Discover Oncology

Research

Reconstruction of cancer marker analysis with holistic anatomical precision implicates heterogeneity development during breast tumor progression

Zhicheng Ge[2] · Jing Wang[1] · Libing He[1] · Meng Zhao[3] · Yang Si[1] · Siyuan Chang[1] · Guoyan Zhang[1] · Shan Cheng[1] · Wei Ding[1]

Received: 7 January 2024 / Accepted: 9 October 2024
Published online: 15 October 2024

Abstract
Biomarkers are not only of significant importance for cancer diagnosis and selection of treatment plans but also recently increasingly used for the evaluation of malignancy development and tumor heterogeneity. Large-size tumors from clinical patients can be unique and valuable sources for the study of cancer progression, particularly to the extent of intratumoral heterogeneity. In the present study, we obtained a series of post-surgery puncture samples from a breast cancer patient with a $4 \times 3.5 \times 2$ cm tumor in its original size. Immunohistochemistry for Ki-67, COX-2, and CA IX was performed and the expression levels within the breast cancer tumor mass were evaluated in the reconstructed 3D models. To further evaluate the intratumoral heterogeneity, we performed high throughput whole transcriptome sequencing of 12 samples from different spatial positions within the tumor tissue. Comparing the reconstructed 3D distribution of biomarkers with projected tumor growth models, asymmetric and heterogeneous expansion of tumor mass was found to be possibly influenced by factors such as blood supply, inflammation and/or hypoxia stimulations, as suggested from the correlation between the results of Ki-67 and CA IX or COX-2 staining. Furthermore, high-throughput RNA sequencing data provided additional information for profiling the intratumoral heterogeneity and expanded the understanding of cancer progression. Digital technology for medical imaging once properly integrated with molecular pathology examinations will become particularly helpful in dissecting out in-depth information for precision medicine. We prospect that this approach, facilitated by rapidly advancing artificial intelligence, could provide new insights for clinical decision-making in the future. Strategies for the continuous development from the present study for better performance and application were discussed.

Keywords Breast cancer · Biomarkers · Tumor heterogeneity · 3D reconstruction · Molecular anatomy

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-024-01442-x.

✉ Shan Cheng, chengs@ccmu.edu.cn | [1]Department of Medical Genetics and Developmental Biology, School of Basic Medical Sciences, Capital Medical University, Beijing 100069, People's Republic of China. [2]Department of General Surgery, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, People's Republic of China. [3]Department of Infection Control, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, People's Republic of China.

1 Introduction

Breast cancer is a devastating disease that is found to be the second leading cause of cancer-associated death in women [1]. Diagnosis of breast cancer at early stages in patients is one of the most important approaches to improve breast cancer treatment outcomes. Among various diagnosis platforms, medical imaging systems have provided valuable data and are intensively used in patients at hospitals and clinics. Various imaging techniques, such as mammography, magnetic resonance imaging (MRI), positron emission tomography (PET), computed tomography (CT), and single-photon emission computed tomography (SPECT) are currently being used for the detection of breast cancers at various stages [2, 3]. Unfortunately, imaging results do not usually give direct biological information about the causal disease. On the other hand, growing evidence indicates that the initiation and progression of breast cancers involve the changes of many biomarker proteins under the influence of various genetic or environmental factors. Thus, histopathological examinations remain the primary method in the diagnosis of breast diseases. Pathologist reports serve as the golden standard in the classification of breast cancer types and grades and greatly impacted the subsequent therapy and treatment [4]. As the field of molecular pathology rapidly progresses, more and more biochemical biomarkers, including proteins, DNAs, mRNAs, and microRNAs are considered and employed as new diagnostic tools for breast cancers [5]. However, imaging and molecular pathology are gapped under a majority of circumstances, where a lot of information is not fully integrated, shared and explored at the case level and this might reduce the possibility to joint-force the power of major diagnostic technologies for a better understanding the etiology and progression of cancers.

In recent years, a critical challenge of tumor heterogeneity of its impacts on treatment outcomes has been recognized in breast cancer patients as well, but the integrated view of tumor heterogeneity development was obscure until high throughput and single-cell level genomic analyses became available and were widely conducted [6–9]. The results from these studies have brought up the classification of breast cancer subtypes based on molecular markers. The major four clinical types classified by molecular markers are Luminal A, Luminal B, HER2 positive, and basal-like. Even within these well-studied breast cancer types, the heterogeneity, particularly intra-tumor heterogeneity within the primary tumors with spatial information is still rather unknown [10]. Besides the common theory of cancer stem cells and clonal evolution in tumor heterogeneity studies, the importance of cancer markers in the tumor microenvironment have also been increasingly recognize [11, 12].

Breast cancer with negative ER, PR, and HER2 expression is defined as triple-negative breast cancer (also called triple-negative basal-like breast cancer, TNBC). The clinical characteristics of breast cancers of this type are highly invasive, metastasis-prone, and unsuitable for endocrine therapy or targeted therapy [13]. Therefore, it is difficult for traditional breast cancer predictors to accurately reflect the biological characteristics of this type of breast cancer. Therefore, the study of the biological behavior of TNBC needs to consider the influence of multiple driving factors on tumor progression. In clinicopathological diagnosis, tissues covering the primary lesion could be selected to determine the expression of biological markers, such as Ki-67 [14]. As a molecular marker for cell proliferation [15], Ki-67 (encoded by *MKI67* gene) expression was found significantly higher in TNBC than in other histological types of breast cancer [16]. The protein levels of Ki-67 can be used as an independent predictor of therapeutic effect and prognosis of various breast cancer types [17], and cancers from other tissue origins as well. Another common cancer marker of diagnostic importance is Cyclooxygenase-2 (COX-2, encoded by the *MT-CO2* gene), which acts as a major rate-limiting enzyme in prostaglandin synthesis. Over-expression of COX-2 can be observed in many solid tumors. COX-2 is known to functionally involve the occurrence and development of breast cancer by inhibiting apoptosis, increasing invasiveness, and promoting angiogenesis of tumors [18]. As another marker for tumor malignancy, Carbonic Anhydrase IX (CA IX, encoded by the *CA9* gene) can induce the degradation of the extracellular matrix and promote the infiltration and migration of tumor cells. CA IX is also considered a reliable hypoxia marker and is closely associated with cancer metastasis [19, 20]. Although with a reasonable collection of cancer markers for selection, it is still difficult to accurately reflect the biological characteristics of breast cancers in general or to predict the progression of breast cancer of specific types. It could be even more challenging to evaluate the behavior of TNBC cells with the existence of intratumoral heterogeneity, therefore, comprehensive profiling and detailed analyses of multiple driving factors (as indicated by the expression cancer markers) which are associated with different aspects of cancer cell phenotypes needs to be considered.

In the present study, the spatial distribution of Ki-67, COX-2, and CA IX was analyzed in a case of large-sized invasive breast cancer (TNBC) at baseline. The patient was not previously received radiotherapy and chemotherapy

procedures. From the reconstructed tumor mass 3D model based on MRI images, the indications of these tumor markers on cell proliferation, inflammation, and hypoxia were evaluated comprehensively about the heterogeneity development of the examined breast tumor. In addition, we performed RNA-seq of samples from different spatial positions within the tumor mass to provide a more complete gene expression profile, revealing the intratumoral heterogeneity and cancer progression. From our report, not only the value of large-sized tumors for understanding tumorigenesis was emphasized, but it was also demonstrated that digital imaging technology in combination with molecular pathology examinations can provide new insight to help pathologists and clinicians evaluate the heterogeneity of tumors. The principle and procedures presented in this study can be further developed and used as an model case for comparable investigations with similar objectives.

2 Materials and methods

2.1 Collection of tissue specimen

A female patient (aged 62) was diagnosed with clinical-stage breast cancer (cT2N0Mx) in the left breast at Beijing Friendship Hospital. Magnetic resonance imaging (MRI) revealed an enhanced solitary mass with a surrounding area of low intensity. The mammogram indicated an irregular mass. From the removed tumor mass of approximately $4 \times 3.5 \times 2$ cm, the tissue puncture samples were collected immediately following surgery and stored under −80°C according to the protocol at the Tissue Bank of Cancer Institute of Capital Medical University. Postoperative pathological diagnosis revealed that the tumor was basal-like invasive carcinoma, ER (-), PR (-), and HER2 (-). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the Beijing Friendship Hospital Ethics Committee (2017-P2-006-01). We also confirm that all methods were performed in accordance with the relevant guidelines and regulations.

2.2 Immunohistochemistry (IHC)

The needle aspiration samples were collected from the breast cancer mass using a pair of puncture needles in opposite directions around the center of the tumor. The distance between the paired puncture needles was 20 mm in the coronal axis (X axis), 17.5 mm in the vertical axis (Y axis), and 10 mm in the sagittal axis (Z axis). The diagram is shown in Fig. 1a, b. Each of the frozen puncture samples was sectioned into 5 fragments and cut into slices at a thickness of 5 μm (Fig. 1c). A fully automated IHC and ISH instrument BOND-MAX (Leica Microsystems, Wetzlar, Germany) was used for the staining of tissue sections, after blocking with 5% bovine serum albumin, antibodies for Ki-67 (1:200) (GB13030-2, Service bio, Wuhan, China), COX-2 (1:300) (ab15191, Abcam, Cambridge, UK) and CA IX (1:400) (NB100-417SS, Novus Biologicals, Colorado, USA) was used for probing with a negative control without primary antibody. Secondary antibody incubation and color development with the diaminobenzidine (DAB) chromogen were carried out using Bond™ Polymer Refine Detection (DS9800, Leica Microsystems, Wetzler, Germany). Digital images were acquired using light microscopy (BX43, Olympus, Tokyo, Japan). Immunohistochemical staining for positive areas in integral optical density (IOD) (integral optical density) was measured and independently graded by three pathologists who were mutually blinded to the sample grouping information.

2.3 Whole transcriptome sequencing and data processing

Needle aspiration samples were collected from the breast cancer mass using a pair of puncture needles, ensuring that the same direction and position were maintained as for the samples used in immunohistochemical (IHC) analysis. The obtained samples were then divided into a total of 12 distinct samples, each representing different spatial positions within the tumor (coronal-medial-superficial, coronal-lateral-superficial, coronal-medial-deep and coronal-lateral-deep for coronal (X) axis; vertical-superficial-foot, vertical-superficial-head, vertical-deep-foot and vertical-deep-head for vertical (Y) axis; sagittal-foot-medial, sagittal-head-medial, sagittal-foot-lateral and sagittal-head-lateral for sagittal (Z) axis, respectively, shown in Fig. 3a). These samples were subsequently subjected to RNA sequencing (RNA-seq) analysis. Total RNA extractions were performed with the RNA-Quick Purification Kit (ES-RN001, Yishan Biotechnology Co., Ltd., Shanghai, China) following the manufacturer's protocol. The RNA concentrations were determined using a Nanodrop ND1000

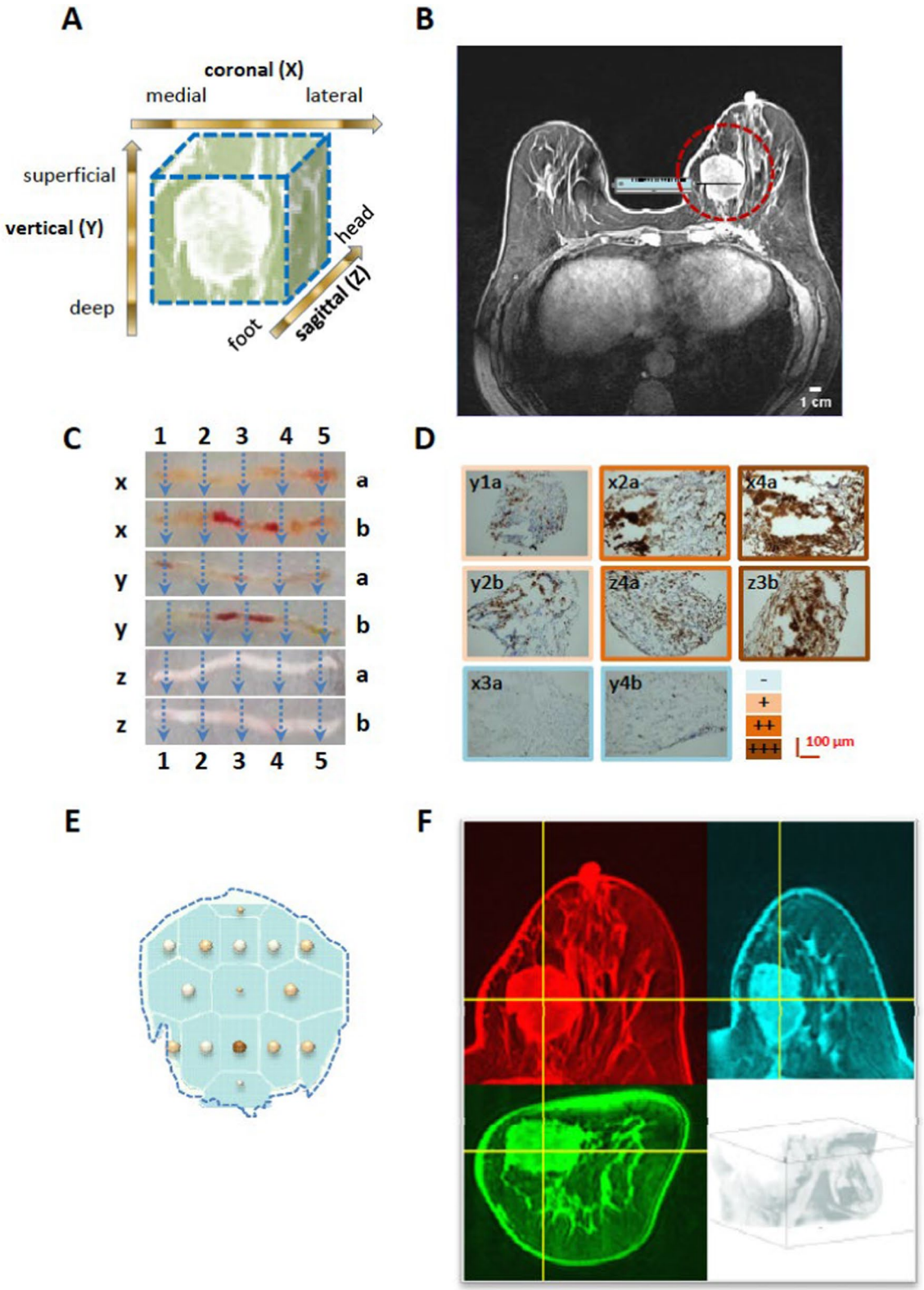

Fig. 1 The anatomical reconstruction for holistic view of cancer markers. **a** The needle puncture aspiration samples of breast cancer tissue were obtained symmetrically from the center of the tumor along the coronal axis (X axis), vertical axis (Y axis), and sagittal axis (Z axis). **b** The diagram of X X-axis puncture showed in the MRI images. **c** Each of the frozen puncture samples were cut into 5 fragments and sliced into sections at a thickness of 5 µm. **d** Marker expressions were detected in the total 30 sampling sites of the breast cancer. **e** The sampling sites were marked on the tumor 3D model. The intensity of the dots represented the variances of marker expression in the sampling sites. The radius of the dots represented the deviations of the positions of the sampling sites. **f** The 3D view of the anatomical reconstruction model

spectrophotometer (Thermo Scientific). The quality assessments and mRNA sequencing libraries were performed in the laboratory of VAHTS Universal V6 RNA-seq Library Prep Kit for Illumina (Vazyme Biotech, Nanjing, China), VAHTS RNA Multiplex Oligos Set1- Set2 for Illumina (Vazyme Biotech, Nanjing, China), VAHTS DNA Clean Beads (Vazyme Biotech, Nanjing, China), VAHTS mRNA Capture Beads (N401-01, Vazyme Biotech, Nanjing, China). All prepared samples were subjected to paired-end multiplex sequenced (2 × 150 bp) on the Illumina Hiseq X10 platform. Approximately 8 GB of sequencing data was generated for each sample. Trimmomatic (Version: 0.39) was selected to input FASTQ sequences for RNA-seq differentially expressed gene flow analysis, and raw data was paired and screened to output clean data. The clean data reads were aligned back to the reference genome using HISAT2 (Version: HISAT2 2.2.0). SAMtools (Version: 1.15.1) was used for binary conversion and sorting. Pathway analysis was performed using GSEA (https://www.gsea-msigdb.org/gsea/index.jsp), all pathway data were downloaded from the C2-Canonical pathways database, and the selected gene sets were shown in Supplementary Table 1. String (https://cn.string-db.org/) was used for functional enrichment analysis of cell proliferation, inflammation, and hypoxia-related genes.

2.4 Image processing and anatomical reconstruction

Image Pro Plus 6.0 (Media Cybernetics, Inc., Rockville, MD, USA) was used for image processing of the evaluation of the staining intensity of cancer markers. ImageJ 2x (National Institutes of Health, Bethesda, MD, USA) was used to reconstruct the holistic view of the intra-tumor distribution of markers according to the spatial coordinate. The plugin modules deployed included packages of stack manipulation, 3D draw shape, 3D watershed Voronoi, and image calculator for various purposes. The MRI images were used with the patient's consent for the study to reconstruct the model of tumor mass in 3D.

2.5 Statistical analysis

SPSS version 20.0 (SPSS, Inc., Chicago, IL, USA) was used for statistical analyses. The coefficient of variation (CV) = $S/\overline{X}$, was calculated to determine the data dispersion, where S represented the standard deviation and $\overline{X}$ represented the arithmetic mean. A threshold of p-value < 0.05 was defined for statistical significance. The Spearman rank test was performed to analyze the correlation between expression levels of breast cancer markers from the sampling sites of 30 in total for each channel. The coefficient numbers evaluated the correlation between different markers in the reconstructed tumor anatomical models and the projected tumor growth model. For comprehensive analysis, data from multiple measurements, including the density readout from MRI images and staining grades of cancer markers, were subjected to clustering using a complete linkage algorithm with Cluster 3.0 software and displayed by Java Treeview (Stanford University, USA).

3 Results

3.1 Spatial analyses of semiquantitative expression of *cancer* in tumor mass

To obtain biological characteristics in aspects of proliferation, inflammation, and hypoxia within the sample breast tumor at different spatial locations, common markers of Ki-67, COX-2, and CA IX were probed by immunohistochemistry and evaluated for grading based on staining intensity (Fig. 1d, Supplementary Fig. 1). To avoid operational errors during experimental procedures, an automated equipment was deployed using the recommended protocols from the manufacturers. The semi-quantification of the staining was determined by summarizing the evaluation from multiple experienced professionals from the pathology department. The coefficient of variation (CV) of Ki-67, COX-2, and CA IX staining from a total of 30 sampling sites was calculated as 36.8%, 44.9%, and 42.6% respectively. The data indicated that intratumorally heterogeneity indeed existed among different spatial locations within the tumor mass, nonetheless, it was an expected result as the varied biological characteristics in tumors were usually seen to be reflected by these commonly used markers. From further analysis of the interconnection between the stained markers, it was interesting to find out that all the compared pairings were statistically different, with Ki-67 and CA IX showing a more significant correlation coefficient (Table 1). The results could be explained as the outcome of the interplay between rapid tumor cell proliferation and hypoxia induction in the local environment. The fast-growing tumor mass often causes local hypoxia conditions from insufficiency of oxygen and nutrients from the blood supply. In turn, hypoxia-induced overexpression

Table 1 Correlation between Ki-67, COX-2, and CA IX expression at different spatial locations

Correlation	Ki-67	COX-2	CA IX
Ki-67	1.00		
COX-2	0.669**	1.00	
CA IX	0.743**	0.686**	1.00

n = 30; spearman rank test, **p < 0.01

of transcription factors, signal molecules and secretory protein or inflammatory factors could stimulate tumor cells to proliferate with acceleration [21]. The discoordinance of these regulatory mechanisms could eventually cause or at least significantly contribute to the development of heterogeneity in tumors.

3.2 Anatomical reconstruction for holistic views of analyzed *cancer* markers

An MRI series consisting of 128 consecutive scans were converted into a 128-frame image stack and subjected to three-dimensional (3D) reconstruction using Image J software (Fig. 1f). Voxels of the tumor mass were extracted by adjusting the threshold values for segmentation, and the size and boundary information of the tumor were obtained and used for creating the cropped subset of the digitally isolated tumor (Fig. 1e). The scaled quantification from the 30 sampling sites of each stained marker was mapped into the 3D reconstructed tumor shape according to the exact measured spatial co-ordinance. The expression variances of the markers and the possible error limits of spatial location were calculated and marked in Fig. 1e. Using information from limited sampling sites to represent the projected marker staining of the whole tumor mass, a 3D Voronoi algorithm was used to determine the section domains where the measured values could best describe the located regions. The watershed filling was performed to reassign the adjacent voxels with the intensity in grayscales for the projected staining of marker expression. The reconstructed marker expression profile in 3D was visualized using the 3D viewer plugin embedded in the software (Fig. 2a). To evaluate the correlation of the marker expression levels, a module for voxel colocalization analysis was used and the Ki-67, COX-2, and CA IX signals were compared pairwise as in separate channels. The correlation coefficients were calculated and shown in Table 2, where the value for the correlation between Ki-67 and CA IX was the highest (R = 0.770). All markers appeared to have a significant correlation with the MRI intensity, especially at the anatomic center region of the tumor, suggesting that the expression of different markers was less divergent at the early stages of tumor development.

3.3 Asymmetrical pattern of *cancer* marker distribution and its implication for tumor growth models

As the large tumor provided a good opportunity to obtain sufficient density data for describing the tumor mass from MRI images, it allowed the study to use hypothetical projection to construct possible tumor expansion models. Theoretically, the growth of tumor mass would be expected to follow a linear propagation profile without apparent intervention. In the case of large tumors with regular shapes, the tumor under investigation in this study could be a suitable sample to mathematically evaluate related models. We first generated a voxel data set calculated by a 3D linear propagation model and simultaneously determined the edges for the stopping of diffusion (Fig. 2b left). We calculated the anatomical center of the tumor from the enhanced MRI images and assumed that the center spot was the primary origin of tumor initiation with the maximum values of detected tumor staining signal. By calculating the correlation coefficient of projected marker distribution and the measured realistic values at the sampling sites using a colocalization analysis program, we could estimate which specific marker better represented the tumor growth profile, or whether the projected model could be improved by adjusting the propagation center position for better data fitness.

Considering the irregularity in MRI pixel density at different positions within the tumor mass, we also attempted to improve the correlation analyses by calculating the spatial center of mass of the tumor and used for reconstruction of the propagation model (Fig. 2b right). No significant improvement was observed as indicated by the small increase in coefficient number (Supplementary Table 1). The results showed the correlation of all three cancer markers of their expression levels exerted the positive values in coefficient measures, indication Ki-67, COX-2 and CA IX indeed significantly contributed to the tumor mass expansion as driving factors commonly acknowledged. However, the R values were relatively low in numbers, suggesting that a simple linear model was insufficient to profile the complex behavior of cancer growth. Interestingly, COX-2 and CA IX gave larger R(obs) of 0.410 and 0.354 respectively than Ki-67 (R = 0.267). As the former is a marker partially reflecting the tumor environmental conditions, we postulated that environmental factors

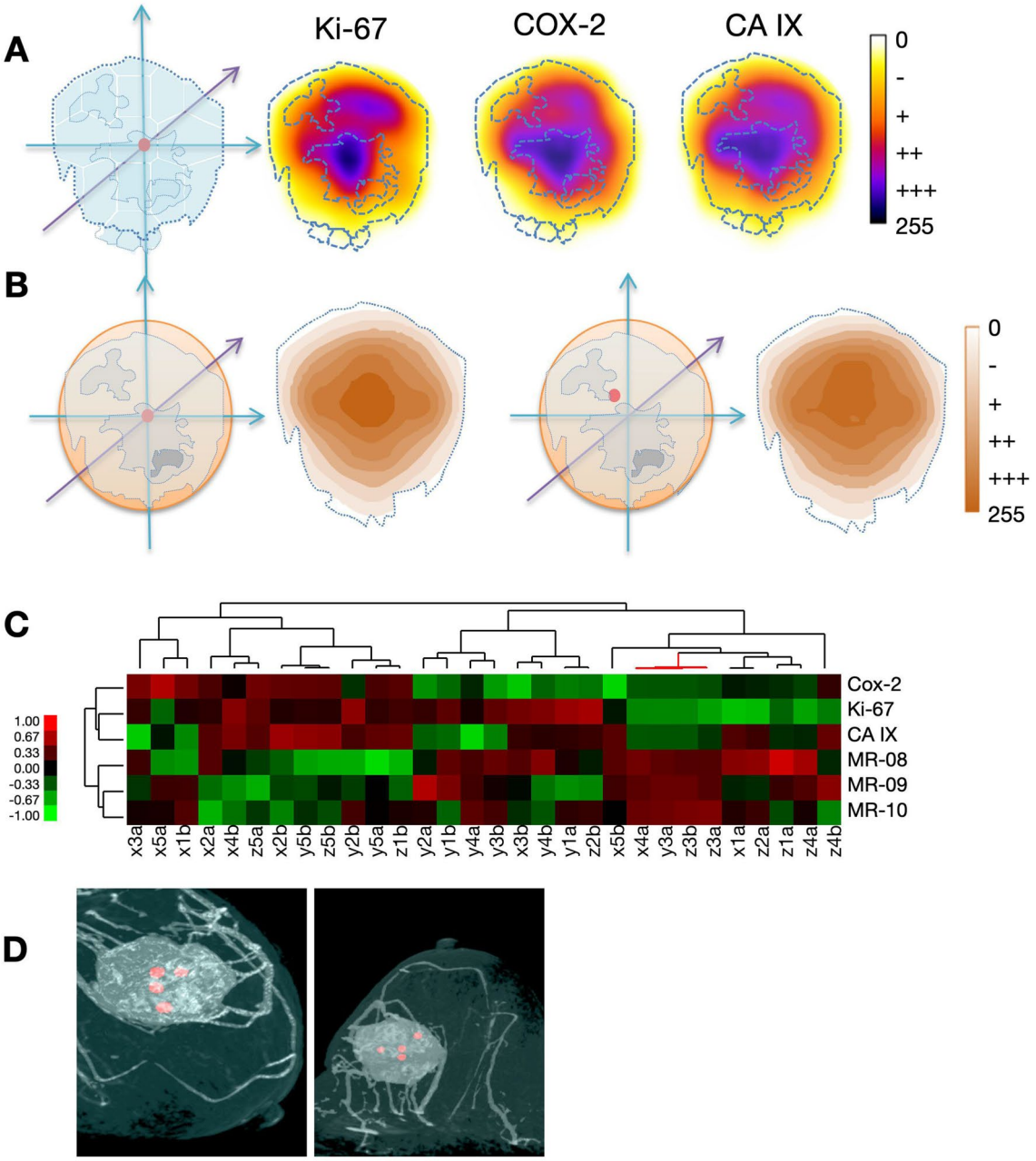

Fig. 2 Holistic view of cancer markers and comprehensive clustering analysis of multi-dimensional data. **a** The reconstructed 3D model for holistic view of Ki-67, COX-2, and CA IX. **b** The construction of the projected tumor growth model amplified from the anatomical center (left) and the density center readout from MRI images (right). **c** Cluster analysis based on multi-dimensional measurements from both imaging and pathology for representation, evaluation, and discovery of tumor heterogeneity. **d** Reconstructed image from vascular enhanced MRI images indicating the tumor mass with the overlayed marks of high heterogeneity sites as identified from **C**

Table 2 Correlation of Ki-67, COX-2, and CA IX distribution in reconstructed model

	Ki-67		COX-2		CA IX	
	R(obs)	*R*(rand) ± SD	*R*(obs)	*R*(rand) ± SD	*R*(obs)	*R*(rand) ± SD
Ki-67	1.00	–	0.626	0.061 ± 0.170	0.770	0.099 ± 0.184
COX-2			1.00	–	0.723	0.024 ± 0.181
CA IX					1.00	–

were unneglectable, including ones released from necrotic cells or the paracrine from adjacent tissues that could influence to a great extent cancer cell proliferation. For COX-2 which produced the best coefficients, it was known to closely relate to the oxygen supply and inflammatory responses requiring blood vessel infiltration (Fig. 2d). Therefore, it implied that a large tumor as its growth in size could be significantly related to the angiogenesis process.

3.4 Comprehensive clustering analysis of multidimensional data

Besides the main set of MRI series used for analyses in this study, we also obtained additional two medical digital imaging sets with different methods of enhancement from the patient. Combining with the pathology grading results of marker staining, multidimensional data grids can be formatted through normalization and subjected to cluster analysis (Fig. 2c). From such comprehensive analysis, the similarity and grouping of the characteristics at each anatomic site could be globally displayed and reviewed (Fig. 2d). The results showed that the sites at adjacent locations tended to be clustered at closer distances, which was expected in large and regular-shaped tumor masses. The Ki-67 associated with CA IX was better than with COX-2, which was also reasonable based on reports of their usage as cancer progression markers. These results were consistent with our previous correlational analyses on Ki-67, COX-2, and CA IX expression, where the signals of COX-2 were more heterogeneous and influenced by peripheral factors. An important readout from the cluster analysis (as marked as the branch in red color) was that the tumor heterogeneity could be manifested as the contrast in measured values between MRI density and marker staining intensity. The intratumoral heterogeneity of the identified sites could be developed surrounding the necrotic regions and did not necessarily have to be adjacent to each other. Here, this type of comprehensive statistical analysis demonstrated its great power for tumor heterogeneity studies, and the resolution and performance will be improved as more channels of data from either imaging or pathology parameters are provided.

3.5 Whole transcriptome sequencing for holistic views of a more complete gene expression profile

To further clarify factors that drive intratumoral heterogeneity, we performed whole transcriptome sequencing of 12 samples from different spatial positions within the breast tumor mass tissue (Fig. 3a). Unlike classic studies that used whole tumor tissue as experimental samples, all samples used for RNA-seq in the present study were obtained by puncturing. Therefore, to avoid operational errors caused by sampling, we first compared the Q30 values of RNA-seq data from the 12 samples as a quality control. According to the result, it was found that the Q30 values of all samples were above 89.98% (89.98%-91.83%), which indicates that the quality of all samples was above the standard for further analysis. Next, we performed a comparison of the global gene expression levels by mean FPKM value among all 12 samples to holistically demonstrate if there are any transcriptome-wide expression differences. The results indicate a significant intratumoral heterogeneity among 12 samples, the sample from the vertical-superficial-foot (ID: Y1) direction has the highest mean expression level (mean FPKM: 518.07) and the sample from sagittal-foot-medial (ID: Z1) direction has the lowest mean expression level (mean FPKM: 97.05).

During immunohistochemistry and staining intensity analysis, Ki-67, COX-2, and CA IX were used as common markers to obtain biological characteristics related to cell proliferation, inflammation, and hypoxia among samples. Based on the results of high-throughput sequencing data, we attempted to present high-dimensional data through pathway analysis to confirm intratumoral heterogeneity in the expression of genes related to cell proliferation, inflammation, and hypoxia. To achieve this, we searched the GSEA C2-Canonical pathways database and retrieved a total of 479 genes in the cell proliferation gene set, 368 genes in the inflammation gene set, and 117 genes in the hypoxia gene set. Since the *CA9* gene was already included in the hypoxia gene set, the *MKI67* gene was incorporated into our set of proliferation-related genes. Similarly, the *MT-CO2* gene was included in our set of inflammation-related genes for subsequent analysis. Both the *MKI67* and *MT-CO2* genes had been subject to examination in previous immunohistochemical (IHC) studies due to their established connections with cell proliferation and inflammatory processes, respectively. According to the pathway analysis results, there is significant heterogeneity among the 12 samples in terms of cell proliferation, inflammation, and hypoxia-related pathways. The sample with the highest expression level of genes in these three gene sets, indicating the most transcription activity, was from the vertical-superficial-foot direction (ID: Y1), with mean FPKM values of 1867.74, 4779.68, and 3706.79 for cell proliferation, inflammation, and hypoxia, respectively. In contrast, the sample with the least transcription activity was from the sagittal-foot-medial direction (ID: Z1), with mean FPKM values of 327.46, 748.34, and 660.43 for cell proliferation, inflammation, and hypoxia, respectively. (Fig. 3a, Supplementary Table 2–4). In addition to the pathway analysis, we also calculated the coefficient of variation to evaluate the axial asymmetry within each axis in large tumors globally and focused on three different pathway gene sets across samples in four directions on each axis. According to the global comparison results, the axial asymmetry

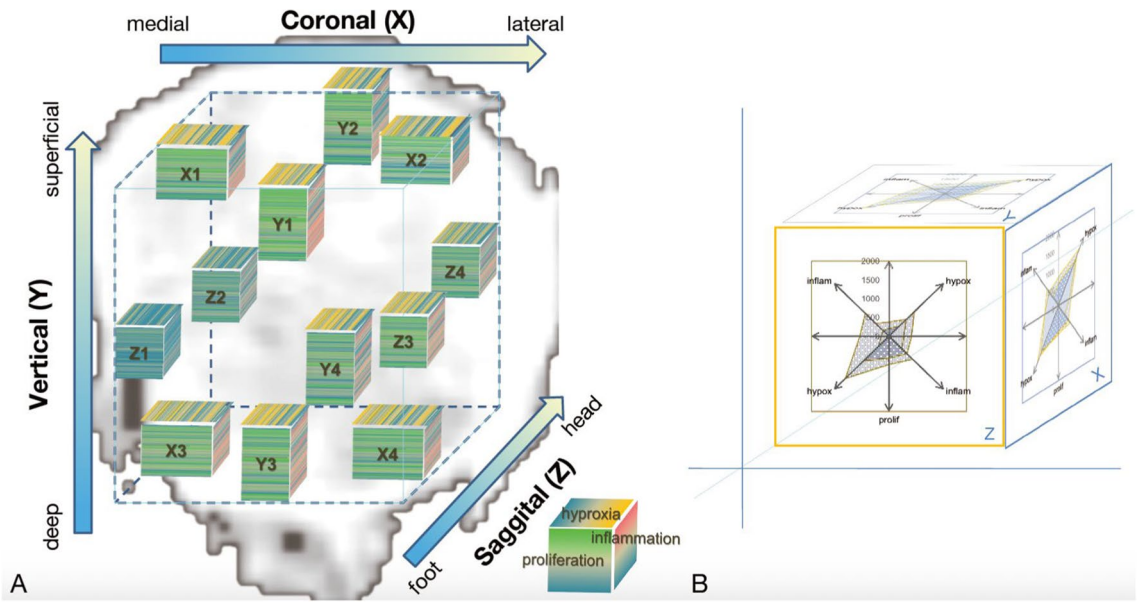

Fig. 3 Holistic view of cancer markers and related gene set analysis of multi-dimensional RNA-seq data. **a** The evaluation of expression heterogeneity of cell proliferation, inflammation, and hypoxia-related gene set within the tumor mass. Each cubic represent the location diagram of the samples selected for whole transcription sequencing including four samples selected from the coronal axis (X1: coronal-medial-superficial close to x1a to x3a in IHC assay, X2: coronal-lateral-superficial close to x3a to x5a, X3: coronal-medial-deep, close to x1b to x3b, X4: coronal-lateral-deep, close to x3b to x5b); four samples selected from vertical axis (Y1: vertical-superficial-foot, close to y1a to y3a in IHC assay, Y2: vertical-superficial-head, close to y3a to y5a, Y3: vertical-deep-foot, close to y1b to y3b, Y4: vertical-deep-head, close to y3b to y5b) and four samples selected from sagittal axis (Z1: sagittal-foot-medial, close to z1a to z3a in IHC assay, Z2: sagittal-head-medial, close to z3a to z5a, Z3: sagittal-foot-lateral, close to z1b to z3b, Z4: sagittal-head-lateral, close to z3b to z5b). The front view, vertical view, and side view of each cubic represents cell proliferation (480 genes), inflammation (369 genes), and hypoxia (117 genes) related gene set, respectively. The color of each dot was assigned by ranking the FPKM value of each gene within 12 samples using 3 color scale formatting (The top color represents values larger than 90%, the center color, represents middle values from 55%-90%, and the bottom color represents values from 10%-55%). Meanwhile, three key genes (*MKI67* gene, *MT-CO2* gene, and *CA9* gene) were highlighted. **b** The evaluation of axis asymmetry of three axes. The front view, vertical view, and side view of the cubic represent Z-axis direction, Y-axis direction, and X-axis direction of the original tumor mass. The radar charts' 3 radial axes respectively represent cell proliferation, inflammation, and hypoxia. Scales on each axis are labeled as FPKM values for the genes of each relevant gene set. The median expression level (FPKM value) of each relevant gene set for a sample is used to graph that sample

among the four samples on the sagittal (Z) axis shows the most variation (with a CV value of 51.91%), while the coronal (X) axis has the least variation (with a CV value of 13.15%). Consistent with the global gene expression results, the axial asymmetry among samples is most pronounced among the four samples on the sagittal (Z) axis (with CV values of 57.53%, 58.24%, and 63.01% for cell proliferation, inflammation, and hypoxia, respectively). These findings suggest that there is a significant degree of heterogeneity in the expression of genes related to cell proliferation, inflammation, and hypoxia both globally within the tumor and across different samples along the sagittal axis. (Fig. 3b and Supplementary Table 5). These findings suggest that there is considerable intratumoral heterogeneity and axial asymmetry in the expression of genes related to cell proliferation, inflammation, and hypoxia within the tumor samples which are consistent with the results obtained from the immunohistochemistry and staining intensity analysis. Meanwhile, the most significant degree of heterogeneity is present across different samples along the sagittal axis.

In addition, our study also compared the correlation among the pathways of proliferation, inflammation, and hypoxia. We found that the proliferation and hypoxia pathways showed the highest correlation coefficients (R = 0.955), suggesting that there may be an interplay between rapid tumor cell proliferation and hypoxia induction in the local environment of breast cancer tissue.

4 Discussion

Cancer is a devastating disease largely because of its complexity, which can be partly demonstrated by the heterogeneity of solid tumors. Heterogeneity can be developed as the tumor grows or rises from the resistance to chemotherapies. Understanding the mechanism of heterogeneity development and progression and applying the obtained knowledge to clinical diagnosis and prognosis is very important to improve the current cancer treatment outcome, ultimately benefiting many patients. Heterogeneity of tumors referred from many studies can be either intratumoral (within a tumor mass) or intertumoral (from the same or different patient), where the latter are better characterized, ever since the high throughput and large scale sequencing technology became available [22–24]. Dozens of biomarkers for cancer driver genes have been identified and a significant portion of which was found to significantly associate with heterogeneity development. However, it is generally agreed that intratumoral heterogeneity is more valuable in providing information linked to the early or persistent factors to drive heterogeneity since the same genetic background and environment exposure of cancerous cells helped to reduce the various influential factor that might affect tumor heterogeneity [25–27]. To this end, large-sized tumors are particularly valuable to be used for the studies of intratumoral heterogeneity. Unfortunately, with the rapid advancement in technology for cancer diagnosis and the aggression for early surgery, large tumors became more and more rare from clinical reports, especially for breast cancers of the preventive screens or routine examinations. In the present case, a basal-like invasive breast cancer with a mass of $4 \times 3.5 \times 2$ cm was extremely useful for research purposes. Besides, the patient was not treated by any interventions which made it perfect for the study of intratumoral heterogeneity based on the characteristics of the tumor tissues.

The digital processing of medical images with 3D reconstruction improves the examination of breast tumors at a better resolution with anatomic information than traditional mammography. MRI imaging can be used to display the signal intensity, edge shape, invasion range, and internal structures of a given tumor [26, 28]. It was reported that differences in structural components identified from MRI between breast tumors and normal breast tissues, or differences in pathohistological structures could link to the difference in relaxation time of T1 and T2 cancers [29]. Despite the low signal areas in MRI images sometimes clinically suggested necrosis of tumors, the density of reconstructed voxels was an accepted index to represent tumor mass in correlation with amounts of cancer cells and was used to indicate tumor growth [30]. In pathology, the growth and size expansion of solid tumor mass can be recognized by the staining intensity of certain cancer markers and predict tumor malignancy. One of the most used markers is Ki-67. Attempts to map Ki-67 distribution into tumor mass have been conducted. However, the comparison with other biomarkers was rarely reported. Currently in breast cancers, no previous investigation has revealed the correlation between cancer markers or tumor mass segments, especially in large-size tumors with the existence of heterogeneity. By introducing other markers of COX-2 and CA IX for their expression levels and correlating to the Ki-67 staining, we found that the asymmetric growth of the examined tumor was significantly under the influence of blood flow, inflammation, and/or hypoxia stimulation. By analyzing RNA-seq data, we could have a similar but transcriptome-wide perspective of understanding the role of cell proliferation, inflammation, and hypoxia-related genes/pathways in the drive of tumor heterogeneity. The comprehensive analyses of multiple markers representing different cancer-driving factors together with the anatomic 3D co-ordinance, as shown by the cluster analysis, enabled quick identification of heterogenic regions inside a tumor mass from a holistic view of the characteristics at whole tissue levels. None of our analyzed markers was likely to be a direct driver for the development of heterogeneity, thus, to expand future investigation by screening other possible causal molecules are necessary to better understand the mechanism of the generation of tumor heterogeneity. Nonetheless, we at least learned that common markers of cancer proliferation could be used to profile and evaluate tissue heterogeneity from this study.

The strength of using large-size tumors for the investigation of cancer heterogeneity is that it avoids the variance among patient individuals. Hypothetically, if progression heterogeneity obeys certain principles, the profiling of intratumoral heterogeneity will have a better chance of revealing it. Several considerations for making such practical attempts are, (1) careful selection of proper markers for examination; (2) the use of correlation analysis to produce numerical indexes for mathematical evaluation; (3) sufficient data in layers, dimensions, and volume to perform statistical analysis; and (4) the possibility to derive theoretical models that could be adjusted through recursive fitting to the measured realistic data. An interesting result from our study was that the environmental factors seemed to play a more important role in defining tumor heterogeneity than initially expected. One of the most prominent aspects could be the process of angiogenesis, which is closely connected to inflammation or hypoxia-related pathways and

turned out not quite as a surprise. This has raised a question about whether genetic alterations remained to be the real cause for heterogeneity development. If one agrees that change in tumor microenvironment is sufficient to induce heterogeneity, it still becomes troubling when tumor heterogeneity itself is widely heterogenic as observed, which does not seem reasonable among patients with the same diagnosis and receiving the same treatments. We argue that the interplay of genetic and environmental factors at the epigenetic level serves as the true driving force of tumor heterogeneity. Since the gene mutations in cancers are bulk in numbers, they tend to be averaged out for identification, thus leading to the genomic studies on cancer heterogeneity often facing tremendous challenges and resulting in vain. Another possibility is that the tumor microenvironment could be also quite different among individuals which was previously not realized [31]. Recent studies in high throughput secretomes started to reveal the variety from case to case and were suggested as a new aspect for consideration in precision medicine.

To evaluate the difference between tumor heterogeneity in general versus intratumoral heterogeneity in our special case, we extracted Ki-67, COX-2, and CA IX expression data from The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/tcga) for breast cancers and found that expression of these markers showed high variance in 849 cases (Supplementary Table 6). Notably, a lower correlation between these markers was found in the recursive tumors, especially of the TNBC type (Supplementary Table 7). All of which are more divergent than the results from our present study. This verified our presumption that intratumoral heterogene it indeed performs better in the mechanistic profiling for tumor heterogeneity. The biological characteristics and prognosis of breast cancer patients with the same histological type, similar differentiation, the same pathological stage, and even the same molecular typing may be different [10, 32]. Another critical point to be discussed is that heterogeneity is usually defined as a pathology term, whereas the current resolution in MRI technology does not reach a similar level for identifying masses in a few cells. It needs to be noted that the heterogeneity measures derived from the present study could be more suitable to apply for large tumors only. Furthermore, it should be highly noted that the inherent heterogeneity of breast tumors presents a significant challenge to treatment decisions based on biopsy or fine-needle aspiration (FNA) samples in the neoadjuvant setting and may also limit the outcomes of tumor microarrays in cancer marker validation studies due to the potential variability within the tumor. As observed during the preparation of tissue microarrays, it is imperative to conduct a comprehensive H&E examination of tissue sections prior to needle puncture sampling. This preliminary step is crucial for verifying whether the selected sampling points can represent the specificity of the entire tumor tissue. It is also likely that the heterogeneity we displayed might differ from molecular analysis by sequencing techniques. By all means, intratumoral heterogeneity deserves more attention, since it is not only in the clinical research and application but also in the research field of basic medicine, such as the origin, plasticity, and potency of mammary cell subsets.

By its nature, the present study is that it was only a retrospective case study. The limitations of our technical or management issues also restricted our research toward a more productive outcome. As the tumor tissues were primarily reserved for clinical pathological assays, we obtained only a limited amount in puncture forms. Ideally, if single-cell sequencing can be performed, it would provide more detailed information on the resolution at cell-type bases, which are frequently used to analyze heterogeneous or complex samples. As shown from several studies, additional trajectory analyses might implicate the possible dynamics of certain subpopulations of cells [33]. However, due to the technical challenges and budget concerns, we were only able to carry out conventional transcriptome sequencing. From our experience, this allowed us to be confident about sequencing data quality by Q30 measures, as well as coverage depth for subsequent pathway enrichment. Another challenge was dealing with the operational errors for reliable results from the 3D tumor mass reconstruction, model construction, and correlation analyses. Finally, the less completed clinical record filing and authority in ethical protocols contributed to the quality of our study as a limiting factor. The MRI images had been erased of the label information as a routine at the hospital imaging department, which was claimed to be an act for the protection of patient privacy. Consequently, professional software for MRI image processing and analysis could not be used without the incorporation of tumor feature information. Despite these limitations, the study provides valuable insights into the intratumoral heterogeneity and axial asymmetry of gene expression related to cell proliferation, inflammation, and hypoxia within breast cancer tissue. It also identifies the interplay between rapid tumor cell proliferation and hypoxia induction in the local environment in breast cancer tissue. The study highlights the importance of individual patient samples for in-depth and comprehensive analyses that can provide critical information to answer important clinical questions.

As a valuable lesson we learned from this study, we would like to emphasize once more the importance of special patient populations or unique clinical samples. In-depth and comprehensive analyses of such tissue samples will provide critical information to answer important clinical questions. Better documentation or follow-up studies and careful experimental design could transform the present investigation from retrospective to prospective, which would be far more beneficial to clinical practice. Furthermore, multi-dimensional comprehensive analyses of the

combined data from both laboratories and hospitals require consensed perspective views and adapted protocols for management. Continuous development needs to be conducted to improve the performance and application of such an approach demonstrated in the present study. If more sets of imaging data, such as PET, enhanced MRI, and SPECT can be joined with more comprehensive pathological information, such as sequencing results of DNA, RNA, and microRNA samples, it will add more weight to this type of case studies. Meanwhile, the design and production of special instruments adapted for such applications could soon be in immediate demand, such as puncture needles with enlarged diameter and scale labels or computer programs embedded with artificial intelligence for machine learning that accept data as provided in this study.

5 Conclusion

Although imaging technology has greatly facilitated the diagnosis of breast tumors, the clinical decision-making for cancer classification and selection of treatment plans mainly relies on the information provided by the pathology department. As the resolution of functional medical imaging and the automated throughput of molecular pathology are rapidly improving, the bridging of these two disciplines with shared and cross-referencing digital data will be extremely helpful in solving many current problems in cancer clinics, including the heterogeneity development during tumor progression. In this study, the expression of Ki-67, COX-2, and CA IX at the sampling sites of a large-size breast tumor was detected by immunohistochemistry and RNA-seq and then superimposed onto the 3D reconstructed tumor mass from MRI images. Simple mathematical models profiling the spatial distribution of cancer marker molecules were used to evaluate the intratumoral heterogeneity. Cluster analysis of multidimensional data from immunohistochemical grading of cancer pathology, pathway analysis data from whole transcriptome sequencing and MRI densitometry information allowed comprehensive representation of a holistic view explaining the synergy and regulation of multiple factors, such as cell proliferation, inflammation or hypoxia that could influence tumor growth. The present study has provided an example of possible generic approaches to investigate tumor biological characteristics using featured clinical tissue samples. We believe that continuous improvements of such protocols and expansion of their application will not only greatly support laboratory cancer-related studies but also may directly bring immediate benefit to patients and hospital doctors as well, particularly in making clinical decisions under the circumstances when significant tumor heterogeneity is observed.

Acknowledgements This work was supported by the National Natural Science Foundation of China (grant number NSFC 81572704) and the Beijing Municipal Natural Science Foundation (grant number 7222001).

Author contributions Zhicheng Ge: Data curation; Formal analysis; Resources; Writing—original draft; Jing Wang: Data curation; Formal analysis; Methodology; Writing—original draft; Libing He: Formal analysis; Methodology; Meng Zhao: Resources; Yang Si: Validation; Siyuan Chang: Visualization; Guoyan Zhang: Validation; Shan Cheng: Conceptualization; Funding acquisition; Supervision; Writing—review & editing. Wei Ding: Conceptualization; Funding acquisition; Supervision; Writing—review & editing.

Data availability The datasets generated during the current study are available in NCBI BioProject which including RNA-seq raw data. The accession number of RNA-seq data is PRJNA928368 and the corresponding sample information is summarized in Supplementary Table 8.

Declarations

Ethics approval and consent to participate The human tissue sample was archived materials obtained with the informed consent of the patient. Their use was approved by the Beijing Friendship Hospital Ethics Committees (2017-P2-006–01).

Competing interests The authors declare no competing interests.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7–34.
2. Padhani AR, Liu G, Koh DM, Chenevert TL, Thoeny HC, Takahara T, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia. 2009;11(2):102–25.
3. Yankeelov TE, Gore JC. Dynamic contrast enhanced magnetic resonance imaging in oncology: theory, data acquisition, analysis, and examples. Curr Med Imaging Rev. 2009;3(2):91–107.
4. Cooks T, Theodorou SD, Paparouna E, Rizou SV, Myrianthopoulos V, Gorgoulis VG, et al. Immunohisto (cyto) chemistry: an old time classic tool driving modern oncological therapies. Histol Histopathol. 2019;34(4):335–52.
5. Jafari SH, Saadatpour Z, Salmaninejad A, Momeni F, Mokhtari M, Nahand JS, et al. Breast cancer diagnosis: imaging techniques and biochemical markers. J Cell Physiol. 2018;233(7):5200–13.
6. Almendro V, Fuster G. Heterogeneity of breast cancer: etiology and clinical relevance. Clin Transl Oncol. 2011;13(11):767–73.
7. Zardavas D, Irrthum A, Swanton C, Piccart M. Clinical management of breast cancer heterogeneity. Nat Rev Clin Oncol. 2015;12(7):381–94.
8. Bolhaqueiro ACF, Ponsioen B, Bakker B, Klaasen SJ, Kucukkose E, van Jaarsveld RH, et al. Ongoing chromosomal instability and karyotype evolution in human colorectal cancer organoids. Nat Genet. 2019;51(5):824–34.
9. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. Cell. 2018;173(4):879–93.
10. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61–70.
11. Miquel-Cases A, Schouten PC, Steuten LM, Retel VP, Linn SC, van Harten WH. (Very) early technology assessment and translation of predictive biomarkers in breast cancer. Cancer Treat Rev. 2017;52:117–27.
12. Shackleton M, Quintana E, Fearon ER, Morrison SJ. Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell. 2009;138(5):822–9.
13. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res. 2007;13(15 Pt 1):4429–34.
14. de Mooij CM, Ploumen RAW, Nelemans PJ, Mottaghy FM, Smidt ML, van Nijnatten TJA. The influence of receptor expression and clinical subtypes on baseline [18F] FDG uptake in breast cancer: systematic review and meta-analysis. EJNMMI Res. 2023;13(1):5.
15. Urruticoechea A, Smith IE, Dowsett M. Proliferation marker Ki-67 in early breast cancer. J Clin Oncol. 2005;23(28):7212–20.
16. Munzone E, Botteri E, Sciandivasci A, Curigliano G, Nole F, Mastropasqua M, et al. Prognostic value of Ki-67 labeling index in patients with node-negative, triple-negative breast cancer. Breast Cancer Res Treat. 2012;134(1):277–82.
17. Wang J, Sang D, Xu B, Yuan P, Ma F, Luo Y, et al. Value of breast cancer molecular subtypes and Ki67 expression for the prediction of efficacy and prognosis of neoadjuvant chemotherapy in a Chinese population. Medicine. 2016;95(18): e3518.
18. Ghosh N, Chaki R, Mandal V, Mandal SC. COX-2 as a target for cancer chemotherapy. Pharmacol Rep. 2010;62(2):233–44.
19. Swietach P, Wigfield S, Cobden P, Supuran CT, Harris AL, Vaughan-Jones RD. Tumor-associated carbonic anhydrase 9 spatially coordinates intracellular pH in three-dimensional multicellular growths. J Biol Chem. 2008;283(29):20473–83.
20. Pastorek J, Pastorekova S. Hypoxia-induced carbonic anhydrase IX as a target for cancer therapy: from biology to clinical use. Semin Cancer Biol. 2015;31:52–64.
21. Chafe SC, Dedhar S. Carving out its niche: a role for carbonic anhydrase IX in pre-metastatic niche development. Oncoimmunology. 2015;4(12): e1048955.
22. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011;472(7341):90–4.
23. Aparicio S, Caldas C. The implications of clonal genome evolution for cancer medicine. N Engl J Med. 2013;368(9):842–51.
24. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med. 2012;366(10):883–92.
25. Martelotto LG, Ng CK, Piscuoglio S, Weigelt B, Reis-Filho JS. Breast cancer intra-tumor heterogeneity. Breast Cancer Res. 2014;16(3):210.
26. Hormuth DA 2nd, Jarrett AM, Lima E, McKenna MT, Fuentes DT, Yankeelov TE. Mechanism-based modeling of tumor growth and treatment response constrained by multiparametric imaging data. JCO Clin Cancer Inform. 2019;3:1–10.
27. Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, et al. A biobank of breast cancer explants with preserved intra-tumor heterogeneity to screen anticancer compounds. Cell. 2016;167(1):260–74.
28. Lima E, Oden JT, Wohlmuth B, Shahmoradi A, Hormuth DA 2nd, Yankeelov TE, et al. Selection and validation of predictive models of radiation effects on tumor growth based on noninvasive imaging data. Comput Methods Appl Mech Eng. 2017;327:277–305.
29. Baikeev RF, Gubanov RA, Sadikov KK, Safina SZ, Muhamadiev FF, Sibgatullin TA. Dynamic properties of water in breast pathology depend on the histological compounds: distinguishing tissue malignancy by water diffusion coefficients. BMC Res Notes. 2014;7:887.
30. O'Connor JP, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome. Clin Cancer Res. 2015;21(2):249–57.
31. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. Nature. 2013;501(7467):346–54.
32. Shipitsin M, Campbell LL, Argani P, Weremowicz S, Bloushtain-Qimron N, Yao J, et al. Molecular definition of breast tumor heterogeneity. Cancer Cell. 2007;11(3):259–73.
33. Wang Q, Sun K, Liu R, Song Y, Lv Y, Bi P, et al. Single-cell transcriptome sequencing of B-cell heterogeneity and tertiary lymphoid structure predicts breast cancer prognosis and neoadjuvant therapy efficacy. Clin Transl Med. 2023;13(8): e1346. https://doi.org/10.1002/ctm2.1346.