Article

Prognostic Value of Copy Number Alteration Burden in Early-Stage Breast Cancer and the Construction of an 11-Gene Copy Number Alteration Model

Dingyuan Wang [1], **Songlin Gao** [2], **Haili Qian** [3], **Peng Yuan** [2,*] **and Bailin Zhang** [1,*]

1 Department of Breast Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

2 Department of VIP Medical Services, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

3 State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China

* Correspondence: yuanpeng01@hotmail.com (P.Y.); bailin_zhang@cicams.ac.cn (B.Z.)

Simple Summary: Breast cancer is a malignancy that poses a significant threat to women's health. The enormous disease burden has forced a wide range of researchers to develop more accurate prognostic models. Copy number alterations, which are amplifications or deletions of DNA fragments, often predict a poor prognosis. Instead, copy number alteration burden, i.e., the level of CNA, may have a good predictive value for disease prognosis. In this study, we developed a prognostic model for early breast cancer based on CNAB and simplified it. It performed excellently in two external validation sets.

Abstract: The increasing burden of breast cancer has prompted a wide range of researchers to search for new prognostic markers. Considering that tumor mutation burden (TMB) is low and copy number alteration burden (CNAB) is high in breast cancer, we built a CNAB-based model using a public database and validated it with a Chinese population. We collected formalin-fixed, paraffin-embedded (FFPE) tissue samples from 31 breast cancer patients who were treated between 2010 and 2014 at the National Cancer Center (CICAMS). METABRIC and TCGA data were downloaded via cBioPortal. In total, 2295 patients with early-stage breast cancer were enrolled in the study, including 1427 in the METABRIC cohort, 837 in the TCGA cohort, and 31 in the CICAMS cohort. Based on the ROC curve, we consider 2.2 CNA/MBp as the threshold for the CNAB-high and CNAB-low groupings. In both the TCGA cohort and the CICAMS cohort, CNAB-high had a worse prognosis than CNAB-low. We further simplified this model by establishing a prognostic nomogram for early breast cancer patients by 11 core genes, and this nomogram was highly effective in both the TCGA cohort and the CICAMS cohort. We hope that this model will subsequently help clinicians with prognostic assessments.

Keywords: breast cancer; copy number alteration burden; prognosis

1. Introduction

Breast cancer is the most prevalent malignancy in women. According to the National Cancer Center of China, the estimated number of new breast cancer cases in China is as high as 304,000 [1]. The increasing burden of breast cancer has prompted a wide range of researchers to search for new prognostic markers [2]. With the promotion of next-generation sequencing technology, an increasing number of multigene models are being established and used in clinical practice due to their accuracy compared to traditional clinical models [3,4].

Somatic cell copy number alterations (CNAs) are one of the hallmarks specific to malignancy, and represent the amplification or deletion of a DNA fragments [5]. The CNA of a gene implies genomic instability and often predicts a poor prognosis [6–9]. It is generally accepted that CNA increases with increasing cancer stage and is higher in patients with advanced breast cancer than in patients with early-stage breast cancer [10]. Copy number alterations of individual genes are often the result of altered chromosomal segments. Takayuki pointed out that loss at 6q, 13q, and 16q, as well as gain at 1q, 6p, 8q, 9p, 11q, 16p, 17q, 19q, and 20q in patients with breast cancer can predict high chromosomal instability, tumor immune escape, and strong tumor aggressiveness [11,12]. Copy number alteration burden (CNAB), as a total level of CNA, will better indicate chromosome stability as well as tumor prognosis. Tumor mutational burden (TMB), another indicator of genomic stability, is often used in prognostic studies of tumors. Breast cancer is one of the tumor types with low mutation frequency, resulting in low TMB with low variability. The mutation frequencies of PI3KCA and TP53, the two most commonly mutated genes in breast cancer, were 34.5% and 34.3%, respectively, and nonsynonymous TMB was only 1.20 Muts/Mb [13]. A study by Liu's team included eight cancer cohorts, including breast cancer (n = 1234), for survival analysis. This study found no significant difference in overall survival between patients with high TMB and low TMB (p = 0.351) [14]. However, there was a significant difference in overall survival between breast cancer patients in the high-CNAB and low-CNAB cohorts (p = 0.004). In addition, CNAB is also a significant predictor of survival for tumor patients with many other cancer types [15–17]. Considering the high frequency of CNA and low frequency of mutation in breast cancer, we believe that the utilization of CNA as a prognostic marker is a very promising topic.

Although transcriptome-based early breast cancer prognostic models continue to be used in the clinic with well-recognized effectiveness, with the improvement of liquid biopsy technology, ctDNA-based prognostic models may be more widely used in the future due to their advantageous features (e.g., they are less invasive and provide reproducible measurements). Compared with transcriptome-based prognostic models, the establishment of CNAB-based prognostic models would be more conducive to the development of future ctDNA prognostic models, thereby allowing clinicians to make prognostic assessments more conveniently.

Admittedly, some studies have revealed that CNAB is a prognostic factor for early-stage breast cancer [2,6,10,18]. However, prognostic evaluation based on whole-exome CNAB would obviously place a great cost burden on patient and be very inconvenient for laboratory physicians. Therefore, we attempted to optimize an all-exon CNAB model. A traditional method of screening for core genes is to find the genes most associated with prognosis by statistical methods. However, this can lead to potential false-positive results. Considering that CNAs are more likely to affect biological pathways only if they alter the transcription of genes, we performed a secondary screen among prognosis-related CNAs, eliminating genes that are not strongly linked to mRNA expression. We finally obtained an 11-gene model to accurately assist clinicians and facilitate treatment decisions.

2. Materials and Methods

2.1. Ethics

The study methodologies conformed to the standards set by the Declaration of Helsinki and were approved by the Ethics Committee of the National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (NCC-2021C-369). All patients signed an informed consent form in writing.

2.2. Study Design

This retrospective study collected formalin-fixed, paraffin-embedded (FFPE) tissue samples from 31 breast cancer patients who were treated between 2010 and 2014 at the Cancer Hospital, Chinese Academy of Medical Sciences (CICAMS). The patients included

all had early invasive breast cancer. Resection specimens rather than core biopsies were used for DNA extraction. The METABRIC cohort and TCGA cohort are two large breast cancer cohorts, and we downloaded clinical information for both cohorts via the cBioPortal website (https://www.cbioportal.org, accessed on 1 July 2022).

For the CICAMS cohort, we retrieved the patients' age, pathological details, and treatment information from their medical records. The main variables in this analysis were (a) the patients' demographic characteristics (age, etc.) and (b) their clinical information (grade, ER status, PR status, HER2 status, surgery, radiotherapy, and chemotherapy). The stage of breast cancer was categorized based on the American Joint Committee on Cancer, Seventh edition [19]. The grade was categorized as grade I to III based on the WHO [20]. Age was categorized as younger than 60 years or 60 years and older. Those with missing or unclear records were categorized as unknown. For the METABRIC and TCGA cohorts, we retrieved the following clinical variables: age, grade, stage, molecular subtype, and treatment strategies.

2.3. Genomic Information Acquisition

Details of the extraction of DNA and next-generation sequencing are available in the Supplementary Methods. CNA, TMB, and mRNA data of the METABRIC cohort and TCGA cohort were obtained from https://www.cbioportal.org. We investigated the transcriptional data by fragments per kilobase of transcript per million mapped reads (FPKM) values. We used these transcriptomic data for subsequent differentially expressed gene analysis and gene ontology analysis. For the assessment of CNA, we used the GISTIC 2.0 criterion [21]. This criterion uses a fixed algorithm to transform the amplification or deletion status of each gene into an integer between −2 and 2.

2.4. CNAB11 Modeling Method

We used the METABRIC training cohort to obtain prognosis-related CNAs by differential CNA analysis with 5-year recurrence/metastasis as the endpoint event. Then, we further screened reliable prognosis-related CNAs with a strong association with mRNA using the Kolmogorov–Smirnov test. Finally, 11 genes (CBWD1, CDC6, CWC25, HS3ST3A1, IFNA2, KDM4C, KRT27, MLLT6, NBR1, NBR2, and ZDHHC21) were included in the CNAB11 model. The model combined these 11 gene CNAs (GISTIC 2.0) to obtain the CNAB11 score. The cutoff for the CNAB11 cluster was then obtained from the receiver operating characteristic curve.

2.5. Statistical Analysis

The differences in clinicopathological characteristics and treatment strategies were compared through the chi-square test or Fisher's exact test for categorical variables and the Wilcoxon rank sum test for ordered variables. Among the groups, prognostic differences were estimated with the log-rank test for categorical variables. Cox regression was used to calculate hazard ratios (HRs) and their 95% confidence intervals (CIs). Relapse-free survival (RFS) was defined as the time from radical resection for breast cancer to the earliest time of recurrence or death from any cause. Overall survival (OS) was defined as the time from the date of diagnosis to the date of death due to any cause or to the last follow-up [22]. Based on the gene expression data of the TCGA cohort, we used the "DESeq2" package in R to analyze the differentially expressed genes (DEGs) between the two subgroups. The screening criteria for DEGs were $p < 0.05$ and absolute log2FC > 2. We used the Kolmogorov–Smirnov test to examine the consistency of the distribution of CNA and mRNA. All analyses were performed in R 4.0.1 (https://www.r-project.org/, accessed on 1 July 2022). The R package "RCircos" was used to generate the circle graphs. GraphPad Prism 6 (https://www.graphpad.com/scientific-software/prism/, accessed on 1 July 2022) was adopted to plot the survival curves. Two-sided tests were used for all analyses. A p value less than 0.05 was considered statistically significant.

3. Results

3.1. Patient Characteristics

A total of 2295 patients with early-stage breast cancer were enrolled in the study, including 1427 in the METABRIC cohort, 837 in the TCGA cohort, and 31 in the CICAMS cohort. The mean age of the patients in the METABRIC cohort was 60.68 years (standard deviation: 12.97 years). A total of 496 (34.76%) patients in the METABRIC cohort were stage I, 816 (57.18%) were stage II, 115 (8.06%) were stage III, 942 (66.01%) received radiotherapy, 1034 (72.46%) received drug therapy, and 818 (57.32%) received mastectomy. The mean age of patients in the TCGA cohort was 58.65 years (standard deviation: 13.18 years). A total of 147 (17.56%) patients in the TCGA cohort were stage I, 494 (59.02%) were stage II, and 196 (23.42%) were stage III. A total of 439 (52.45%) patients in the TCGA cohort received radiotherapy, 65 (7.77%) received drug therapy, and 382 (45.64%) received mastectomy. The mean age of patients in the CICAMS cohort was 53.44 years (standard deviation: 10.02 years). Of these, 16 (51.61%) were stage II, 15 (48.39%) were stage III, 13 (41.94%) received radiotherapy, 27 (87.1%) received medication, and all received mastectomy (Table 1).

Table 1. Clinical characteristics and treatment strategies.

	METABRIC		TCGA		CICAMS	
	Cases	**%**	**Cases**	**%**	**Cases**	**%**
Overall	1427		837		31	
Age						
<60	647	45.34%	445	53.17%	22	70.97%
≥60	780	54.66%	392	46.83%	9	29.03%
Subtype						
HoR [†] +HER2 [‡] −	1025	71.83%	527	62.96%	15	48.39%
HER2+	177	12.4%	174	20.79%	14	45.16%
TNBC [§]	225	15.77%	136	16.25%	2	6.45%
Grade						
I	116	8.13%			0	0%
II	549	38.47%			17	54.84%
III	715	50.11%			14	45.16%
Stage						
I	496	34.76%	147	17.56%	0	0%
II	816	57.18%	494	59.02%	16	51.61%
III	115	8.06%	196	23.42%	15	48.39%
Radiotherapy						
Yes	942	66.01%	439	52.45%	13	41.94%
No	485	33.99%	398	47.55%	18	58.06%
Drug therapy						
Yes	1034	72.46%	65	7.77%	27	87.1%
No	393	27.54%	772	92.23%	4	12.9%
Chemotherapy						
Yes	308	21.58%			24	77.42%
No	1119	78.42%			7	22.58%
Hormone Therapy						
Yes	875	61.32%			14	45.16%
No	552	38.68%			17	54.84%
Surgery						
Mastectomy	818	57.32%	382	45.64%	31	100%
Lumpectomy	609	42.68%	194	23.18%	0	0%
Unknown						

[†] HoR, hormone receptor; [‡] HER2, human epidermal growth factor receptor 2; [§] TNBC, triple-negative breast cancer.

In addition, the median CNABs of patients in the METABRIC, TCGA, and CICAMS databases were 2.1, 3.8, and 1.5 CNA/Mbp, respectively (Figure 1a). The genes with more than 50% CNA in patients from METABRIC (Figure 1b), TCGA (Figure 1c), and CICAMS

(Figure 1d) are presented using circle plots. The percentage of CNA per chromosome for these three cohorts is also presented in bar graphs (Figure 1e–g).
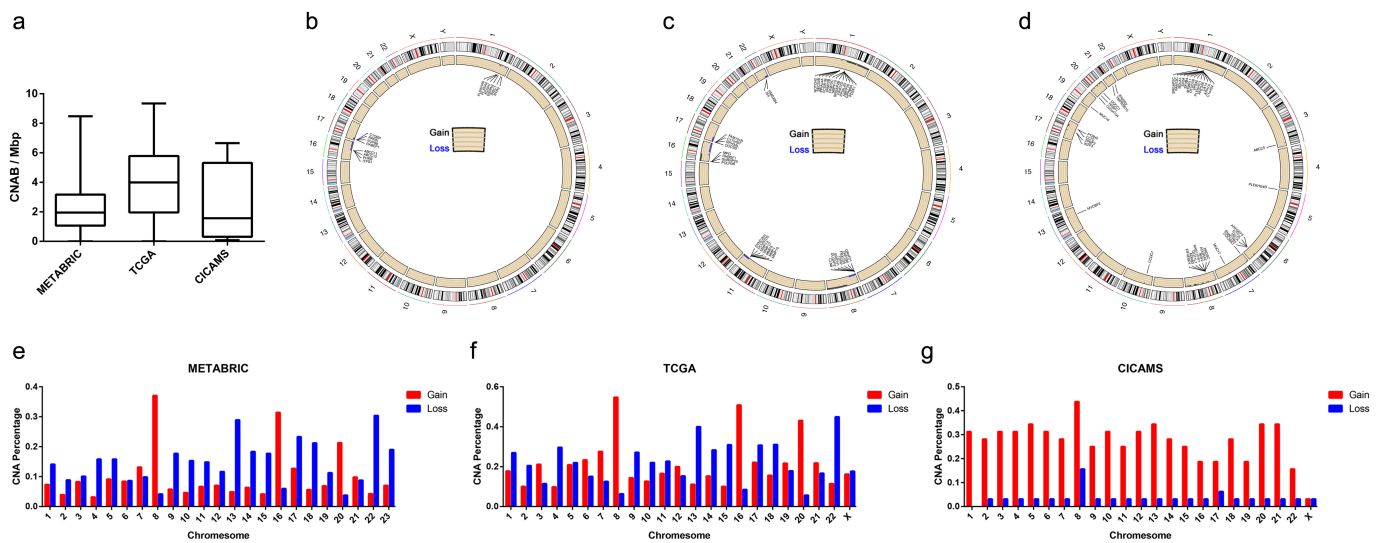

Figure 1. Copy number alteration of the study population. (**a**) Copy number alteration burden of three breast cancer cohorts; (**b**–**d**) the genes with more than 50% copy number alteration in patients from METABRIC, TCGA, and CICAMS; (**e**–**g**) the percentage of CNA per chromosome from METABRIC, TCGA, and CICAMS.

3.2. Differences in Clinical Characteristics between Different CNAB Subgroups

Patients in METABRIC were divided into a training cohort ($n = 714$) and a test cohort ($n = 713$). Receiver operating curve (ROC) analysis was performed in the METABRIC training cohort with 5-year OS, 10-year OS, 5-year RFS, and 10-year RFS as the endpoints. The 5-year RFS had the highest AUC value (Supplementary Figure S1, AUC = 0.611) when the CNAB threshold was 2.2 CNA/Mbp. Therefore, we divided the cohorts into a CNAB-high group and a CNAB-low group according to whether the CNAB was higher than 2.2 CNA/Mbp.

The differences in clinical characteristics between the CNAB-high group and CNAB-low group of the METABRIC cohort are shown in Supplementary Table S1. In the test cohort, there were 341 patients with hormone receptor (HoR)+ human epidermal growth factor receptor 2 (HER2)-, 34 with HER2+, and 43 with triple-negative breast cancer (TNBC) in the CNAB-low group, and 169 with HoR+HER2−, 58 with HER2+, and 68 with TNBC in the CNAB-high group. The differences between the two groups were statistically significant ($p < 0.001$). The CNAB-low group had a lower grade than the CNAB-high group ($p < 0.001$).

The differences between the CNAB-high and CNAB-low groups in the TCGA cohort are shown in Supplementary Table S2. There were 213 (84.86%) patients in the CNAB-low group with a molecular subtype of HoR+HER2−, higher than in the CNAB-high group ($p < 0.001$). The CNAB-low group had a lower stage than the CNAB-high group ($p = 0.006$).

3.3. Survival Analysis between CNAB Groups

In the METABRIC test cohort, the CNAB-high group had a worse prognosis than the CNAB-low group, with both shorter RFS and shorter OS (Figure 2a,b). After adjusting for age, subtype, grade, stage, and treatment strategies, the multivariate Cox regression model showed a 33% (RFS) and 25% (OS) higher risk for the CNAB-high group compared to the CNAB-low group (Supplementary Table S3). Similarly, in the TCGA cohort, the CNAB-high group showed a worse prognosis than the CNAB-low group in terms of both RFS and OS (Figure 2c,d). Adjusted by age, stage, and treatment strategy, the CNAB-high group had shorter RFS (HR = 1.62, 95% CI: 1.08–2.46, $p = 0.021$) and shorter OS (HR = 1.94, 95% CI: 1.15–3.28, $p = 0.013$) than the CNAB-low group (Supplementary Table S4). In the

CICAMS cohort, the CNAB-high group had a worse prognosis for RFS than the CNAB-low group (HR = 6.28, 95% CI: 1.08–36.84, *p* = 0.042) (Figure 2e).
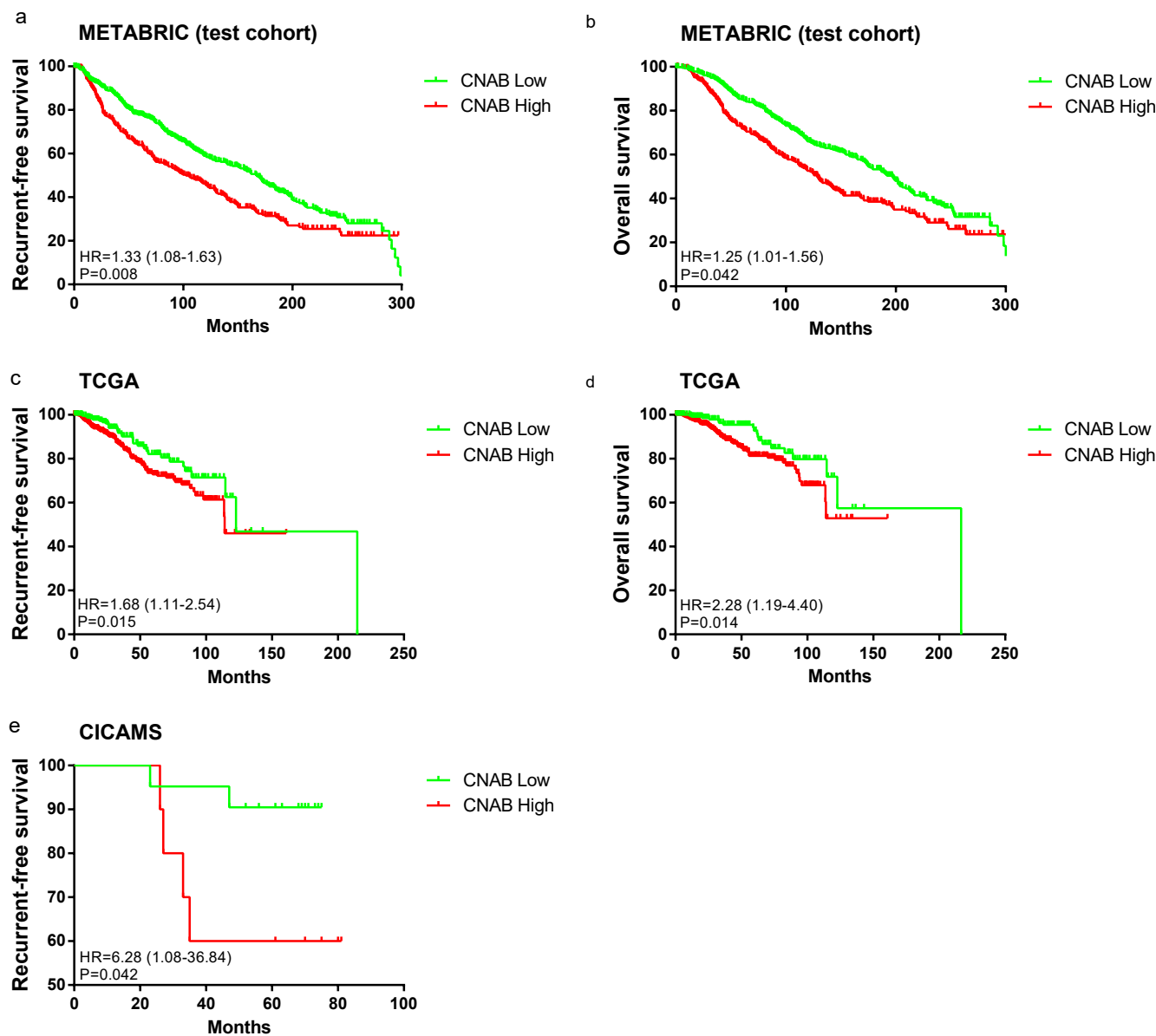

Figure 2. Survival analysis on relapse-free survival and overall survival of patients in the (**a**,**b**) METABRIC test cohort, (**c**,**d**) TCGA, and (**e**) CICAMS between the CNAB-high group and CNAB-low group.

Furthermore, we used different cutoff values to dichotomize CNAB, and the cutoff was taken as 2.2 or 2.6 CNA/Mbp as a valid predictor of prognosis in both the METABRIC test cohort and TCGA cohort. When the threshold is below 2.2 or above 2.6 CNA/Mbp, there will no longer be a statistically significant difference in RFS and OS in the CNAB-high group compared to the CNAB-low group (Supplementary Table S5). We also performed survival analyses for different subgroups (Supplementary Tables S6 and S7). Some subgroups no longer had statistically significant survival differences due to the reduction in the number of events after subgrouping.

3.4. Combined Survival Analysis of TMB and CNAB

We performed a combined survival analysis of TMB and CNAB in the TCGA cohort. TMB was higher in the CNAB-high group according to the Wilcoxon test (*p* < 0.001)

(Supplementary Figure S2). The TCGA cohort was divided into four groups according to the quartiles of TMB. Survival analysis showed no prognostic value for RFS or OS (Supplementary Figure S3). Then, the cohort was divided into high and low groups based on the median TMB. Combined survival analysis with CNAB showed that the TMB-high/CNAB-low group had the best prognosis for OS, which was statistically significant (p = 0.039) (Supplementary Figure S3).

3.5. Construction and Validation of the CNAB11 Model

In the METABRIC training cohort, chi-squared tests were performed for whole-exon CNAs between the 5-year relapse group and the 5-year nonrelapse group, with a total of 449 genes selected for corrected p values < 0.05 (Supplementary Table S8). Meanwhile, we conducted a concordance test of CNA and mRNA for each patient in the METABRIC training cohort, with a total of 726 genes screened for the Kolmogorov–Smirnov test p < 0.05 (Supplementary Table S9). We took the intersection of these two gene sets to obtain 11 genes (CBWD1 (9p24.3), CDC6 (17q21.2), CWC25 (17q12), HS3ST3A1 (17p12), IFNA2 (9p21.3), KDM4C (9p24.1), KRT27 (17q21.2), MLLT6 (17q12), NBR1 (17q21.31), NBR2 (17q21.31), and ZDHHC21 (9p22.3)). We summed the number of copy number changes for these 11 genes to obtain the CNAB11 score. In the METABRIC training cohort, we used ROC to conclude that the cutoff should be taken as 6.

We divided the METABRIC, TCGA, and CICAMS cohorts into high and low groups based on the CNAB11 score. Survival analysis showed that in the METABRIC test cohort, the CNAB11-high group had a significantly worse prognosis than the CNAB11-low group (RFS: HR = 1.35, 95% CI: 1.09–1.66, p = 0.005; OS: HR = 1.25, 95% CI: 1.01–1.56, p = 0.044). Similarly, in the TCGA cohort, the prognosis was worse in the CNAB11-high group than in the CNAB11-low group (RFS: HR = 1.55, 95% CI: 1.08–2.23, p = 0.017; OS: HR = 2.59, 95% CI: 1.46–4.59, p = 0.001). In the CICAMS cohort, the CNAB11-high group had a shorter RFS than the CNAB11-low group (HR = 5.94, 95% CI: 1.08–32.72, p = 0.017) (Figure 3).

Then, we built the nomogram in the METABRIC training cohort based on a multivariate Cox regression model (Figure 4a). In the TCGA cohort and the CICAMS cohort, we calculated the score for each patient individually based on the nomogram. The 5-year RFS was then predicted for both cohorts. ROC showed that the nomogram was a good predictor of 5-year RFS in patients in both the TCGA cohort (AUC = 0.72) and the CICAMS cohort (AUC = 0.83) (Figure 4b).

3.6. Differences in CNA and Expression Profiles of Different CNAB11 Groups

We analyzed the CNA differences between the CNAB11 high and CNAB11 low populations using Fisher's test. For the TCGA cohort, the differential CNA between the CNAB11-high and CNAB11-low populations was distributed on almost all chromosomes, but was most significant on chromosomes 6, 9, 17, and 20 (Figure 5a). For the CICAMS cohort, differential CNA was scattered on most chromosomes, but most densely on chromosome 14 (Figure 5b).

We analyzed the expression profiles of the CNAB11-high group and CNAB11-low group in the TCGA cohort to screen for differentially expressed genes (Supplementary Table S10). The heatmap (Figure 6a) and volcano map (Supplementary Figure S4) are shown. We further analyzed the differentially expressed genes by Gene Ontology (GO), and the results showed that they were enriched in 33 GO terms, such as "positive regulation of establishment of protein localization to telomere" (Figure 6b).
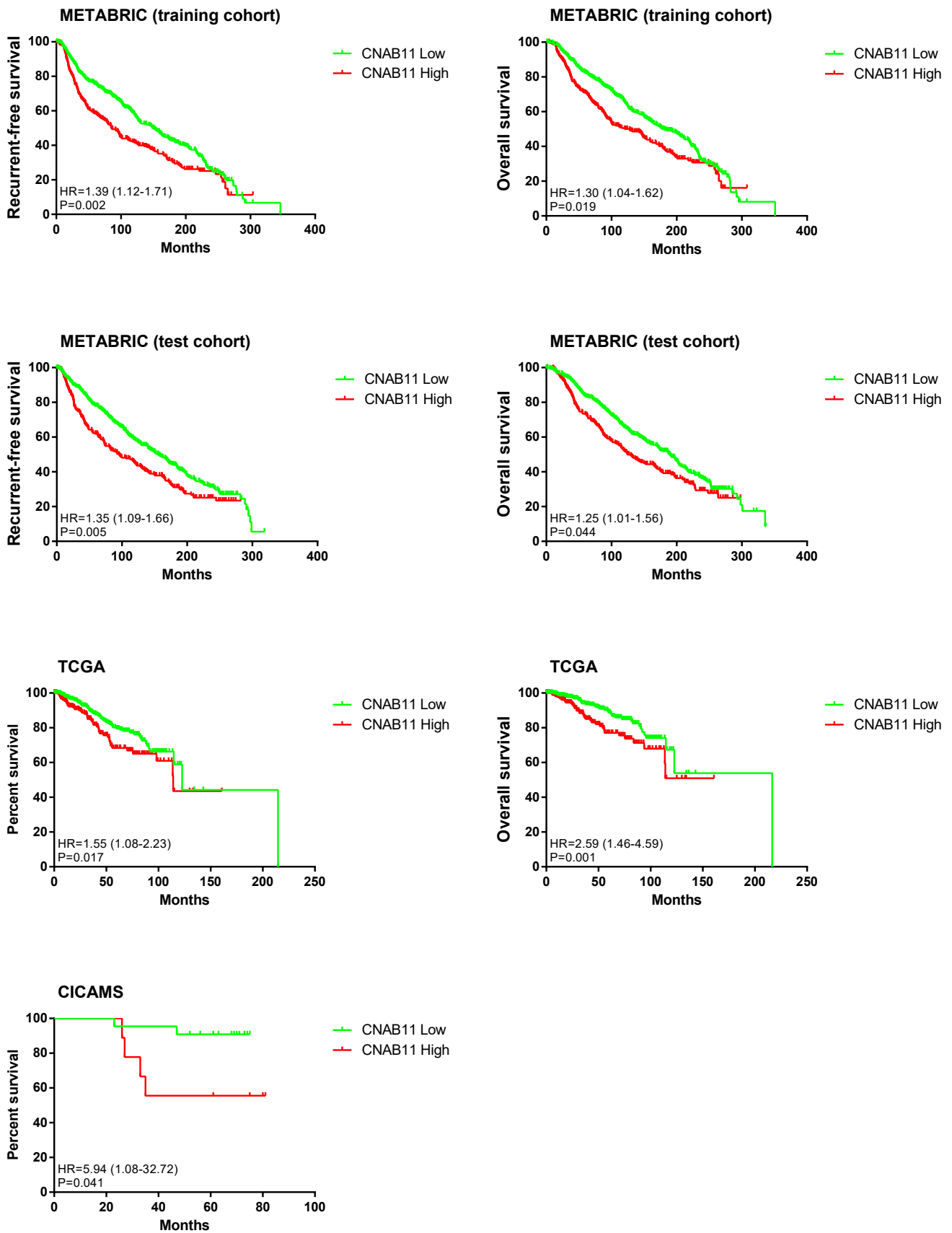
Figure 3. Survival analysis on relapse-free survival and overall survival of patients in METABRIC, TCGA, and CICAMS between the CNAB11-high group and CNAB11-low group.
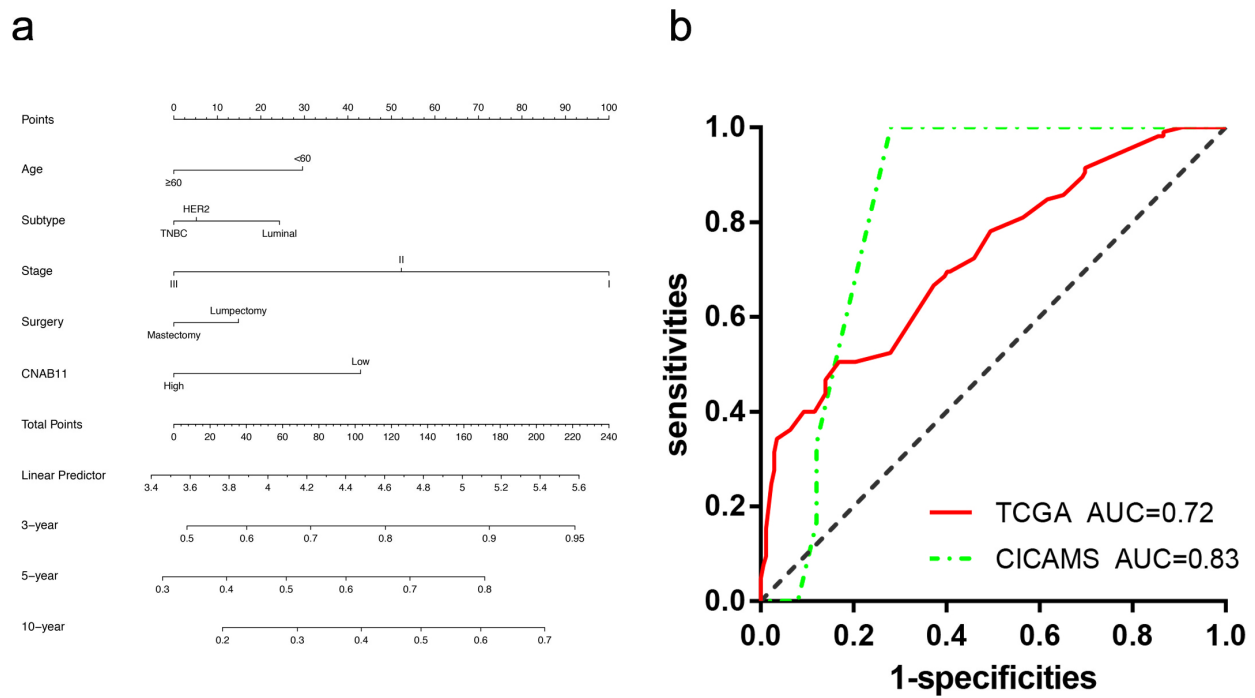
Figure 4. (**a**) Nomogram of early-stage breast cancer based on CNAB11 score and clinical characteristics. (**b**) Receiver operating curve of 5-year relapse-free survival in TCGA and CICAMS.
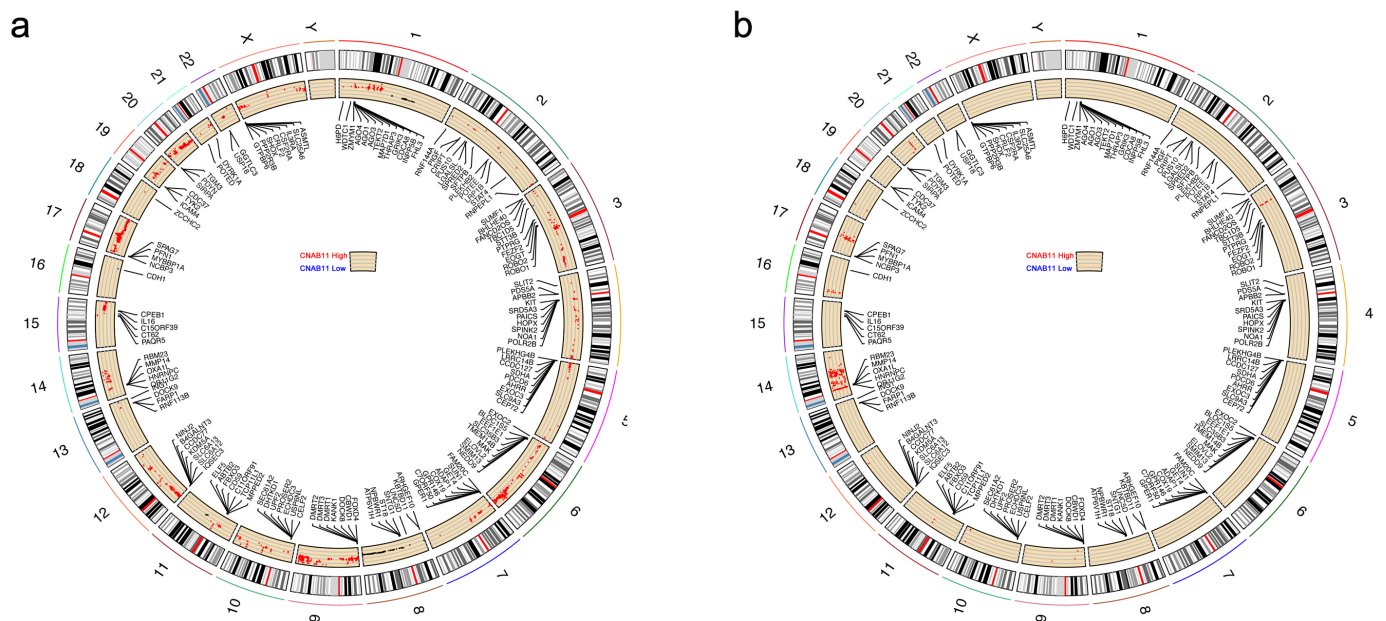

Figure 5. The differential CNA between the CNAB11-high and CNAB11-low populations in (**a**) the TCGA cohort and (**b**) the CICAMS cohort.
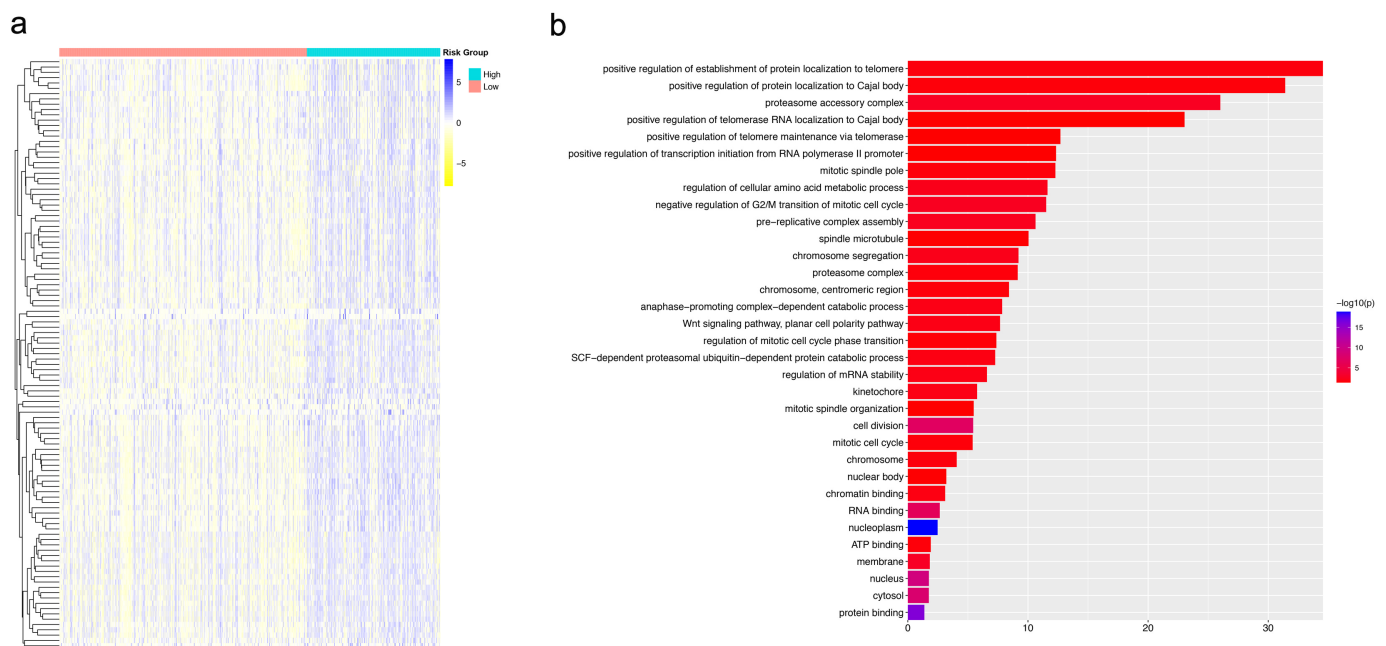
Figure 6. (**a**) Heatmap of the expression profile difference between the CNAB11-high group and the CNAB11-low group. (**b**) Gene ontology analysis between the CNAB11-high group and the CNAB11-low group.

4. Discussion

With the above results, it is evident that CNAB has better prognostic prediction than TMB for early breast cancer patients. Furthermore, our simplified CNAB11 model is similar to the CNAB model in that both have good predictive effects. The METABRIC cohort is dominated by loss, while CICAMS is dominated by gain. However, CNAB is a metric that homogenizes METABRIC and CICAMS well. The CNAB and CNAB11 models built by the METABRIC cohort can be used for the CICAMS cohort, even though they have different CNA types. Therefore, we believe that the CNAB-based model has generalizability across different racial cohorts.

Several studies have confirmed the close association between somatic copy number alterations in some genes and the prognosis of patients [15–17]. In this paper, we also found prognosis-associated CNAs in breast cancer. Notably, the genes in CNAB11 were mainly distributed at 17q and 9p. This is consistent with the findings of Budczies and Ueno [11,12]. Therefore, in the future, if histological specimens are used for prognostic evaluation, fluorescence in situ hybridization can be considered for the corresponding regions. Since ctDNA tends to be fragmented, we believe that it may be more accurate to use the corresponding 11 probes for detection if this model is to be used for liquid biopsies in the future.

Somatic cell copy number alteration burden has also been demonstrated as a new prognostic marker in many cancer types [2,6,8,10,14,18,23–25]. Few investigators have focused on the predictive role of copy number variant burden on the prognosis of patients with early-stage breast cancer. In fact, CNAB has more predictive potential for prognosis than TMB due to the higher rate of copy number alterations than mutations in breast cancer. Therefore, we believe that CNAB-based prognostic models should be investigated and explored by more researchers.

For early-stage breast cancer, there are several widely accepted prognostic models based on gene expression profiles, such as Oncotype DX. The TAILORx study showed that Oncotype DX can accurately predict survival in ER (+), HER2 (−), and axillary lymph node-negative breast cancer, thus guiding the choice for chemotherapy [26]. The expression levels of genes are continuous variables that are more suitable for accurate modeling than mutations or copy number alterations. With the improvement of DNA sequencing technology

in recent years, blood tissue-based ctDNA testing has become increasingly mature [27,28]. Wang et al. concluded that the DNA sequencing results of tumor tissue specimens and blood specimens are similar [29]. Therefore, prognostic models based on TMB or CNAB are expected to be tested noninvasively with ctDNA in the future. Surgical resection specimens are difficult to obtain for both early-stage breast cancer patients receiving neoadjuvant therapy and advanced breast cancer patients, at which point, prognosis-specific CNA by blood ctDNA is an excellent option. Thus, we suggest that CNAB-based prognostic models are very promising in all cancer types. In addition, DNA sequencing-based models can be used in conjunction with transcriptome-based models. Both models complement each other and may play an increasingly important role in the prognosis of early breast cancer patients in the future.

However, the cost of whole-exome sequencing is becoming increasingly low. Nevertheless, for some patients, it is not appropriate to perform whole-exome sequencing for prognostic assessment. Therefore, we simplified the CNAB model and selected 11 promising genes. Kong and Mahadevappa demonstrated experimentally that CDC6 plays an important role in the progression of breast cancer and can help predict the prognosis of breast cancer patients [30,31]. Several studies also found that HS3ST3A1 was a novel tumor regulator and a good predictor of survival in both lung cancer patients and breast cancer patients [32,33]. In breast cancer, glioma, and colorectal cancer, KDM4C is involved in biological processes such as tumorigenesis and metastasis [34–36]. MLLT6 has been confirmed to play an important role in immune maintenance [37]. NBR1 and NBR2 are closely related to the regulation of BRCA1 and have important roles in the genesis and progression of many tumors [38]. Few studies have shown the prognostic value of CBWD1, CWC25, IFNA2, KRT27, or ZDHHC21 in breast cancer, and subsequent studies are expected. There are two potential limitations to this article. First, the number of patients in CICAMS was relatively small, which may have resulted in potential selection bias. Second, the 11CNAB model we constructed was not validated by further animal experiments, and we will follow up with related research.

5. Conclusions

For early-stage breast cancer, CNAB is a better prognostic factor than TMB and has shown great results in the European-dominated METABRIC cohort, the American-dominated TCGA cohort, and the CICAMS cohort. We subsequently constructed an 11-gene CNA-based prognostic model. This model was shown to have promising prognostic indications in both the TCGA cohort and the CICAMS cohort.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers14174145/s1, Supplementary Figure S1: Receiver operating characteristic curve of the 5-year relapse-free survival of patients in the METABRIC cohort. Supplementary Figure S2: Tumor mutational burden of patients in different copy number alteration burden groups. Supplementary Figure S3: Survival analysis for different tumor mutational burden groups and combined survival analysis of tumor mutational burden and copy number alteration burden of patients in TCGA cohort. Supplementary Figure S4: Volcano plot of the expression profile difference between the CNAB11 high group and the CNAB11 low group. Supplementary Table S1: Clinical characteristics differences between the high and low CNAB groups of the METABRIC cohort. Supplementary Table S2: Baseline differences between the high and low CNAB groups of TCGA cohort. Supplementary Table S3: Univariate and multivariate Cox regression model on METABRIC test cohort. Supplementary Table S4: Univariate and multivariate Cox regression model on TCGA cohort. Supplementary Table S5: Sensitivity Analysis. Supplementary Table S6: Subgroup survival analysis on RFS. Supplementary Table S7: Subgroup survival analysis on OS. Supplementary Table S8: Chi-Square test of METABRIC training cohort between relapse group and non-relapse group. Supplementary Table S9: KS test of METABRIC cohort between CNA and mRNA z score. Supplementary Table S10: DEGs between CNAB11 high group and CNAB11 low group. Supplementary Method: DNA Extraction and Targeted Next-Generation Sequencing.

References

1. Zhang, S.; Sun, K.; Zheng, R.; Zeng, H.; Wang, S.; Chen, R.; Wei, W.; He, J. Cancer incidence and mortality in China, 2015. *J. Natl. Cancer Cent.* **2021**, *1*, 2–11. [CrossRef]
2. Zhang, L.; Feizi, N.; Chi, C.; Hu, P. Association Analysis of Somatic Copy Number Alteration Burden With Breast Cancer Survival. *Front. Genet.* **2018**, *9*, 421. [CrossRef] [PubMed]
3. Glas, A.M.; Floore, A.; Delahaye, L.J.M.J.; Witteveen, A.T.; Pover, R.C.F.; Bakx, N.; Lahti-Domenici, J.S.T.; Bruinsma, T.J.; Warmoes, M.O.; Bernards, R.; et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics.* **2006**, *7*, 278. [CrossRef]
4. Cobleigh, M.A.; Tabesh, B.; Bitterman, P.; Baker, J.; Cronin, M.; Liu, M.-L.; Borchik, R.; Mosquera, J.-M.; Walker, M.G.; Shak, S. Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. *Clin. Cancer Res.* **2005**, *11 Pt 1*, 8623–8631. [CrossRef] [PubMed]
5. Franch-Expósito, S.; Bassaganyas, L.; Vila-Casadesús, M.; Hernández-Illán, E.; Esteban-Fabró, R.; Díaz-Gay, M.; Lozano, J.J.; Castells, A.; Llovet, J.M.; Castellví-Bel, S.; et al. CNApp, a tool for the quantification of copy number alterations and integrative analysis revealing clinical implications. *eLife* **2020**, *9*, e50267. [CrossRef] [PubMed]
6. Hieronymus, H.; Murali, R.; Tin, A.; Yadav, K.; Abida, W.; Moller, H.; Berney, D.; Scher, H.; Carver, B.; Scardino, P.; et al. Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. *eLife* **2018**, *7*, e37294. [CrossRef] [PubMed]
7. Sansregret, L.; Vanhaesebroeck, B.; Swanton, C. Determinants and clinical implications of chromosomal instability in cancer. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 139–150. [CrossRef]
8. Smith, J.C.; Sheltzer, J.M. Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. *eLife* **2018**, *7*, e39217. [CrossRef]
9. Stopsack, K.H.; Whittaker, C.A.; Gerke, T.A.; Loda, M.; Kantoff, P.W.; Mucci, L.A.; Amon, A. Aneuploidy drives lethal progression in prostate cancer. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 11390–11395. [CrossRef]
10. Jin, X.; Yan, J.; Chen, C.; Chen, Y.; Huang, W.-K. Integrated Analysis of Copy Number Variation, Microsatellite Instability, and Tumor Mutation Burden Identifies an 11-Gene Signature Predicting Survival in Breast Cancer. *Front. Cell Dev. Biol.* **2021**, *9*, 721505. [CrossRef]
11. Ueno, T.; Emi, M.; Sato, H.; Ito, N.; Muta, M.; Kuroi, K.; Toi, M. Genome-wide copy number analysis in primary breast cancer. *Expert Opin. Ther. Targets* **2012**, *16* (Suppl. 1), S31–S35. [CrossRef] [PubMed]
12. Budczies, J.; Denkert, C.; Győrffy, B.; Schirmacher, P.; Stenzinger, A. Chromosome 9p copy number gains involving PD-L1 are associated with a specific proliferation and immune-modulating gene expression program active across major cancer types. *BMC Med. Genomics* **2017**, *10*, 74. [CrossRef] [PubMed]
13. Ciriello, G.; Gatza, M.L.; Beck, A.H.; Wilkerson, M.D.; Rhie, S.K.; Pastore, A.; Zhang, H.; McLellan, M.; Yau, C.; Kandoth, C.; et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell.* **2015**, *163*, 506–519. [CrossRef] [PubMed]

14. Liu, L.; Bai, X.; Wang, J.; Tang, X.-R.; Wu, D.-H.; Du, S.-S.; Du, X.-J.; Zhang, Y.-W.; Zhu, H.-B.; Fang, Y.; et al. Combination of TMB and CNA Stratifies Prognostic and Predictive Responses to Immunotherapy Across Metastatic Cancer. *Clin. Cancer Res.* **2019**, *25*, 7413–7423. [CrossRef] [PubMed]

15. Liang, L.; Fang, J.-Y.; Xu, J. Gastric cancer and gene copy number variation: Emerging cancer drivers for targeted therapy. *Oncogene* **2016**, *35*, 1475–1482. [CrossRef] [PubMed]

16. Wang, H.; Liang, L.; Fang, J.-Y.; Xu, J. Somatic gene copy number alterations in colorectal cancer: New quest for cancer drivers and biomarkers. *Oncogene* **2016**, *35*, 2011–2019. [CrossRef] [PubMed]

17. Nibourel, O.; Guihard, S.; Roumier, C.; Pottier, N.; Terre, C.; Paquet, A.; Peyrouze, P.; Geffroy, S.; Quentin, S.; Alberdi, A.; et al. Copy-number analysis identified new prognostic marker in acute myeloid leukemia. *Leukemia* **2017**, *31*, 555–564. [CrossRef] [PubMed]

18. Pladsen, A.V.; Nilsen, G.; Rueda, O.M.; Aure, M.R.; Borgan, Ø.; Liestøl, K.; Vitelli, V.; Frigessi, A.; Langerød, A.; Mathelier, A.; et al. DNA copy number motifs are strong and independent predictors of survival in breast cancer. *Commun. Biol.* **2020**, *3*, 153. [CrossRef]

19. The American Joint Committee on Cancer. *AJCC Cancer Staging Manual*, 7th ed.; Springer: New York, NY, USA, 2010.

20. WHO. *World Organization of Classification of Tumours: Pathology and Genetics of Tumors of the Breast and Female Genital Organs*, 3rd ed.; IARC Press: Lyon, France, 2003.

21. Beroukhim, R.; Getz, G.; Nghiemphu, L.; Barretina, J.; Hsueh, T.; Linhart, D.; Vivanco, I.; Lee, J.C.; Huang, J.H.; Alexander, S.; et al. Assessing the significance of chromosomal aberrations in cancer: Methodology and application to glioma. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20007–20012. [CrossRef]

22. Tolaney, S.M.; Garrett-Mayer, E.; White, J.; Blinder, V.S.; Foster, J.C.; Amiri-Kordestani, L.; Hwang, E.S.; Bliss, J.M.; Rakovitch, E.; Perlmutter, J.; et al. Updated Standardized Definitions for Efficacy End Points (STEEP) in Adjuvant Breast Cancer Clinical Trials: STEEP Version 2.0. *J. Clin. Oncol.* **2021**, *39*, 2720–2731. [CrossRef]

23. Pariyar, M.; Johns, A.; Thorne, R.F.; Scott, R.J.; Avery-Kiejda, K.A. Copy number variation in triple negative breast cancer samples associated with lymph node metastasis. *Neoplasia* **2021**, *23*, 743–753. [CrossRef] [PubMed]

24. Schrank, T.P.; Lenze, N.; Landess, L.P.; Hoyle, A.; Parker, J.; Lal, A.; Sheth, S.; Chera, B.S.; Patel, S.N.; Hackman, T.G.; et al. Genomic heterogeneity and copy number variant burden are associated with poor recurrence-free survival and 11q loss in human papillomavirus-positive squamous cell carcinoma of the oropharynx. *Cancer* **2021**, *127*, 2788–2800. [CrossRef] [PubMed]

25. Bassaganyas, L.; Pinyol, R.; Esteban-Fabró, R.; Torrens, L.; Torrecilla, S.; Willoughby, C.E.; Franch-Expósito, S.; Vila-Casadesús, M.; Salaverria, I.; Montal, R.; et al. Copy-Number Alteration Burden Differentially Impacts Immune Profiles and Molecular Features of Hepatocellular Carcinoma. *Clin. Cancer Res.* **2020**, *26*, 6350–6361. [CrossRef] [PubMed]

26. Sparano, J.A.; Gray, R.J.; Makower, D.F.; Pritchard, K.I.; Albain, K.S.; Hayes, D.F.; Geyer, C.E.; Dees, E.C.; Perez, E.A.; Olson, J.A.; et al. Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *New Engl. J. Med.* **2015**, *373*, 2005–2014. [CrossRef]

27. Pessoa, L.S.; Heringer, M.; Ferrer, V.P. ctDNA as a cancer biomarker: A broad overview. *Crit. Rev. Oncol. /Hematol.* **2020**, *155*, 103109. [CrossRef]

28. Clatot, F. Review ctDNA and Breast Cancer. *Recent Results Cancer Res.* **2020**, *215*, 231–252. [CrossRef]

29. Wang, Z.; Duan, J.; Cai, S.; Han, M.; Dong, H.; Zhao, J.; Zhu, B.; Wang, S.; Zhuo, M.; Sun, J.; et al. Assessment of Blood Tumor Mutational Burden as a Potential Biomarker for Immunotherapy in Patients With Non-Small Cell Lung Cancer With Use of a Next-Generation Sequencing Cancer Gene Panel. *JAMA Oncol.* **2019**, *5*, 696–702. [CrossRef]

30. Kong, X.; Duan, Y.; Sang, Y.; Li, Y.; Zhang, H.; Liang, Y.; Liu, Y.; Zhang, N.; Yang, Q. LncRNA-CDC6 promotes breast cancer progression and function as ceRNA to target CDC6 by sponging microRNA-215. *J. Cell. Physiol.* **2019**, *234*, 9105–9117. [CrossRef]

31. Mahadevappa, R.; Neves, H.; Yuen, S.M.; Bai, Y.; McCrudden, C.M.; Yuen, H.F.; Wen, Q.; Zhang, S.-D.; Kwok, H.F. The prognostic significance of Cdc6 and Cdt1 in breast cancer. *Sci. Rep.* **2017**, *7*, 985. [CrossRef]

32. Nakano, T.; Shimizu, K.; Kawashima, O.; Kamiyoshihara, M.; Kakegawa, S.; Sugano, M.; Ibe, T.; Nagashima, T.; Kaira, K.; Sunaga, N.; et al. Establishment of a human lung cancer cell line with high metastatic potential to multiple organs: Gene expression associated with metastatic potential in human lung cancer. *Oncol. Rep.* **2012**, *28*, 1727–1735. [CrossRef]

33. Mao, X.; Gauche, C.; Coughtrie, M.; Bui, C.; Gulberti, S.; Merhi-Soussi, F.; Ramalanjaona, N.; Bertin-Jung, I.; Diot, A.; Dumas, D.; et al. The heparan sulfate sulfotransferase 3-OST3A (HS3ST3A) is a novel tumor regulator and a prognostic marker in breast cancer. *Oncogene* **2016**, *35*, 5043–5055. [CrossRef] [PubMed]

34. Garcia, J.; Lizcano, F. KDM4C Activity Modulates Cell Proliferation and Chromosome Segregation in Triple-Negative Breast Cancer. *Breast Cancer Basic Clin. Res.* **2016**, *10*, 169–175. [CrossRef] [PubMed]

35. Wu, X.; Li, R.; Song, Q.; Zhang, C.; Jia, R.; Han, Z.; Zhou, L.; Sui, H.; Liu, X.; Zhu, H.; et al. JMJD2C promotes colorectal cancer metastasis via regulating histone methylation of MALAT1 promoter and enhancing β-catenin signaling pathway. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 435. [CrossRef] [PubMed]

36. Chen, Y.; Fang, R.; Yue, C.; Chang, G.; Li, P.; Guo, Q.; Wang, J.; Zhou, A.; Zhang, S.; Fuller, G.N.; et al. Wnt-Induced Stabilization of KDM4C Is Required for Wnt/β-Catenin Target Gene Expression and Glioblastoma Tumorigenesis. *Cancer Res.* **2020**, *80*, 1049–1063. [CrossRef]

37. Sreevalsan, S.; Döring, M.; Paszkowski-Rogacz, M.; Brux, M.; Blanck, C.; Meyer, M.; Momburg, F.; Buchholz, F.; Theis, M. MLLT6 maintains PD-L1 expression and mediates tumor immune resistance. *EMBO Rep.* **2020**, *21*, e50155. [CrossRef]

38. Marsh, T.; Debnath, J. Autophagy suppresses breast cancer metastasis by degrading NBR1. *Autophagy* **2020**, *16*, 1164–1165. [CrossRef]