

RESEARCH

Open Access



# Machine learning unveils key Redox signatures for enhanced breast Cancer therapy

Tao Wang<sup>1</sup>, Shu Wang<sup>2</sup>, Zhuolin Li<sup>2</sup>, Jie Xie<sup>2</sup>, Kuiying Du<sup>2\*</sup> and Jing Hou<sup>2\*</sup>

## Abstract

**Background** Breast cancer remains a leading cause of mortality among women worldwide, necessitating innovative prognostic models to enhance treatment strategies.

**Methods** Our study retrospectively enrolled 9,439 breast cancer patients from 12 independent datasets and single-cell data from 12 patients (64,308 cells). Moreover, 30 in-house clinical cohort were collected for validation. We employed a comprehensive approach by combining ten distinct machine learning algorithms across 108 different combinations to scrutinize 88 pre-existing signatures of breast cancer. To affirm the efficacy of our developed model, immunohistochemistry assays were performed. Additionally, we investigated various potential immunotherapeutic and chemotherapeutic interventions.

**Results** This study introduces an Artificial Intelligence-aided Redox Signature (AIARS) as a novel prognostic tool, leveraging machine learning to identify critical redox-related gene signatures in breast cancer. Our results demonstrate that AIARS significantly outperforms existing prognostic models in predicting breast cancer outcomes, offering a robust tool for personalized treatment planning. Validation through immunohistochemistry assays on samples from 30 patients corroborated our results, underscoring the model's applicability on a wider scale. Furthermore, the analysis revealed that patients with low AIARS expression levels are more responsive to immunotherapy. Conversely, those exhibiting high AIARS were found to be more susceptible to certain chemotherapeutic agents, including vincristine.

**Conclusions** Our study underscores the importance of redox biology in breast cancer prognosis and introduces a powerful machine learning-based tool, the AIARS, for personalized treatment strategies. By providing a more nuanced understanding of the redox landscape in breast cancer, the AIARS paves the way for the development of redox-targeted therapies, promising to enhance patient outcomes significantly. Future work will focus on clinical validation and exploring the mechanistic roles of identified genes in cancer biology.

**Keywords** Breast cancer, Machine learning, Redox biology, Gene signatures, Personalized therapy

\*Correspondence:  
Kuiying Du  
571058891@qq.com  
Jing Hou  
hjhlingtong@163.com

<sup>1</sup>Research Laboratory Center, Guizhou Provincial People's Hospital, Guiyang 550002, China

<sup>2</sup>Department of Breast Surgery, Guizhou Provincial People's Hospital, Guiyang 550002, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Introduction

Breast cancer remains a formidable challenge in oncology, with its complex etiology and varied patient outcomes underscoring the urgent need for innovative diagnostic and prognostic strategies [1]. Despite significant advances in understanding the disease's molecular underpinnings, the quest for personalized treatment approaches that can accurately predict individual responses and survival outcomes is ongoing [2]. Central to this challenge is the role of redox biology—a fundamental aspect of cellular physiology that, when dysregulated, profoundly influences cancer development, progression, and response to therapy [3].

Redox reactions are pivotal in maintaining cellular homeostasis, and their imbalance has been implicated in the pathophysiology of numerous diseases, including cancer [4]. In breast cancer, oxidative stress and altered redox signaling are increasingly recognized as key drivers of tumorigenesis, tumor aggressiveness, and resistance to standard therapies [5, 6]. This burgeoning field of research suggests that redox dysregulation not only contributes to the molecular heterogeneity of breast cancer but also holds the potential to unlock novel prognostic and therapeutic avenues [7].

However, the integration of redox-related molecular signatures into breast cancer prognosis models remains nascent. Current prognostic tools often rely on traditional clinical and pathological markers, which do not fully capture the molecular complexity of the disease or its dynamic interaction with redox processes. This gap in our prognostic capabilities limits our ability to tailor treatments to individual patients' molecular profiles, potentially leading to suboptimal outcomes.

Herein, we propose a novel approach that leverages machine learning to distill complex redox-related gene signatures into a predictive model for breast cancer prognosis. By harnessing the power of computational algorithms to analyze large-scale omics data, we aim to uncover the intricate relationships between redox dysregulation and breast cancer outcomes. This endeavor seeks to bridge the existing gap in our prognostic tools, offering a more nuanced understanding of breast cancer biology and paving the way for redox-targeted therapeutic strategies.

## Materials and methods

### Data acquisition

We retrospectively collected 12 independent breast cancer cohorts from The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO) and MetaGxData [8]. Samples with complete survival data were used for further analysis. Patients in this study were enrolled from multiple datasets with varying clinical histories, including those who had undergone treatments such as

chemotherapy or radiation therapy. However, detailed treatment histories, including information on prior radiation therapy, were not uniformly available across all datasets. As a result, we were unable to conduct a specific analysis of the relationship between AIARS and radiation therapy, which we acknowledge as a limitation of the current study. Overall, we enrolled 9,439 patients from 12 cohorts for prognostic analysis, TCGA-BRCA ( $n=1076$ ), GSE202203 ( $n=3206$ ), GSE1456 ( $n=159$ ), GSE 20,685 ( $n=327$ ), GSE96058 ( $n=3409$ ), GSE131769 ( $n=298$ ), GSE86166 ( $n=330$ ), GSE21653 ( $n=244$ ), GSE88770 ( $n=108$ ), GSE58812 ( $n=107$ ), GE20711 ( $n=88$ ) and PNC ( $n=87$ ).

### Identification of aberrant expressed redox genes in breast cancer

To isolate redox-related gene signatures, we employed a differential gene expression analysis. Genes involved in redox processes were curated from the GeneCards databases. Differential expression analysis was conducted using three independent datasets (TCGA-BRCA, GSE93601, and GSE76250) identifying genes with significant expression differences between normal and tumor tissues, with a focus on those implicated in redox regulation.

### Machine learning derived redox signature

To formulate a redox signature unique to breast cancer, we followed the methodology described by Liu et al. [9]. Our strategy involved the integration of ten different computational tools: Random Forest (RF), Least Absolute Shrinkage and Selection Operator (LASSO), Gradient Boosting Machine (GBM), Survival Support Vector Machine (Survival-SVM), Supervised Principal Component (SuperPC), Ridge Regression, Partial Least Squares Cox Regression (plsRcox), CoxBoost, Stepwise Cox regression, and Elastic Net (Enet). Particularly, tools like RF, LASSO, CoxBoost, and Stepwise Cox were utilized for their ability to reduce dimensionality and select relevant variables. These were synergized into 108 distinct combinations for the purpose of generating a predictive signature. Through the evaluation of all cohorts, including TCGA and five GEO datasets, we determined the most consistent prognostic model by calculating the average Concordance Index (C-index). This process led to the establishment of a redox-specific signature aimed at predicting outcomes in breast cancer.

### Genomic alteration analysis

To elucidate genetic variations between the two AIARS groups, we analyzed both genetic mutation levels and Copy Number Alterations (CNA) using the TCGA-BRCA database.

The Tumor Mutation Burden (TMB) of both high- and low AIARS breast cancer patients was derived from the raw mutation file. Employing the maftools landscape, we visualized the most frequently mutated genes (mutation rate > 5%). Moreover, patient-specific mutational signatures were obtained using the deconstructSigs package [10]. Notably, we highlighted four prominent mutational signatures (SBS3, SBS1, SBS12, SBS11) within the TCGA-BRCA dataset that exhibited increased mutation frequencies. We pinpointed the five most common regions of amplification and deletion and specifically spotlighted the four predominant genes in chromosomal regions 8q24.21 and 5q11.2.

### Single-cell data processing

To prepare the dataset for single-cell RNA sequencing analysis, we utilized Seurat (v4.0) to process the data contained within GSE161529 [11]. This involved the removal of genes that did not show any expression, focusing on those that exhibited nonzero expression levels. Normalization of the expression matrix was achieved through the application of Seurat's "SCTransform" function. For reducing the dimensionality of the dataset, PCA and UMAP techniques were applied. To identify distinct cellular groupings, we employed the "FindNeighbors" and "FindClusters" operations within Seurat. To enhance the integrity and reliability of our dataset, we used the DoubletFinder package to remove potential doublets [12]. Cells that did not adhere to the defined quality standards, such as having a mitochondrial gene content over 15% or containing fewer than 500 genes, were excluded. Through these rigorous quality control measures, 64,308 cells were maintained for subsequent examination. Cell types were identified by manually annotating them based on the presence of established marker genes.

### Inference of regulons and their activity

In our study, we adapted the Single-Cell rEgulatory Network Inference (SCENIC) methodology to construct gene regulatory networks (GRNs) from single-cell RNA sequencing data [13]. SCENIC involves a three-step process: initially, it identifies co-expression modules between transcription factors (TFs) and their potential target genes. In the next step, for each module, it determines the direct target genes, selecting those for which the motif of the associated TF is notably enriched. A regulon is subsequently defined comprising a TF and its direct targets. Finally, the regulatory activity score (RAS) is computed for each cell by measuring the area under the recovery curve. The conventional SCENIC protocol, however, struggles with scalability for extensive datasets and is susceptible to variations in sequencing depth. To enhance both scalability and robustness, we introduced a modification by partitioning data into metacells before

applying SCENIC to these gene expression profiles [14]. This adjustment significantly improves data quality and reduces computational demands, marking a notable advancement in the application of SCENIC to single-cell RNAseq data analysis.

### Regulon clustering

This study employs a comprehensive computational strategy to delineate the regulatory crosstalk between TFs and their corresponding target genes, concentrating on the clustering of TFs. Initially, the method involves filtering TF-target interaction data to isolate pairs that surpass a predefined significance threshold (> 1), thus ensuring the examination prioritizes regulatory interactions of utmost relevance. Subsequent analysis is dedicated to identifying pivotal regulatory TFs by calculating the extent of their target gene regulation, spotlighting these as hub genes within the regulatory network for detailed exploration. To represent the intricate web of TF-target interactions, we construct an undirected graph model, with the spatial arrangement of this graph refined via a force-directed algorithm to intuitively portray the network's architecture, accentuating the interplay between TFs and their targets. To further enrich our understanding of the network's structure, we apply the Leiden algorithm for community detection, unveiling the modular configuration of the transcription factors based on their regulatory interconnections. This process assigns each TF to a distinct cluster, facilitating a nuanced analysis of the regulatory landscape.

### Cell-cell communication analysis

Utilizing the "CellChat" R package, we created CellChat objects based on the UMI count matrices for each respective group [15]. The database "CellChatDB.human" was employed as the reference for ligand-receptor interactions. The analysis of intercellular communication was conducted using the default settings provided by the package. To evaluate and compare the counts and intensities of interactions, we merged the CellChat objects from each group using the "mergeCellChat" function. The "netVisual\_diffInteraction" function enabled us to visualize variations in the number and intensity of interactions among specific cell types across different groups. Furthermore, we identified changes in signaling pathways with the "rankNet" function and illustrated the distribution of signaling gene expression among the groups using both "netVisual\_bubble" and "netVisual\_aggregate" functions.

We further applied NicheNet package to analyze intercellular communication through the lens of ligand activity and the expression patterns of specific downstream targets that are regulated by these key ligands [16]. This approach allows for a detailed understanding of the

signaling processes that underpin interactions between different cell types, leveraging information about ligand-target relationships to infer communication pathways within the cellular microenvironment.

#### Evaluation of the TME disparities and immunotherapy response

In our pursuit to thoroughly and precisely assess immune cell infiltration levels, we analyzed the presence of adverse infiltrated immune cells across several algorithms [17], including MCPcounter [18], EPIC [19], xCell [20], CIBERSORT [21], quanTIseq [22], and TIMER [23], among patients categorized by the AIARS. To accurately portray the immune landscape and architecture within the tumor microenvironment (TME), we also evaluated the ESTIMATE and TIDE indices [24, 25]. These measures are instrumental in offering vital insights into the potential for immunotherapy and understanding the prognostic implications for breast cancer patients. Furthermore, we quantified immune checkpoints, which serve as indicators of the immune state and provide preliminary predictions of patient responsiveness to immune checkpoint inhibitor (ICI) therapy. This comprehensive approach to evaluating the immune profile within the TME is critical for advancing personalized medicine and enhancing treatment strategies for breast cancer patients.

#### Determination of therapeutic targets and drugs for high AIARS patients

In the quest to identify therapeutic targets and drugs for high AIARS patients, our methodology began with filtering out duplicate compounds from the Drug Repurposing Hub, resulting in a curated list of 6,125 compounds (<https://clue.io/repurposing>). The selection of therapeutic targets associated with breast cancer outcomes was determined through Spearman correlation analysis. This analysis focused on the relationship between the AIARS and gene expression levels, selecting genes with a correlation coefficient greater than 0.15 and a *P*-value less than 0.05. Furthermore, genes exhibiting a correlation coefficient below −0.30 and a *P*-value below 0.05 were identified as associated with poor prognosis. The significance of these genes was further assessed by exploring the relationship between CERES scores from the Cancer Cell Line Encyclopedia (CCLE) and risk scores, particularly for brain cells [26].

To refine predictions regarding drug responsiveness, we leveraged data from the Cancer Therapeutics Response Portal (CTRP) and the PRISM project, both of which provide extensive drug screening and molecular data across various cancer cell lines. Differential expression analysis was carried out between bulk samples and cell lines. The pRRophetic package was utilized to implement a ridge regression model for predicting drug

response. This model, trained using expression data and drug response metrics from solid Cancer Cell Lines (CCLs), exhibited excellent predictive accuracy, validated through 10-fold cross-validation [27].

Moreover, to pinpoint the most promising therapeutic drugs for breast cancer, we engaged in Connectivity Map (CMap) analysis. This involved comparing gene expression profiles across different risk subgroups and submitting the top 300 genes (comprising 150 up-regulated and 150 down-regulated genes) to the CMap website (<https://clue.io/query>). Intriguingly, we found that a negative CMap score suggested a higher therapeutic potential against breast cancer, indicating an inverse relationship between the CMap score and a compound's effectiveness as a potential treatment.

#### Patient stratification

To examine gene expression in breast cancer specimens, RNA was extracted employing TRIzol reagent (Invitrogen, Carlsbad, CA, USA), followed by cDNA synthesis and quantitative reverse transcription PCR (qRT-PCR) using GoScript reverse transcriptase and Master Mix (Promega), in adherence to the guidelines provided by the manufacturer. The CFX96 Touch Real-Time PCR Detection System (BioRad, Hercules, CA, USA) was utilized for data acquisition. Gene expression quantification was conducted through the  $2^{-\Delta\Delta C_q}$  method, with GAPDH serving as the normalization control. Subsequently, patients were segregated based on their gene expression profiles, utilizing a predefined formula derived from the AIARS. This stratification was instrumental in identifying patients with differential risk profiles, facilitating tailored therapeutic strategies.

#### Immunohistochemistry experiment

We collected tissue samples from 30 breast cancer patients undergoing surgery at Guizhou Provincial People's Hospital. These samples were then subjected to Hematoxylin and Eosin (H&E) staining, following established protocols. The diagnosis was independently confirmed by two pathologists, with cohort details provided in Table S1. For the immunohistochemistry (IHC) analysis, we employed procedures for paraffin-embedded samples as outlined in our earlier studies [28, 29], with specific antibodies detailed in Table S2. Adhering to standardized protocols and scoring systems, protein expression levels were independently assessed by two pathologists, consistent with methodologies from our prior research [29].



## Results

### Construction of artificial intelligence-derived redox signature

The overall design of this study is displayed in Fig. 1. In our comprehensive investigation of redox biology within breast cancer, we curated a set of redox-related genes using the GeneCards database to conduct differential expression analyses across three independent datasets: TCGA-BRCA, GSE93601, and GSE76250 (Figure S1). These analyses revealed significant differences in expression patterns between tumor and normal tissues. Leveraging this gene set, we developed an artificial intelligence-aided redox signature (AIARS) by employing 108 different algorithmic strategies, validated through ten-fold cross-validation. We computed the mean C-index for each algorithm within the TCGA-BRCA training cohort and five additional external cohorts to ascertain the effectiveness of each algorithm (Fig. 2A). The RSF algorithm, exhibiting the highest mean C-index, was chosen as the definitive model for our study (Fig. 2B, C). To further validate the prognostic significance of the identified redox genes, we performed univariate Cox regression analysis (Fig. 2D) and determined the AIARS score for every sample across the six participating cohorts, thereby providing a nuanced understanding of their prognostic potential (Figure S2).

### Evaluation of AIARS with 88 published signatures in breast cancer

To establish the prognostic independence of the AIARS from other clinical indices, we employed Cox univariate and multivariate analyses (Figure S3A). A nomogram integrating AIARS, stage, and age was then developed to accurately forecast the overall survival (OS) of breast cancer patients at various intervals (Figure S3B-E). The kernel-smoothing hazard plot highlighted that patients with high AIARS scores faced a greater likelihood of recurrence and poorer outcomes (Figure S3F).

In assessing the predictive capacity of AIARS, we meticulously reviewed and gathered data from 88 published signatures, testing their efficacy across 10 independent breast cancer cohorts. The univariate Cox regression analysis underscored AIARS's unique consistency in demonstrating statistical significance across all cohorts (Fig. 3A). Subsequent comparisons of the predictive accuracy between AIARS and these 88 signatures, using the C-index across the 10 breast cancer cohorts (Fig. 3B), unequivocally showed AIARS's superior performance, underscoring its robustness and effectiveness in prognostication within the breast cancer context.

### Genetic alteration landscape of AIARS

Subsequently, extensive endeavors were implemented using the multi-omics analysis to identify genomic

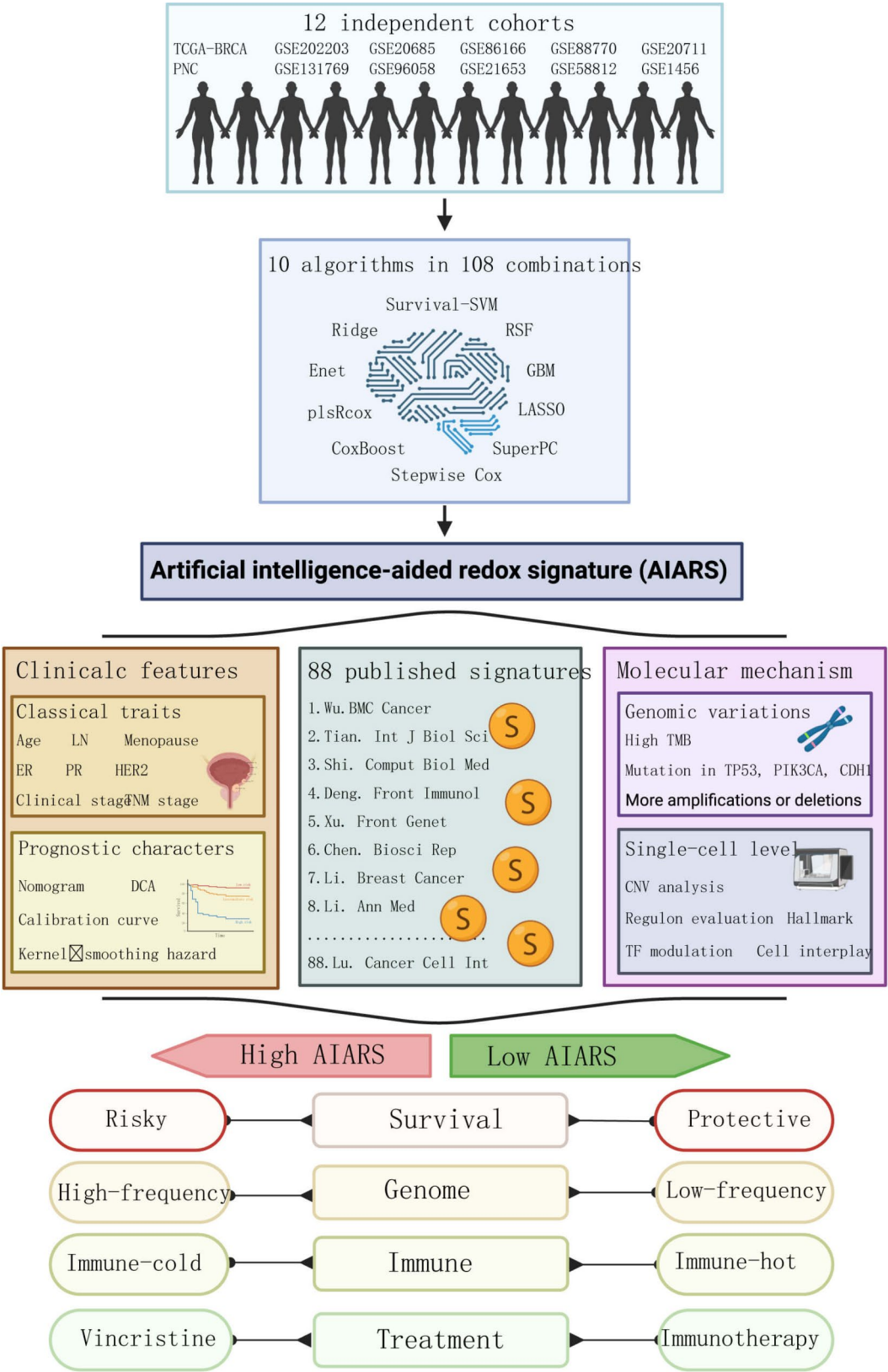
alterations. The result from tumor mutational burden (TMB) announced that the TMB of high AIARS patients was strikingly higher, accompanied by multiple mutation signatures, such as the combination of SBS2 and SBS13, SBS7b and SBS7d (Fig. 4A, C, D). By examining the 10 critical oncogenic signaling pathways, our analysis revealed that classical tumor suppressor genes such as TP53, CREBBP, and TGFBR1/2 exhibited higher mutation rates in the group with elevated AIARS levels compared to the group with lower AIARS levels. Conversely, genes like PIK3CA/B, CDH1, and AKTs showed more frequent mutations in the low AIARS group (Fig. 4A, B).

In delving into the Copy Number Alterations (CNA) landscape between these two groups, it was observed that the group with higher AIARS levels demonstrated significantly more amplifications or deletions at the chromosome arm level. Notable amplifications included regions 3q26.32, 6p23, 8q24.21, 10p15.1, and 12p13.33, while significant deletions were identified in 5q11.2, 5q21.3, 17q21.31, 19p13.3, and 19q13.32 (Fig. 4A, E). The above findings implied the unfavorable prognosis in high AIARS patients, which were able to be demonstrated due to the markable gain of multiple oncogenic genes (PVT1, MYC, CCD26, and GSDMC) in 3q26.32, accompanied by many losses of genes (GPBP1, RAB3C, DDX4, and ITGA1) in 5q11.2 (Fig. 4A).

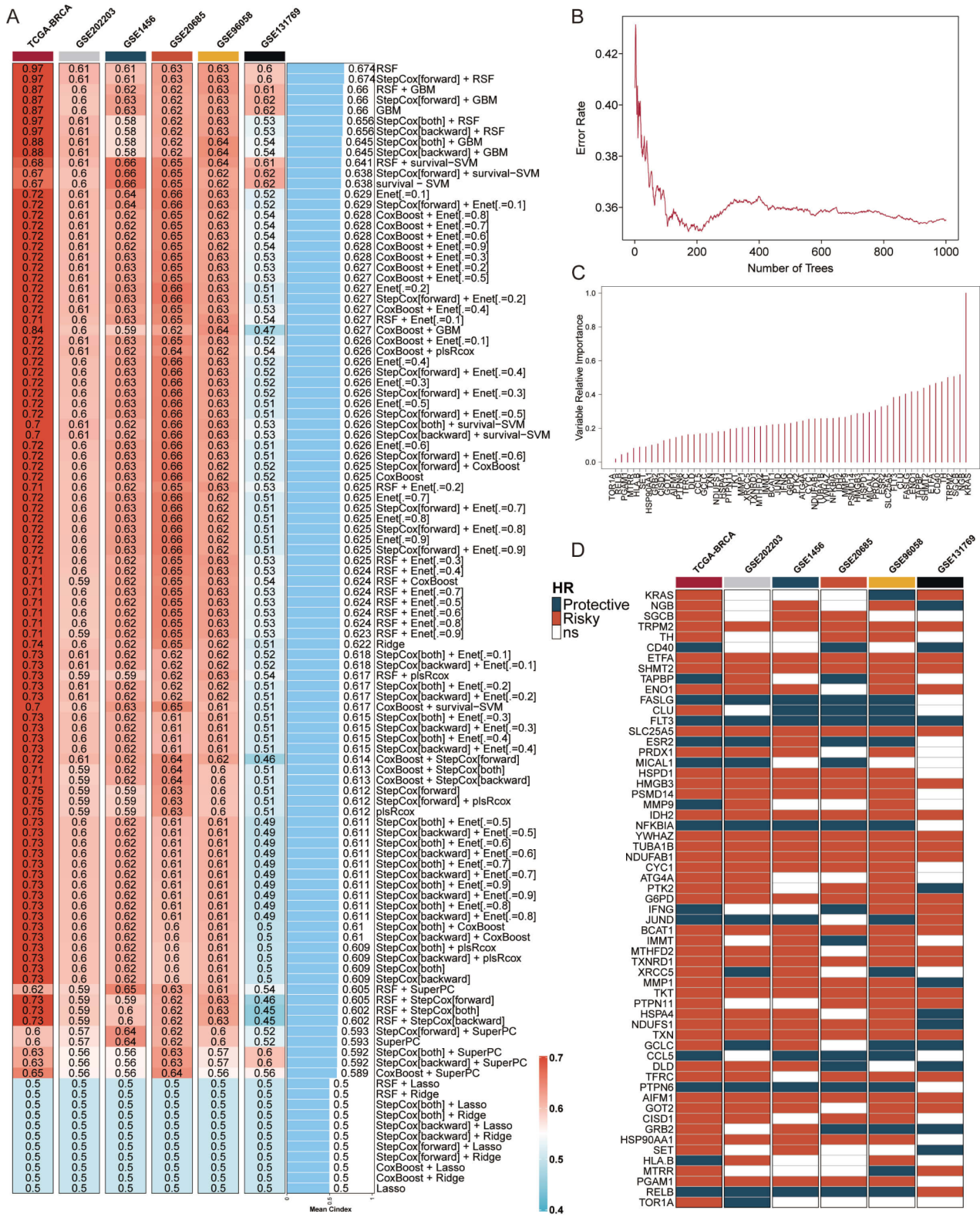
### Deciphering biological mechanisms of AIARS at single-cell level

The single-cell transcriptome analysis was introduced to meticulously assess the AIARS in 12 breast cancer patients, of which included 6 tumor and 6 normal tissues (Figure S4A, B). We identified 13 clusters and 7 cell types (Fig. 5A, B; Figure S4C, D). The representative markers of each cell type were also shown (Fig. 5C; Figure S4E). Moreover, we summarized the distributions of 7 cell types between the tumor and normal tissues, revealing that T cells, macrophages, and epithelial cells accounted for higher proportions in breast cancer patients, while other cells were primarily enriched in normal tissue (Fig. 5D).

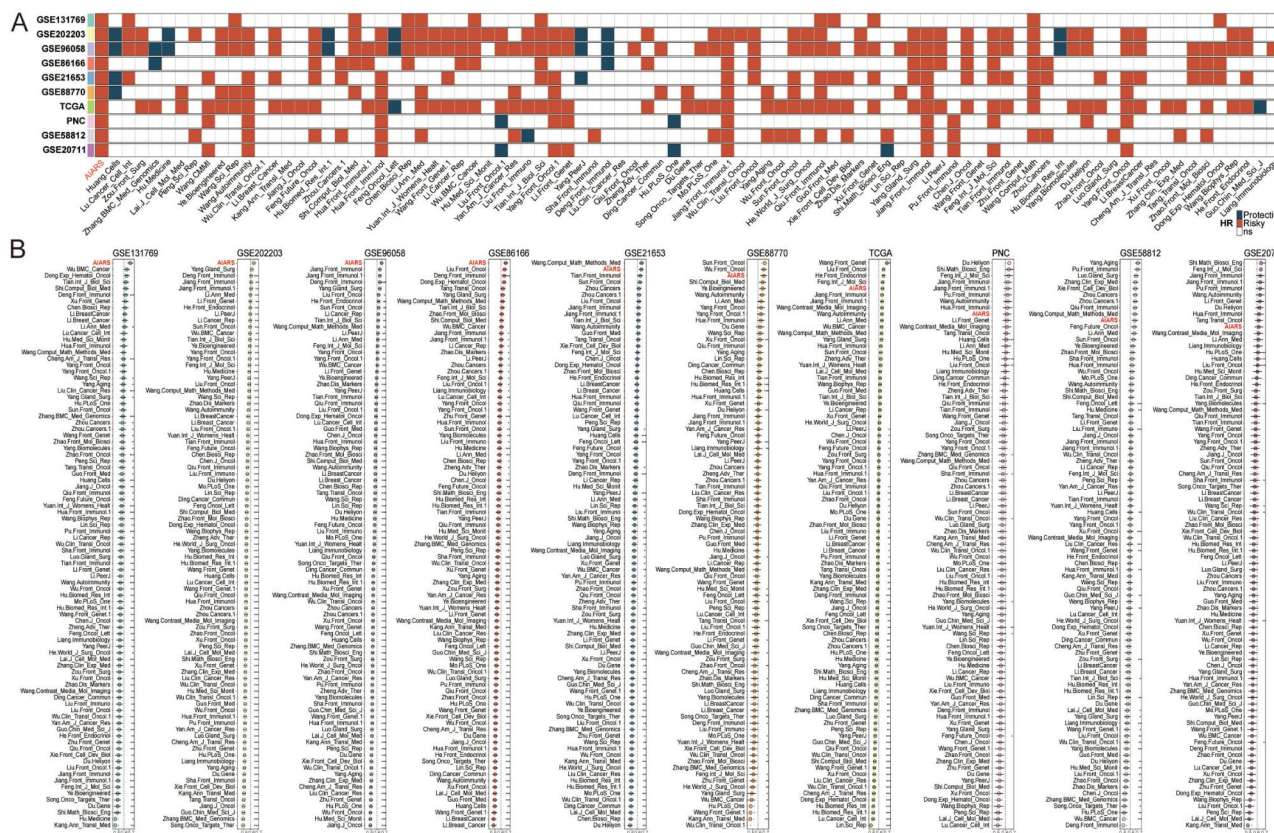
We then estimated the AIARS score for each cell, and observed a significant various distribution between the two AIARS groups (Fig. 5E, F). The differential expression analysis and GSEA were applied to elucidate the potential functional pathways of AIARS (Figure S4F, G). Take epithelial cells (donor for tumor cells) for example, high AIARS group was remarkably enriched in apoptosis, proteasome, focal adhesion, and tight junction. While the low AIARS group was predominantly associated with the oxidative phosphorylation, and reactive oxygen (Figure S4G). The tumor cells were further distinguished using the copyKat algorithm based on the CNA (Fig. 5G). We observed a higher AIARS score in tumor-aneuploid than



**Fig. 1** The overall flow of this study







**Fig. 3** Comparisons of signatures across multiple breast cancer studies. **(A)** Univariate Cox regression analysis of AIARS and 88 published signatures. **(B)** C-indices of AIARS and 88 published signatures in GSE13176, GSE20220, GSE96058, GSE86166, GSE21653, GSE88770, TCGA-BRCA, PNC, GSE58812 and GSE20711. Z-score test:  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$

in tumor-diploid, implying the significance of AIARS in breast cancer progression (Fig. 5H).

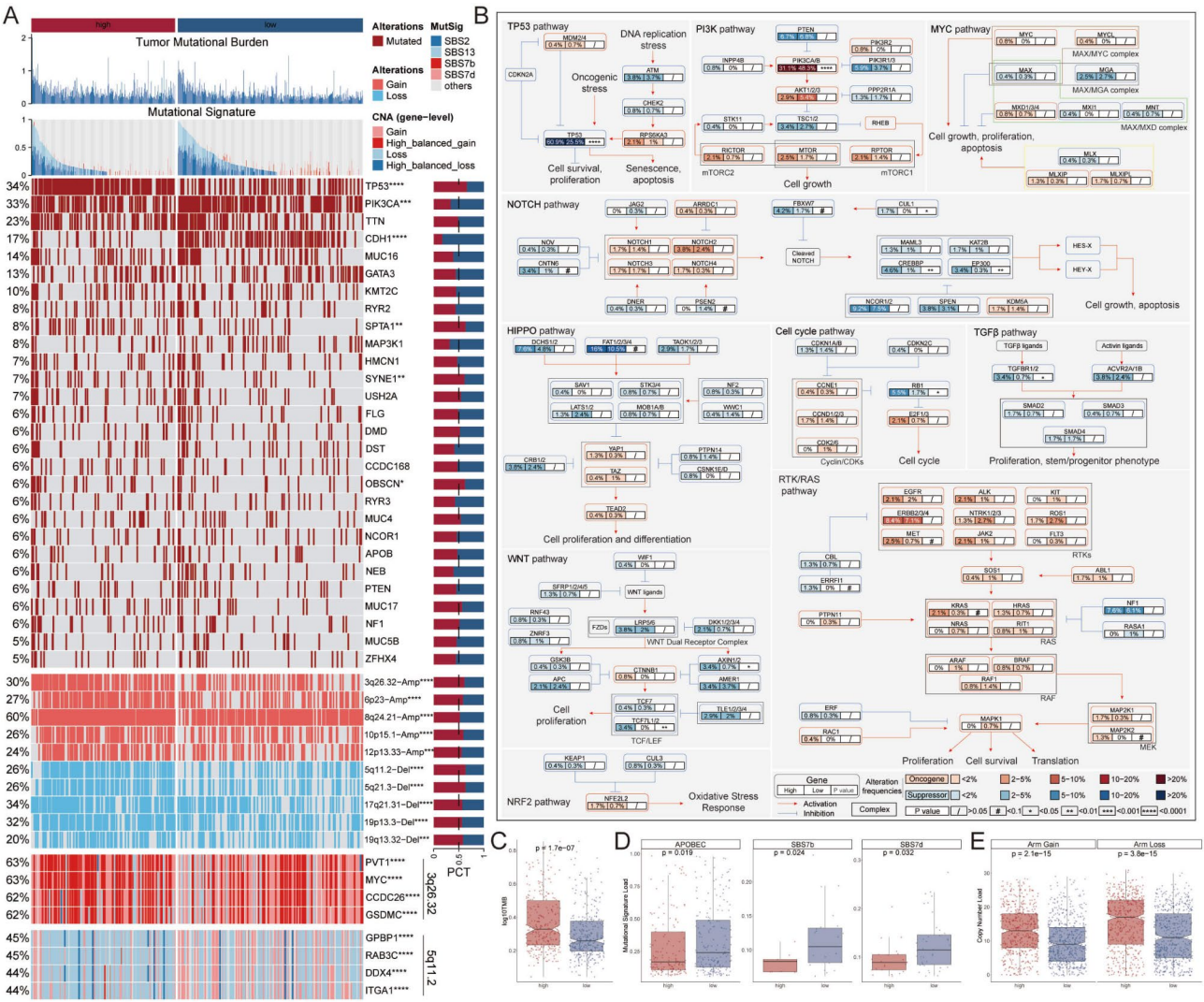
### Exploration of specific regulons for cell identity and AIARS

To comprehensively construct the gene regulatory networks, we applied the SCENIC pipeline to analyze single-cell RNA-seq data with cis-regulatory sequence information. In brief, the gene expression data was transformed to the regulon activity score (RAS) of transcription factors (TF) (Fig. 6A, B). We further performed variance decomposition analysis based on principal component analysis (PCA) to explore the specific regulons for AIARS and cells. Results showed that PC1 accounted for cell type-specific TFs, while PC2 was correlated with AIARS-specific TFs (Fig. 6C, D; Figure S5A, B). The variations of gene expression and TF activity for RUNX3 and IRF3 were demonstrated (Figure S5C, D). RUNX3 was activated in the high AIARS group across all cell types, but the expression of RUNX3 was not significantly changed. Whereas the opposite was true for IRF3 (Figure S5C, D).

In our quest to pinpoint pivotal regulators of cell identity, we evaluated the activity of each regulon across various cell types, assigning a regulon specificity score (RSS)

using Jensen-Shannon divergence to gauge its association with specific cell identities (Fig. 6E) [13, 30]. Focusing on regulons with the highest RSS values, we delved into their functional characteristics (Fig. 6F; Figure S5E), identifying CREB3L4, SPDEF, and GATA3 as key regulons uniquely associated with epithelial cells (Fig. 6E). Support from UMAP plots further highlighted the specificity of these regulons' activities to epithelial cells (Fig. 6F). Regulons specific to other cell types were also presented and analyzed (Fig. 6E; Figure S5E).

Understanding that transcription factors often collaborate to modulate gene expression, we systematically explored the combinatory patterns of these regulatory elements. By assessing the atlas-wide similarity of RAS scores for each regulon pair through the Leiden algorithm, we observed a remarkable organization of these 313 regulons into 10 distinct modules, demonstrating intricate patterns of regulation within the cellular landscape (Fig. 6G; Figure S5F). Therein, modules D and E contributed chiefly to AIARS (Fig. 6H; Figure S5F). We next focused on the exact TFs that drive epithelial cells transcriptomic changes by AIARS. GSEA analysis identified multiple pathway variations (Fig. 6I). For example, TNFA signaling via NFkB was activated in the epithelial



**Fig. 4** Genomic landscape and pathway alterations in AIARS. **(A)** Depicts tumor mutational burden and mutational signatures across the genome, with a detailed heatmap below illustrating the frequency of gene alterations such as gains and losses. Bar graphs on the side quantify the alterations for specific genes. **(B)** Mutation map of genes across ten canonical cancer pathways, illustrating the complexity of oncogenic signaling and the interplay between different pathways in tumorigenesis. Each pathway is detailed with gene names and alterations that contribute to cancer development, including cell survival, proliferation, and apoptosis. **(C-E)** Box plots comparing TMB **(C)**, mutational signatures **(D)** and CNV **(E)** across different AIARS groups

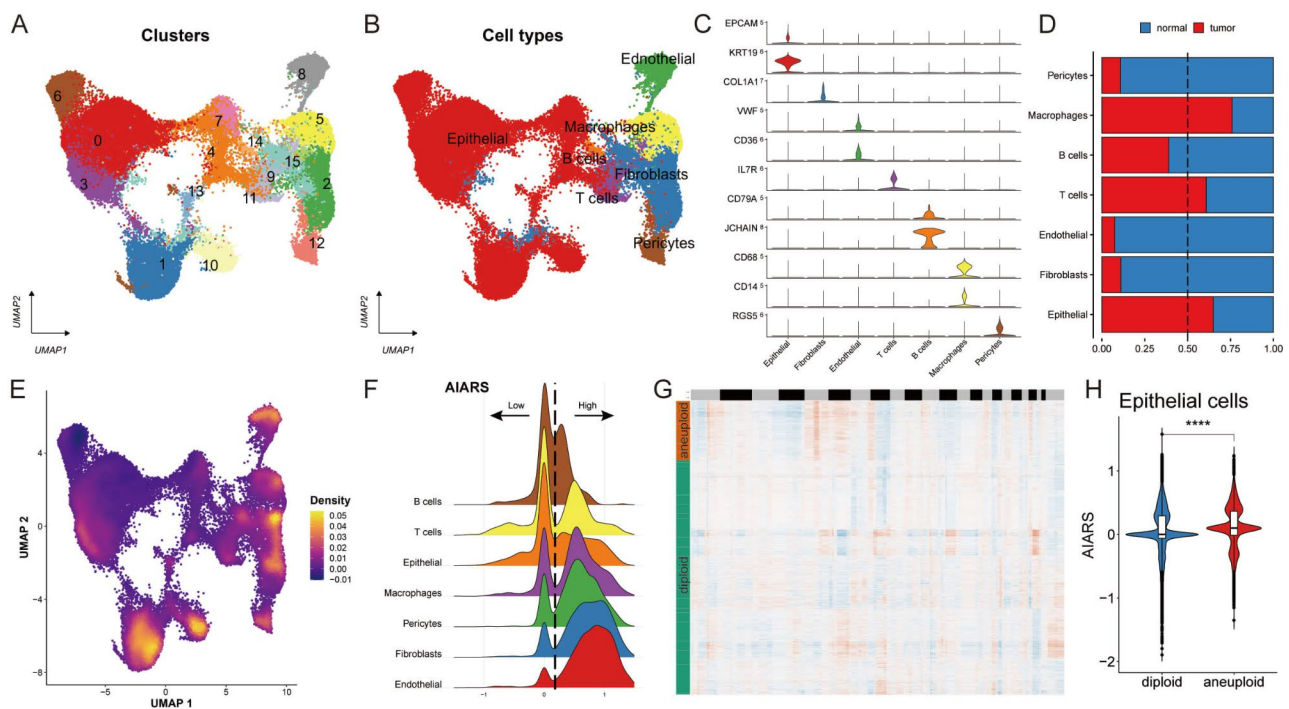
cells with high AIARS (Fig. 6I, J). Further analysis identified TFs that contributed to this pathway, which also involved in AIARS progression (Fig. 6K). Detailed regulation network was shown in Fig. 6L. We also confirmed the TFs that prevents AIARS progression, for example, G2M checkpoint (Figure S5G, H).

**Cell-cell communication between the AIARS groups**

Since cell-cell interaction plays a significant role during breast cancer progression, the CellChat analysis was applied to assess the communication across seven cell types of AIARS. The interaction numbers and strength were assessed between two AIARS groups, unveiling the stronger cell-cell communication in the high AIARS (Fig. 7A). The complicated interactions across each cell

were further emphasized via the interaction network. The result outlined that massive interactions were detected among epithelial cells, fibroblasts and endothelial cells in the high AIARS group, accompanied by the weaker interaction strength among T cells, and B cells (Fig. 7B). Subsequently, the interaction of 38 signaling pathways was illuminated between the two groups, of which, except for FN1, MIF, APP, CD99 and ADGRE5, 33 pathways were principally activated in the high AIARS cells, such as collagen, laminin, and cxcl pathways (Fig. 7C). For example, epithelial cells were allivated form the signals of fibroblasts, and the gene expression of collagen was significantly different between two AIARS groups in epithelial cells (Figure S6A, B). Moreover, the outgoing and incoming interaction strength was deployed to monitor





**Fig. 5** Single-cell analysis of cellular heterogeneity and AIARS level. **(A)** UMAP visualization showing various clusters within a dataset, each color representing a different cluster, suggesting distinct subpopulations based on gene expression or cellular phenotypes. **(B)** UMAP plot color-coded to represent different cell types within the sample. **(C)** Violin plots for specific marker genes across different cell types, showing the distribution of expression levels, which can indicate the relative abundance or activity of these cells. **(D)** Bar chart comparing the proportion of each cell type between normal and tumor samples. **(E)** UMAP density plot indicating the expression of the AIARS across different cell populations, with warmer colors representing higher expression levels. **(F)** Violin plot displaying the distribution of AIARS across various cell types. **(G)** Heatmap generated by CopyKat depicting the inferred copy number variations across a range of genes (rows) and individual cells or samples (columns), with the color intensity reflecting the degree of aneuploidy or copy number alterations. **(H)** Violin plot contrasting the AIARS score between diploid and aneuploid cells within the epithelial cell population, indicating significant differences between these two genomic states

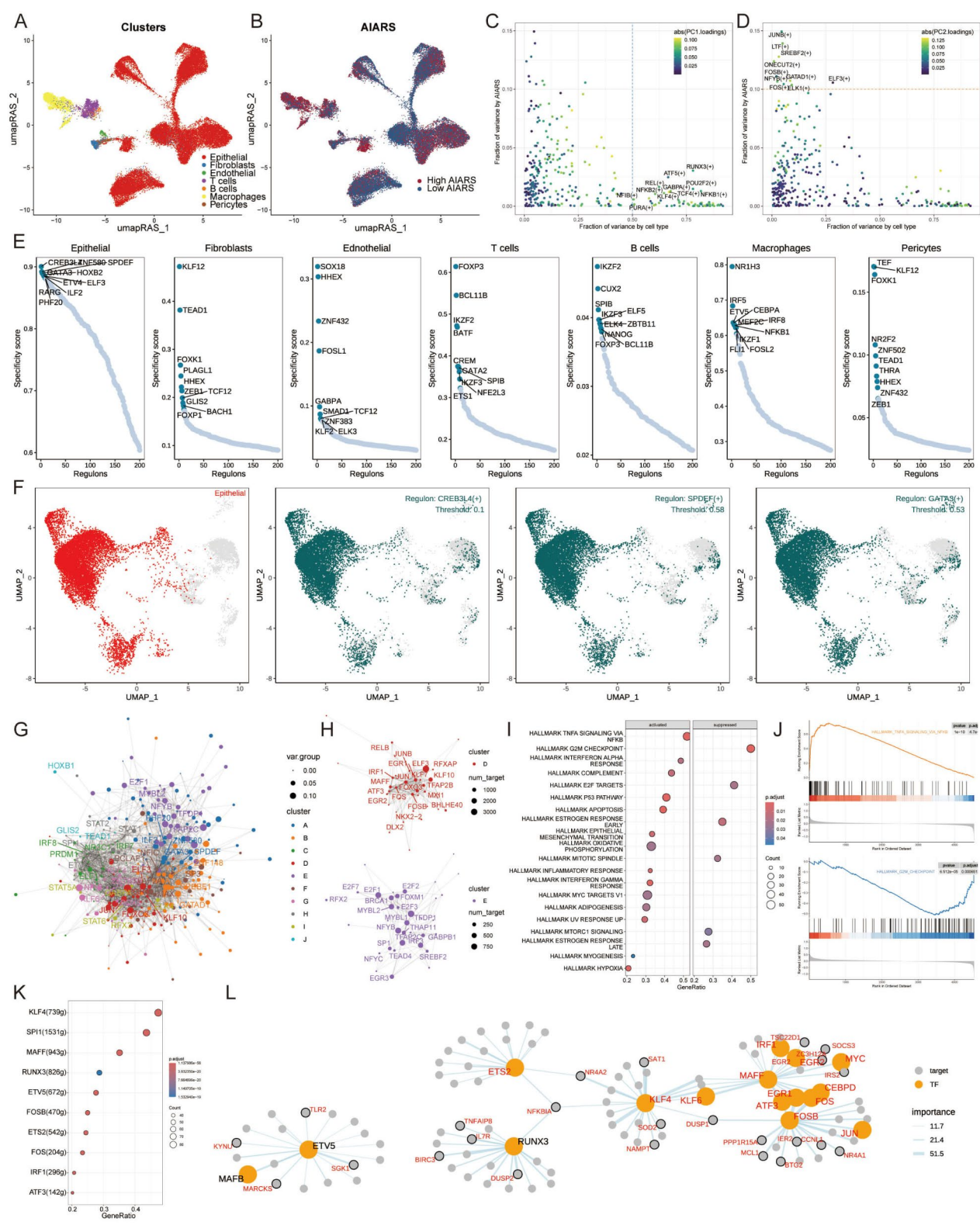
the cell-cell interactions. Compared with the low AIARS cells, stronger incoming interactions of epithelial cells, and endothelial cells were found in high AIARS cells, accompanied by weaker communication with T cells and B cells (Fig. 7D). In epithelial cells, various pathways were specific in the high AIARS group, such as APP, laminin and THBS pathway (Fig. 7E; Figure S6C, D).

We further performed nichenetr analysis to further explore the effects of different cell types on epithelial cells in the TME. We inferred potential ligands that may regulate epithelial cells from different AIARS group (Figure S6E). One of the most interesting ligands due to its role in cancer promotion in the context of tumorigenesis is CCN1. In fact, the CCN1 gene is significantly overexpressed in fibroblasts and endothelial cells compared with T cells and B cells (Figure S6F), and it was also observed to be overexpressed in endothelial cells of high AIARS (Figure S6G, H). Depper analyses revealed the differential activity of ligand-receptor pairs for AIARS, we summarize the top-predicted links in the circos plot (Fig. 7G). We observed high interactions of CCN1-SDC4 pairs in the high AIARS cells, which indicated the fibroblasts and endothelial cells were the main sender cells to

influence the pathway changes of epithelial cells (Fig. 7F). It is interesting that the involved transcription factors, such as TP53, MYC and RAC1, were mutated have higher mutation rates in high AIARS patients (Figs. 4B and 7H).

#### Evaluating potential immunotherapy targets for AIARS

Since the immune microenvironment is involved in tumor progression, its evaluation could enhance the understanding of breast cancer outcomes. We applied 6 algorithms to assess the immune infiltration of AIARS. A higher proportion of immune cell infiltrations were found in patients with low AIARS, such as T cells, B cells, NK cells and other immune cell types. In contrast, only a few immune cells infiltrated in high AIARS patients (Fig. 8A). It is noteworthy that the expression levels of immune checkpoint inhibitors (ICIs) are acknowledged as key indicators to apprise the responsiveness to immunotherapy. A large number of ICIs were minor in the low AIARS patients, for instance, PD-L1, PD-1, CTLA4, HAVCR2, LAG3, and TIGIT, while less ICIs were observed in high AIARS groups (Fig. 8B). IHC was performed to support the above results using the representative cell markers and clinical ICIs (Fig. 8C, D).



**Fig. 6** (See legend on next page.)

(See figure on previous page.)

**Fig. 6** AIARS-specific regulon activity analysis. **(A)** umapRAS plot displaying distinct clusters within a cell population, with each color representing a unique cluster, potentially indicative of cell subtypes or states. **(B)** umapRAS plot shows the levels of the AIARS across the cell population, with color intensity indicating score magnitude. **(C)** Variance analysis plot demonstrating the effect of both cell types and AIARS on transcription factor activity, with color mapping to PC1, which highlights the primary variance driven by these two factors. **(D)** Variance analysis plot with color mapping to PC2. **(E)** Rank for regulons in each cell type based on RSS. **(F)** UMAP plots for epithelial cell highlighted, each highlighting cells where a particular regulon is active. **(G)** Network graph of transcription factor modules identified through the Leiden algorithm, depicting 313 regulons organized into 10 major modules. **(H)** Network graph focusing on modules D and E, which are chiefly contributing to AIARS. **(I)** GSEA indicating pathway variations related to AIARS in epithelial cells. **(J)** Representative pathways activated (TNFA signaling) and inhibited (G2M checkpoint) in the context of high AIARS, illustrating the differential pathway engagement within epithelial cells. **(K)** TFs that contribute to TNFA signaling via the NFkB pathway, highlighting their involvement in the progression of AIARS. **(L)** Detailed regulatory network illustrating the interactions of transcription factors involved in TNFA signaling of AIARS progression

Previous findings demonstrated the indicative roles for response to immunotherapy. We determined the responsiveness via multiple analyses, including ESTIMATE and TIDE. It was summarized that low AIARS patients were superior in ESTIMATE score, immune score, and stromal score, while high AIARS patients occupied greater advantage in tumor purity, hinting that the former was more likely to be prone to immunotherapy (Fig. 9A). We also observed the higher TIDE value and dysfunction value in low AIARS patients, but the exclusion value did not show remarkable differences between the AIARS group (Fig. 9B). Interestingly, it was underlined that the outcomes of patients with combined low AIARS with high TIDE outperformed other types of patients (Fig. 9C). The results from anti-cancer immunity interpreted that except from the tumor stages 2 and 7, low AIARS patients possessed higher activity of anti-tumor immunity in the other five stages in comparison with high AIARS patients (Fig. 9D).

ICIs have emerged as a transformative approach in cancer immunotherapy over the past several decades, yet their effectiveness in solid tumors, including breast cancer, remains limited [31]. In our study, we sought to explore the predictive capability of AIARS levels regarding the efficacy of immune checkpoint blockade therapies. Utilizing SubMap algorithms, we predicted the response to immunotherapy and discovered that patients exhibiting low AIARS levels were significantly more likely to benefit from treatments with anti-PD-1 (Bonferroni corrected  $p=0.003$ ) and anti-MAGE-A3 (Bonferroni corrected  $p=0.007$ ) drugs (Fig. 9E). Further investigation into AIARS levels within the IMvigor210 (anti-PD-L1) and GSE78220 (anti-PD-1) cohorts [32, 33] revealed that patients with low AIARS exhibited notably better survival rates and clinical outcomes compared to those with high AIARS levels in both the IMvigor210 (Fig. 9F-I) and GSE78220 (Fig. 9J-M) studies. These findings collectively suggest that patients with lower AIARS levels may derive significant advantages from ICI therapy, underscoring the potential of AIARS as a predictive marker for immunotherapy responsiveness.

### Identification of therapeutic agents for the high AIARS patients

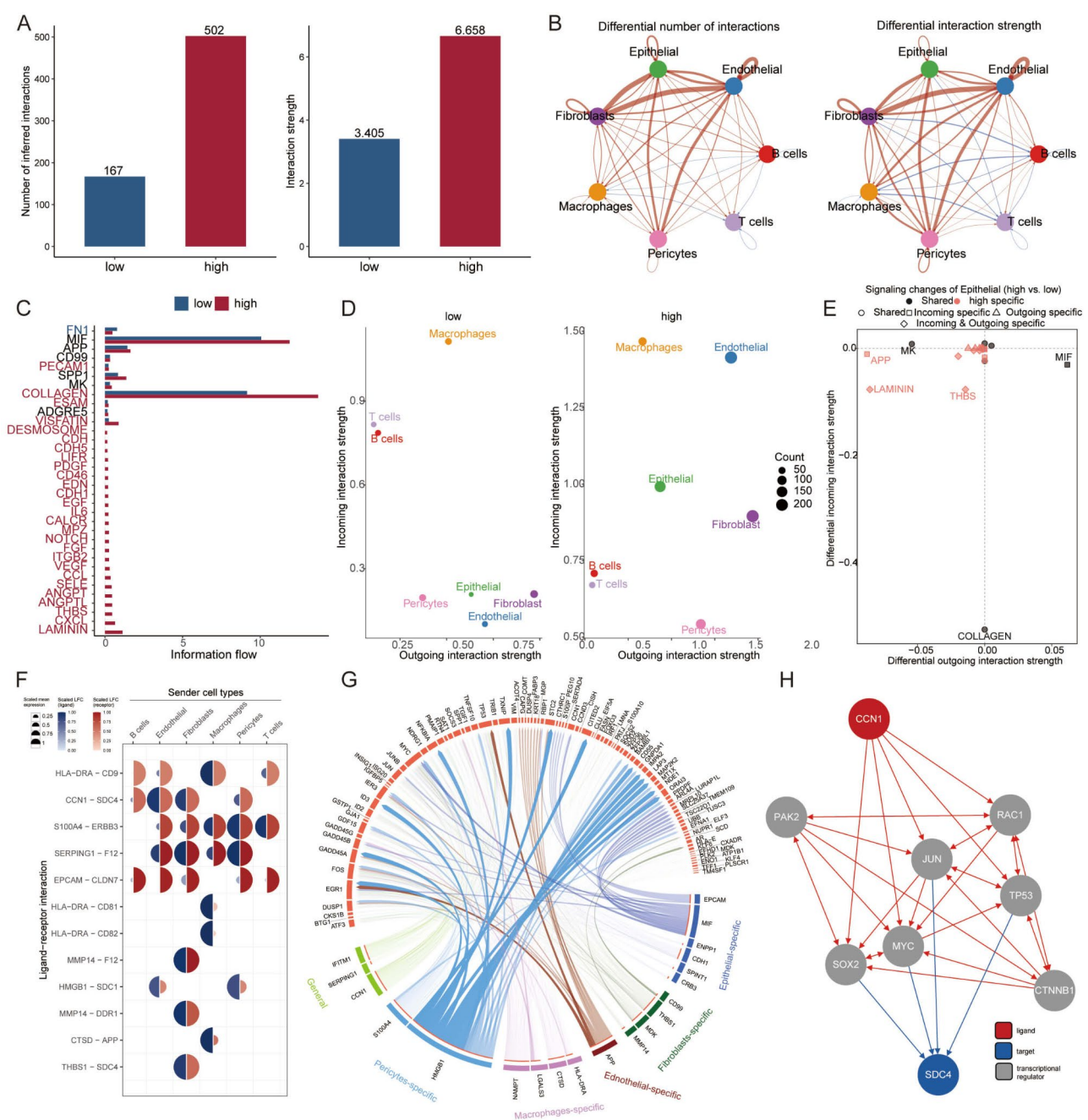
Chemotherapy is the standard treatment against cancer, we then developed potential agents for BC patients with high AIARS using sensitivity data from multiple dataset. We initially identify the therapeutic targets utilizing Spearman's correlation analysis, suggesting that AIARS exhibited positive relevance to the abundance of six targets (PAICS, HSD17B10, AHCY, PARS2, GART, WDR5), and notably negative relations to CERES score of them, hinting that these six targets could serve as potential therapeutic targets for high AIARS patients (Fig. 10A). Furthermore, it was highlighted that these six therapeutic targets were tightly associated with multiple drug action pathways, hence they were considered critical therapeutic targets for BC patients with high AIARS (Fig. 10B).

Following by, nine compounds separately were obtained from CTPR (SB-743921, paclitaxel, vincristine, and BI2536) and PRISM datasets (romidepsin, deforolimus, JQ1-(+), ispinesib, and vincristine). Comparing the AUC value of each compound in two AIARS group, it was emphasized that a higher AUC value was recognized in low AIARS patients, revealing an unfavorable effect for chemotherapy in this group (Fig. 10C). A multiple-perspective analysis was executed to determine optimal therapeutic drugs from these nine candidates, of which related information of each compound, including clinical status, experimental evidence, mRNA expression levels, and CMap score, were completely listed. According to the CMap score, it was dictated that vincristine was eventually selected as potential therapeutic drugs for high AIARS patients (Fig. 10D).

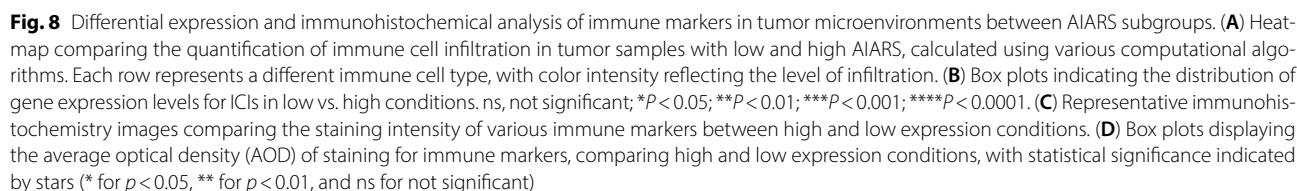
### Discussion

This study marks a considerable advancement in fusing redox biology with machine learning to refine the prognosis of breast cancer. Through the development of a predictive model grounded in redox-related gene signatures, we have uncovered potential molecular mechanisms correlated with disease outcomes. The effectiveness of our model not only reiterates the crucial role of redox processes in breast cancer but also illustrates the capability of computational algorithms in untangling the complexity of biological data sets.

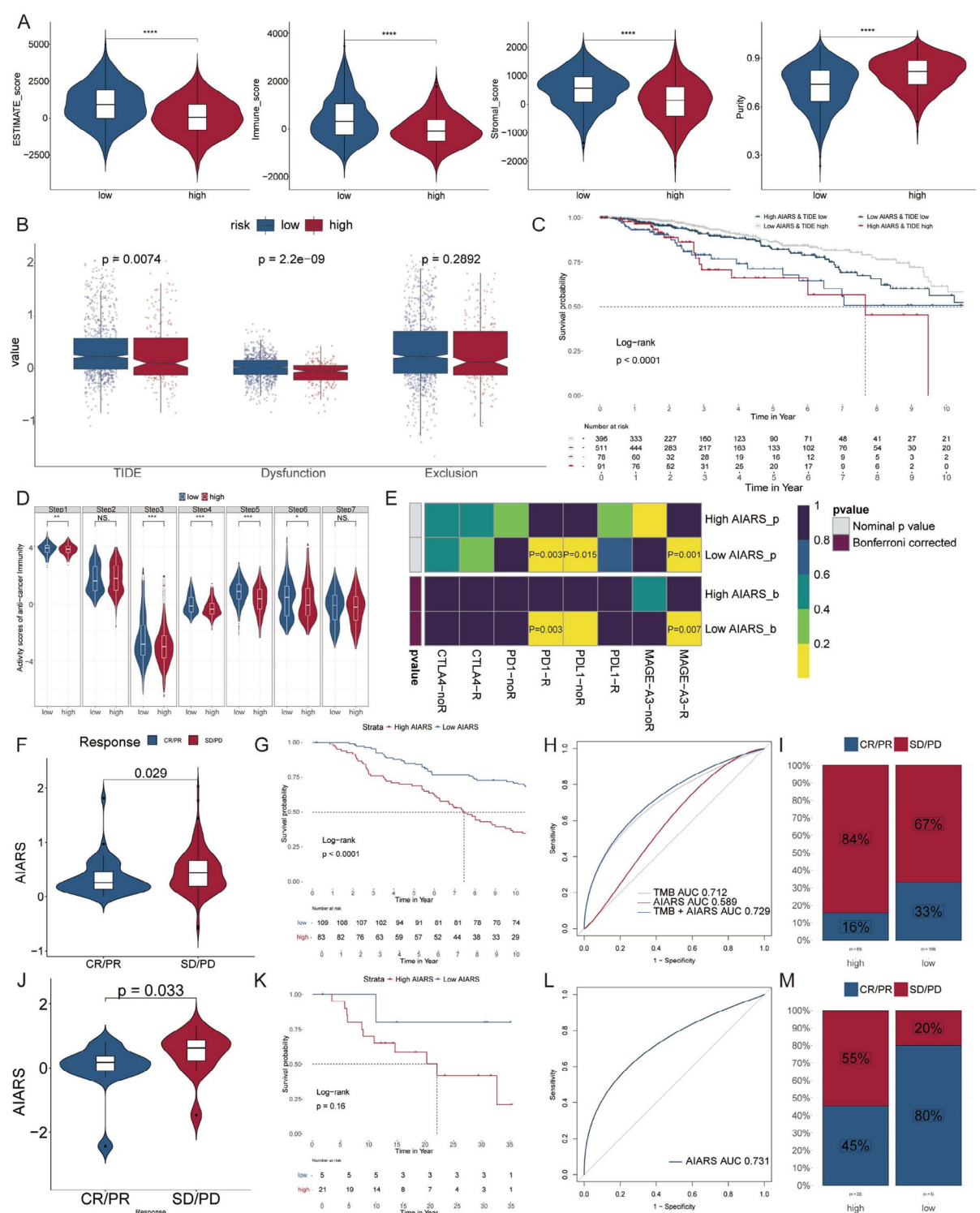




**Fig. 7** Dissecting cell-cell interactions and signaling pathway dynamics in AIARS-modulated breast cancer progression. **(A)** Bar graphs representing the number of cell-cell interactions, with a significant increase in interactions noted within the high AIARS group compared to the low. **(B)** Network diagrams illustrating differential interaction strength among various cell types, highlighting more complex interactions in high AIARS, especially between epithelial, fibroblasts, and endothelial cells. **(C)** A bar chart showing the information flow of 38 signaling pathways, indicating the majority are more active in the high AIARS group, with particular emphasis on collagen and laminin pathways. **(D)** Scatter plots comparing the outgoing and incoming interaction strength between cell types in low and high AIARS, demonstrating stronger interactions with epithelial and endothelial cells in the high AIARS group. **(E)** Pathway specificity in epithelial cells in high AIARS, with pathways like APP, laminin, and THBS being notably specific. **(F)** potential ligand-receptor interactions inferred by nichenetr analysis, with specific focus on the activity differences between cell types in different AIARS groups. **(G)** Circos plot summarizing top-predicted ligand-receptor pairs, pointing out the heightened interactions, especially those involving CCN1-SDC4 in high AIARS cells. **(H)** Network graph revealing mutations in transcription factors such as TP53, MYC, and RAC1, which are associated with higher mutation rates in patients with high AIARS, impacting cell signaling pathways







**Fig. 9** AIARS correlation with immune infiltration and response to immunotherapy in breast cancer. **(A)** Violin plots comparing ESTIMATE, immune, and stromal scores between low and high AIARS groups. **(B)** Box plots illustrating TIDE values, dysfunction, and exclusion metrics in low vs. high AIARS patients. **(C)** Kaplan-Meier survival curves for breast cancer patients stratified by AIARS and TIDE. **(D)** Violin plots depicting the activity of anti-tumor immunity across tumor stages. **(E)** Heatmap demonstrating the predictive power of AIARS for responsiveness to different ICIs treatment. The responses (R) and non-responses (noR) to therapies such as anti-PD-1, and anti-MAGE-A3 are stratified by AIARS, with lower AIARS associated with higher responsiveness. **(F, J)** The violin chart displaying the relation between AIARS and anti-PD1 **(F)** and anti-PDL1 **(J)** responses. **(G, K)** The survival possibility of low- and high-patients in anti-PD1 **(g)** and anti-PDL1 **(k)** cohorts. **(H, L)** Estimating the predictive ability of AIARS via AUC value combining with TMB or without TMB in anti-PD1 **(H)** and anti-PDL1 **(L)** cohorts. **(I, M)** The percentages of CR/PR and SD/PD in anti-PD1 **(I)** and anti-PDL1 **(M)** cohorts based on AIARS



**Fig. 10** Screening therapeutic targets and drugs for high AIARS breast cancer patients. **(A)** Scatter plots with Spearman's correlation coefficients showing the association between AIARS and the abundance of six potential therapeutic targets in breast cancer patients. The negative correlation with CERES scores indicates these targets may be particularly relevant for patients with high AIARS. **(B)** Network analysis highlighting the connections between the six therapeutic targets and their related drug action pathways. The targets are shown to have significant implications for the development of therapeutics for BC patients with high AIARS, underscoring their potential as critical intervention points. **(C)** Box plots comparing the AUC values of nine compounds derived from the CTRP and PRISM datasets in low vs. high AIARS patient groups. A higher AUC value in low AIARS patients suggests less favorable chemotherapy outcomes in this subgroup. **(D)** Summary table outlining the multi-perspective analysis of nine candidate compounds, showing their clinical status, experimental evidence, mRNA expression levels, and Connectivity Map (CMap) scores. Vincristine emerges as a potentially suitable therapeutic agent for high AIARS patients, as suggested by the CMap score

A particularly noteworthy discovery from our investigation was the identification of several redox-related genes previously unrecognized in the context of breast cancer prognosis. This revelation suggests a more complex redox landscape within breast cancer than once understood, potentially involving unexplored biological pathways. These insights necessitate a reassessment of the existing paradigms of redox biology in breast cancer, underscoring the imperative for further research into these novel genes and their contribution to disease progression.

Our findings indicate that patients with low AIARS scores tend to have higher levels of immune infiltration and are more responsive to immunotherapy, while those with high AIARS scores exhibit greater sensitivity to chemotherapeutic agents such as vincristine. This contrast suggests that redox biology may significantly influence both chemotherapy resistance and immunotherapy response in breast cancer patients. It is well-established that increased levels of reactive oxygen species (ROS) and dysregulated redox signaling can promote cancer cell survival, drug resistance, and treatment failure. The overproduction of ROS leads to the activation of several downstream pathways that enhance the tumor's ability to evade cell death. For example, oxidative stress can induce DNA repair mechanisms, protect tumor cells from apoptosis, and increase the expression of drug-efflux pumps, all of which contribute to chemoresistance. Patients with high AIARS scores, which reflect heightened redox activity, may exhibit resistance to chemotherapy due to these enhanced survival mechanisms. Interestingly, the increased mutation burden observed in high AIARS patients also supports the notion that redox dysregulation may foster genetic instability, further complicating treatment outcomes [6]. Conversely, our data suggest that patients with low AIARS scores exhibit a better response to immune checkpoint inhibitors (ICIs) such as anti-PD-1 and anti-PD-L1 therapies. Lower redox activity might mitigate the immunosuppressive tumor microenvironment typically induced by oxidative stress, allowing immune cells—especially T cells—to infiltrate tumors more effectively. Furthermore, redox dysregulation is known to inhibit antitumor immunity by promoting immunosuppressive signaling and reducing the activity of effector immune cells. Therefore, tumors with lower AIARS may be less equipped to evade immune surveillance, making them more susceptible to ICIs [4].

The findings from our study highlight the complex interplay between redox biology and therapeutic outcomes. While we have shown that AIARS is a useful predictor of both chemotherapy and immunotherapy responses, further research is needed to fully elucidate the molecular mechanisms underlying these effects. For instance, *in vitro* and *in vivo* studies could investigate

the direct impact of redox modulation on chemotherapy resistance and immune cell activity. Additionally, combining redox-modulating agents with ICIs could be explored as a therapeutic strategy for patients with high AIARS, potentially overcoming redox-mediated resistance and enhancing treatment efficacy.

Despite the model's promising predictive power, we recognize its limitations. Utilizing multiple datasets from varied sources enhances the robustness and applicability of our findings, yet it introduces the challenge of data heterogeneity. Additionally, the intricate nature of the machine learning models utilized in this study poses hurdles for clinical application. The opaque “black box” nature of these algorithms complicates their interpretability, making it challenging for clinicians to integrate them into decision-making processes without comprehensive understanding.

Our investigation also revealed discrepancies with existing literature, particularly in the prognostic value assigned to certain established redox-related genes. These variances could be attributed to differences in patient cohorts, therapeutic approaches, or analytical methods. Such discrepancies highlight the diversity of breast cancer and the impact of external factors on disease outcomes, emphasizing the need for tailored medicine strategies. While our current analysis was limited by the lack of longitudinal molecular data, future studies will aim to investigate AIARS score changes in response to therapy. Understanding how redox-related gene signatures evolve before and after treatment could shed light on the mechanisms driving drug resistance, particularly in high-AIARS patients who are more susceptible to chemotherapy.

Looking ahead, our study establishes a foundation for subsequent research directions. Prospective studies with larger, multi-institutional cohorts are crucial for validating the predictive accuracy of our model and confirming its generalizability across diverse patient populations. Furthermore, experimental validation of the newly identified redox-related genes from this study may unveil novel therapeutic targets, paving the way for the development of redox-based treatments for breast cancer. The integration of our model into clinical practice will necessitate dedicated efforts to improve its interpretability and flexibility, ensuring it can reliably inform therapeutic decisions and enhance patient care.

## Conclusions

In conclusion, our study not only bridges a critical gap in breast cancer research by linking redox biology with machine learning but also sets the stage for the next era of precision oncology. By tailoring treatment strategies to the molecular profile of individual tumors, we move

closer to realizing the promise of personalized medicine in breast cancer care.

#### Abbreviations

PCA	Principal Component Analysis
OS	Overall Survival
TME	Tumor microenvironment
TCGA	The Cancer Genome Atlas
GEO	Gene Expression Omnibus
RF	Random Forest
LASSO	Least Absolute Shrinkage and Selection Operator
GBM	Gradient Boosting Machine
SuperPC	Supervised Principal Component
plsRcox	Partial Least Squares Cox Regression
Enet	Elastic Net
ICI	Immune checkpoint inhibitor
CNA	Copy Number Alterations

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12935-024-03534-8>.

**Additional file 1: Figure S1.** Differential expression of redox genes in breast cancer across various datasets. **Figure S2.** Construction and validation of AIARS. **Figure S3.** Prognostic characters of AIARS. **Figure S4.** Comprehensive cellular profiling of breast cancer tissue. **Figure S5.** Transcription factor activity correlation and contribution analysis in cell types. **Figure S6.** Cellular interactions in tumor microenvironments based on AIARS.

**Additional file 2: Table S1:** Clinical Characteristics of BC patients. **Table S2.** Antibodies used in this study.

#### Acknowledgements

Not Applicable.

#### Author contributions

JH and KD are the designer of the study. TW analyzed the data and wrote the manuscript. SW, ZL, and JX contributed to the study design, information collection. All authors read and approved the final manuscript.

#### Funding

This study was supported by the Talent Fund of Guizhou Provincial People's Hospital [2022-33], and National Natural Science Foundation of China [82260502 and 82272656].

#### Data availability

No datasets were generated or analysed during the current study.

#### Declarations

##### Ethics approval and consent to participate

The study was conducted following the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Guizhou Provincial People's Hospital. Written informed consent was obtained from all subjects.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 14 July 2024 / Accepted: 14 October 2024

Published online: 09 November 2024

#### References

- Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. *Lancet*. 2021;397:1750–69.
- Wang D, Liu B, Zhang Z. Accelerating the understanding of cancer biology through the lens of genomics. *Cell*. 2023;186:1755–71.
- Lenicke C, Cochemé HM. Redox metabolism: ROS as specific molecular regulators of cell signaling and function. *Mol Cell*. 2021;81:3691–707.
- Muri J, Kopf M. Redox regulation of immunometabolism. *Nat Rev Immunol*. 2021;21:363–81.
- Jeziarska-Drutel A, Rosenzweig SA, Neumann CA. Role of oxidative stress and the microenvironment in breast cancer development and progression. *Adv Cancer Res*. 2013;119:107–25.
- Gorini C, Harris IS, Mak TW. Modulation of oxidative stress as an anticancer strategy. *Nat Rev Drug Discov*. 2013;12:931–47.
- Jorgenson TC, Zhong W, Oberley TD. Redox imbalance and biochemical changes in cancer. *Cancer Res*. 2013;73:6118–23.
- Gendoo DMA, Zon M, Sandhu V, Manem VSK, Ratanasirigulchai N, Chen GM, Waldron L, Haibe-Kains B. MetaGxData: clinically annotated breast, ovarian and pancreatic Cancer datasets and their Use in Generating a Multi-cancer Gene signature. *Sci Rep*. 2019;9:8770.
- Liu Z, Guo C, Dang Q, Wang L, Liu L, Weng S, Xu H, Lu T, Sun Z, Han X. Integrative analysis from multi-center studies identifies a consensus machine learning-derived lncRNA signature for stage II/III colorectal cancer. *EBioMedicine*. 2022;75:103750.
- Wang L, Liu Z, Liang R, Wang W, Zhu R, Li J, Xing Z, Weng S, Han X, Sun YL. (2022) Comprehensive machine-learning survival framework develops a consensus model in large-scale multicenter cohorts for pancreatic cancer. *Elife* 11.
- Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X, Bryant VL, Penington JS, Di Stefano L, Ribera T, Wilcox N, Mann S, Papenfuss GB, Lindeman AT, Smyth GJ, G. K., and, Visvader JE. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J*. 2021;40:e107333.
- McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in single-cell RNA sequencing data using Artificial Nearest neighbors. *Cell Syst*. 2019;8:329–e337324.
- Suo S, Zhu Q, Saadatpour A, Fei L, Guo G, Yuan GC. Revealing the critical regulators of cell identity in the mouse cell Atlas. *Cell Rep*. 2018;25:1436–e14451433.
- Baran Y, Bercovich A, Sebe-Pedros A, Lubling Y, Giladi A, Chomsky E, Meir Z, Hoichman M, Lifshitz A, Tanay A. MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol*. 2019;20:206.
- Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. 2021;12:1088.
- Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods*. 2020;17:159–62.
- Zeng D, Ye Z, Shen R, Yu G, Wu J, Xiong Y, Zhou R, Qiu W, Huang N, Sun L, Li X, Bin J, Liao Y, Shi M, Liao W. IOBR: Multi-omics Immuno-Oncology Biological Research to Decode Tumor Microenvironment and signatures. *Front Immunol*. 2021;12:687975.
- Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautès-Fridman C, Fridman WH, de Reyniès A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17:218.
- Racle J, Gfeller D. EPIC: a Tool to Estimate the proportions of different cell types from bulk gene expression data. *Methods Mol Biology (Clifton N J)*. 2020;2120:233–48.
- Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017;18:220.
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12:453–7.
- Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, Krogsdam A, Loncova Z, Posch W, Wilflingseder D, Soppor S, Ijsselstein M, Brouwer TP, Johnson D, Xu Y, Wang Y, Sanders ME, Estrada MV, Ericsson-Gonzalez P, Charoentong P, Balko J, de Miranda N, Trajanoski Z. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med*. 2019;11:34.
- Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. TIMER: a web server for Comprehensive Analysis of Tumor-infiltrating Immune cells. *Cancer Res*. 2017;77:e108–10.

24. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, Wucherpfennig KW, Liu XS. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med*. 2018;24:1550–8.
25. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
26. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia NV, Montgomery PG, Cowley GS, Pantel S, Goodale A, Lee Y, Ali LD, Jiang G, Lubonja R, Harrington WF, Strickland M, Wu T, Hawes DC, Zhivich VA, Wyatt MR, Kalani Z, Chang JJ, Okamoto M, Stegmaier K, Golub TR, Boehm JS, Vazquez F, Root DE, Hahn WC, Tsherniak A. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet*. 2017;49:1779–84.
27. Yang C, Huang X, Li Y, Chen J, Lv Y, Dai S. (2021) Prognosis and personalized treatment prediction in TP53-mutant hepatocellular carcinoma: an in silico strategy towards precision oncology. *Brief Bioinform* 22.
28. Wang T, Li T, Li B, Zhao J, Li Z, Sun M, Li Y, Zhao Y, Zhao S, He W, Guo X, Ge R, Wang L, Ding D, Liu S, Min S, Zhang X. Immunogenomic Landscape in breast Cancer reveals immunotherapeutically relevant Gene signatures. *Front Immunol*. 2022;13:805184.
29. Wang T, Ba X, Zhang X, Zhang N, Wang G, Bai B, Li T, Zhao J, Zhao Y, Yu Y, Wang B. Nuclear import of PTPN18 inhibits breast cancer metastasis mediated by MVP and importin  $\beta$ 2. *Cell Death Dis*. 2022;13:720.
30. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*. 2011;25:1915–27.
31. Katz H, Alsharedi M. Immunotherapy in triple-negative breast cancer. *Med Oncol*. 2017;35:13.
32. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE, Koeppen III, Astarita H, Cubas JL, Jhunjhunwala R, Banchereau S, Yang R, Guan Y, Chalouni Y, Ziai C, Şenbabaoğlu J, Santoro Y, Sheinson S, Hung D, Giltane J, Pierce JM, Mesh AA, Lianoglou K, Riegler S, Carano J, Eriksson RAD, Höglund P, Somarriba M, Halligan L, van der Heijden DL, Lorient MS, Rosenberg Y, Fong JE, Mellman L, Chen I, Green DS, Derleth M, Fine C, Hegde GD, Bourgon PS, R., and, Powles T. TGF $\beta$  attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*. 2018;554:544–8.
33. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, Berent-Maoz B, Pang J, Chmielowski B, Cherry G, Seja E, Lomeli S, Kong X, Kelley MC, Sosman JA, Johnson DB, Ribas A, Lo RS. Genomic and transcriptomic features of response to Anti-PD-1 therapy in metastatic melanoma. *Cell*. 2016;165:35–44.

### Publisher's note

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