang H, Xu R, Zhou J, Lu J, et al. Formin protein DIAPH1 positively regulates PD-L1 expression and predicts the therapeutic response to anti-PD-1/PD-L1 immunotherapy. Clin Immunol. 2022;246:109204.
30. McCracken MN, Cha AC, Weissman IL. Molecular pathways: activating T cells after cancer cell phagocytosis from blockade of CD47 "don't eat me" signals. Clin Cancer Res. 2015;21:3597–601.
31. Muraoka RS, Dumont N, Ritter CA, Dugger TC, Brantley DM, Chen J, et al. Blockade of TGF-beta inhibits mammary tumor cell viability, migration, and metastases. J Clin Investig. 2002;109:1551–9.
32. Mir H, Kapur N, Gales DN, Sharma PK, Oprea-Ilies G, Johnson AT, et al. CXCR6-CXCL16 axis promotes breast cancer by inducing oncogenic signaling. Cancers. 2021;13:3568.
33. Xiao G, Wang X, Wang J, Zu L, Cheng G, Hao M, et al. CXCL16/CXCR6 chemokine signaling mediates breast cancer progression by pERK1/2-dependent mechanisms. Oncotarget. 2015;6:14165–78.
34. Pitt JM, Marabelle A, Eggermont A, Soria JC, Kroemer G, Zitvogel L. Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. Ann Oncol. 2016;27:1482–92.
35. Zhang X, Zhao L, Zhang H, Zhang Y, Ju H, Wang X, et al. The immunosuppressive microenvironment and immunotherapy in human glioblastoma. Front Immunol. 2022;13:1003651.
36. Ruffin AT, Li H, Vujanovic L, Zandberg DP, Ferris RL, Bruno TC. Improving head and neck cancer therapies by immunomodulation of the tumour microenvironment. Nat Rev Cancer. 2022;23:173–88.
37. Ding S, Chen X, Shen K. Single-cell RNA sequencing in breast cancer: understanding tumor heterogeneity and paving roads to individualized therapy. Cancer Commun. 2020;40:329–44.
38. Wang S, Xiong Y, Zhang Q, Su D, Yu C, Cao Y, et al. Clinical significance and immunogenomic landscape analyses of the immune cell signature based prognostic model for patients with breast cancer. Brief Bioinform. 2021;22:bbaa311.
39. Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell. 2021;39:845–65 e7.
40. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A comprehensive Pan-cancer molecular study of gynecologic and breast cancers. Cancer Cell. 2018;33:690–705 e9.
41. Cui K, Yao S, Liu B, Sun S, Gong L, Li Q, et al. A novel high-risk subpopulation identified by CTSL and ZBTB7B in gastric cancer. Br J Cancer. 2022;127:1450–60.
42. Cui K, Yao S, Zhang H, Zhou M, Liu B, Cao Y, et al. Identification of an immune overdrive high-risk subpopulation with aberrant expression of FOXP3 and CTLA4 in colorectal cancer. Oncogene. 2021;40:2130–45.
43. Gonzalez-Amaro R, Cortes JR, Sanchez-Madrid F, Martin P. Is CD69 an effective brake to control inflammatory diseases? Trends Mol Med. 2013;19:625–32.
44. Cibrian D, Sanchez-Madrid F. CD69: from activation marker to metabolic gatekeeper. Eur J Immunol. 2017;47:946–53.
45. Gorabi AM, Hajighasemi S, Kiaie N, Gheibi Hayat SM, Jamialahmadi T, Johnston TP, et al. The pivotal role of CD69 in autoimmunity. J Autoimmun. 2020;111:102453.
46. Hu ZW, Sun W, Wen YH, Ma RQ, Chen L, Chen WQ, et al. CD69 and SBK1 as potential predictors of responses to PD-1/PD-L1 blockade cancer immunotherapy in lung cancer and melanoma. Front Immunol. 2022;13:952059.
47. Kim HD, Jeong S, Park S, Lee YJ, Ju YS, Kim D, et al. Implication of CD69(+) CD103(+) tissue-resident-like CD8(+) T cells as a potential immunotherapeutic target for cholangiocarcinoma. Liver Int. 2021;41:764–76.
48. Salazar J, Guardiola M, Ferre R, Coll B, Alonso-Villaverde C, Winklhofer-Roob BM, et al. Association of a polymorphism in the promoter of the cellular retinoic acid-binding protein II gene (CRABP2) with increased circulating low-density lipoprotein cholesterol. Clin Chem Lab Med. 2007;45:615–20.
49. Tang X, Liang Y, Sun G, He Q, Hou Z, Jiang X, et al. Upregulation of CRABP2 by TET1-mediated DNA hydroxymethylation attenuates mitochondrial apoptosis and promotes oxaliplatin resistance in gastric cancer. Cell Death Dis. 2022;13:848.
50. Liu CL, Hsu YC, Kuo CY, Jhuang JY, Li YS, Cheng SP. CRABP2 is associated with thyroid cancer recurrence and promotes invasion via the integrin/FAK/AKT pathway. Endocrinology. 2022;163:bqac171.
51. Xie T, Tan M, Gao Y, Yang H. CRABP2 accelerates epithelial mesenchymal transition in serous ovarian cancer cells by promoting TRIM16 methylation via upregulating EZH2 expression. Environ Toxicol. 2022;37:1957–67.
52. Wu JI, Lin YP, Tseng CW, Chen HJ, Wang LH. Crabp2 promotes metastasis of lung cancer cells via HuR and integrin beta1/FAK/ERK signaling. Sci Rep. 2019;9:845.
53. Zhao Y, Sun H, Zheng J, Shao C, Zhang D. Identification of predictors based on drug targets highlights accurate treatment of goserelin in breast and prostate cancer. Cell Biosci. 2021;11:5.

AUTHOR CONTRIBUTIONS

YZ, XP and TX designed and performed the experiments. JM, YC, LC, YW and JL conducted the data analysis of single-cell sequencing. JM, ZQ, YJ and PZ performed the statistics and analysis of clinical data. JL and YJ helped the research methods. YC handled the processing of single-cell RNA sequencing. JM, YC, LC and YW organised the data and wrote the manuscript. YZ, XP and TX supervised the study. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Precision Medicine Project of Wuxi Municipal Health Commission (J202106), the Maternal and Child Health Research Project of Jiangsu Province (F202034), the Major project of Wuxi Science and Technology Bureau (N20201006), the 333 Project of Province (BRA2020380), the Wujieping Project (320.6750.2022-19-38).

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The collection of the cohort 1 and the cohort 6 was approved by the institutional review board at Wuxi Maternal and Child Health Hospital (2021-01-0927-28), and the

collection of the cohort 8 was approved by the Clinical Research Ethics Committee in Outdo Biotech (YB-M-05-02). The cohort 2, the cohort 3, the cohort 4, the cohort 5, and the cohort 7 were public cohorts and no ethical approval was needed. All experiments were performed in accordance with the Declaration of Helsinki, and informed consent was obtained from all subjects.

CONSENT FOR PUBLICATION

Not applicable.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-023-02432-6.

Correspondence and requests for materials should be addressed to Tiansong Xia, Xiang Pan or Yan Zhang.