Integrative Clonal Evolution Analysis on SNV and CNV Levels in Multifocal Breast Cancer

Yan Gao

Tianjin Medical University Cancer Institute and Hospital

Ning Wu

Tianjin Medical University Cancer Institute and Hospital

Qingke Duan

Tianjin Medical University Cancer Institute and Hospital

Hong Liu

Tianjin Medical University Cancer Institute and Hospital

Bo Xu (✉ xubo@tmu.edu.cn)

Tianjin Medical University Cancer Institute and Hospital

Research Article

Keywords: multifocal breast cancer, single cell genome sequencing, clonal evolution, single nucleotide variations (SNVs), copy number variations (CNVs)

Posted Date: September 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-698940/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: For multifocal or multicentric breast cancer with two or more separated malignant foci in ipsilateral breast, the origin of its multiple foci is still debatable: do they share a common origin or have they developed independently?

Methods: By bulk exome sequencing combined with single cell whole genome sequencing using multiple annealing and looping-based amplification cycles (MALBAC), we obtained genomic information of two separated breast tumors and one metastatic lymph node in a multifocal breast cancer patient. We performed single nucleotide variation (SNV) and single cell copy number variation (CNV) analyses of the two breast tumors and lymph node.

Results: On SNV level, we found common origin and single focus metastasis in this patient, concretely speaking that the two breast tumors were formed by the spread of one single tumor to another, instead of being developed independently. Furthermore, the lymph node was metastasized from one of the two tumors instead of co-metastasis by both of them. We also found that the SNV subclone architecture always kept stable between the two breast tumors as well as from the breast tumors to lymph node. On CNV level, in contrast, the frequency of CNV subclones changed dramatically from the breast tumors to lymph node. Among them, frequency of Clone 2 increased significantly, indicating metastatic advantages of it. By combining distinct clonal evolution patterns of SNVs and CNVs, we built an integrative evolutionary model for this patient. CNV subclones were determined at a relative early time and kept unchanged, and CNV Clone 2 with metastatic advantages was separated from the common ancestors during intramammary spread and metastasis. In the meantime, SNVs were accumulated gradually.

Conclusions: In conclusion, we found distinct clonal evolution patterns on SNV and CNV levels in multifocal breast cancer and identified a CNV clone with metastatic advantages.

Background

Multifocal or multicentric breast cancer is defined as two or more malignant tumor foci, separated by normal mammary gland, presenting within the same quadrant or different quadrants in ipsilateral breast [1]. The incidence of multifocal and multicentric breast cancer is around 20% [2, 3]. Compared with unifocal breast cancer, multifocal or multicentric breast cancer has more axillary lymph node metastasis and worse clinical outcome [4–7]. For origin of multifocal or multicentric breast cancer two hypotheses exist: multiple foci may develop their own malignant tumors independently, or form through spread of one single malignant tumor. So far, several studies have tried to explain this on the genetic level [8–10]. One study based on targeted sequencing of different lesions from 36 multifocal or multicentric breast cancer patients found that 67% patients had mutations shared by all lesions and 33% patients did not share any mutations [8], indicating that both of the two hypotheses existed. In addition, the more axillary lymph node metastasis in multifocal or multicentric breast cancer implies that it may undergo a distinct process compared with unifocal breast cancer.

The origin and clonal evolution of multifocal or multicentric breast cancer can be studied on somatic genomic alterations. Single nucleotide variations (SNVs) and copy number variations (CNVs) are two major types of genomic alterations. SNVs are widely used in evolutionary history studies of cancer because it is commonly accepted that SNVs are accumulated gradually without reverse mutations. The phylogenetic trees of lung cancer, prostate cancer, renal cancer and other cancer types have been well established [11–14]. CNVs also play important roles during cancer genesis and development. It was reported that some tumors, belonging to the C class, were dominated by CNVs [15]. In breast cancer, CNV subclones within an individual have been exhibited [16, 17].

Heterogeneity is a key feature of breast cancer on both inter- and intra- patient levels [18, 19], especially for multifocal or multicentric breast cancer with multiple tumor foci. Due to heterogeneity it is difficult to study the subclone architecture and clonal evolution by bulk sequencing. Single cell sequencing can solve this problem, especially on CNV level. For SNVs, bulk sequencing shows a mixed result, and we can still decode the subclonal SNVs to some extent by bioinformatic method. But for CNVs, bulk sequencing can only gain the average level of copy numbers for each genomic region, making it almost impossible to obtain the subclonal CNVs. In 2011, the first single cell genome sequencing method for human cells, single-nucleus sequencing (SNS), was developed [16]. But the low genome coverage limited its application in many fields. After that another method called multiple annealing and looping-based amplification cycles (MALBAC) improved the coverage to 93% by using random priming and quasilinear preamplification [20].

In this study, we enrolled one patient diagnosed with multifocal breast cancer and collected samples from her right upper, outer breast tumors and metastatic axillary lymph node. We studied the subclone architecture and constructed phylogenetic tree on SNV level through bulk exome sequencing of the two breast tumors and lymph node. To further analyze clonal evolution from the breast tumors to lymph node on CNV level, we also performed single cell whole genome sequencing of total 125 cells using MALBAC. By integrating the results on both SNV and CNV levels, we revealed the origin and lymph node metastasis in this multifocal breast cancer patient.

Methods

Patient recruitment and sample collection

A multifocal breast cancer patient, female, 32 years old, was enrolled in our study. Three separated breast tumor foci were detected by CT images, including the right upper (12:00 direction, 6cm from the papilla, size 2×2cm), outer upper (10:00 direction, 4cm from the papilla, size 1.5×2cm) and outer (9:00 direction, 4cm from the papilla, size 2×2cm) breast tumors. All of the three breast tumors were diagnosed as Luminal B subtype, IIIA stage and invasive ductal carcinoma. Multiple right axillary lymph nodes were diagnosed positive. The frozen biopsy samples from the right upper, outer breast tumors and one right positive axillary lymph node (size 2×2cm) were collected. Before the sample collection, this patient had

not received systematic prior treatment. This study was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital with approval number bc2020003.

Frozen Tissue Dissociation And Single Nucleus Isolation

Half of each frozen biopsy sample was used for single cell sequencing. We dissociated the frozen samples into nucleus suspension following the published paper [16]. Tissue samples were first placed in the dish with 1-2ml of freshly prepared NST-DAPI buffer. After minced by surgical scissors, the samples were continually kept on ice for 15min to dissociate thoroughly. Then the nucleus suspensions were filtered through 70µm cell strainers. DAPI + single nuclei were sorted into 96 well plate by using BD FACSAria III system. 96 single nuclei from each sample (a total of 288 nuclei) were isolated.

Single Nucleus Whole Genome Amplification And Sequencing

Single nuclei were lysed in 5ul lysis buffer containing Triton X-100, DTT and QIAGEN protease. Then we performed whole genome amplification by using MALBAC following the published paper [20]. First, nine cycles of quasilinear amplification using random primers 5N3G and 5N3T were carried out. After that to gain enough DNAs for library preparation, 20 cycles of PCR amplification targeting the common sequence of 5N3G and 5N3T primers were followed. The quality of MALBAC products was checked by qPCR at eight randomly selected loci. Single nuclei with Ct number < 30 at no less than five loci passed the quality control. 44, 71 and 32 single nuclei from the upper, outer breast tumors and lymph node passed to the next step respectively.

DNAs from the selected single nucleus were used to construct libraries. First we used Covaris S220 system to fragment these DNAs to around 300bp. Then end repair, 3' dA tailing, adapter ligation and PCR enrichment were performed by using KAPA Hyper Prep Kit and NEBNext Multiplex Oligos for Illumina. The quality of libraries was checked by qPCR and fragment analysis using LabChip NGS3K. The qualified libraries were sequenced by Illumina Hiseq X Ten 2×150bp.

Single Cell Cnv Analyses

For raw sequencing data, quality control and MALBAC primer sequence removal were performed by using Trimmonatic-0.36 [21]. The quality control including removal reads containing adapter contamination, low quality nucleotides and unrecognizable nucleotide. Then clean reads were mapped to hg19 human reference genome by Burrows-Wheeler Aligner (BWA) 0.1.22 [22]. The CNV regions were identified with Control-FREEC [23]. Cells with complete deletion of large chromosome fragment (the whole chromosome or the whole chromosome arm) were identified to fail the single cell amplification, and cells with no obvious CNV regions were identified to be non-tumor cells. These cells were removed from the following

analyses. A total of 22 cells were removed, including four cells to fail the single cell amplification and 18 cells to be non-tumor cells.

Clustering on CNV level was performed using Euclidean distance method and complete cluster method at a bin size of 1M. To infer the clonal evolution process from the breast tumors to lymph node, we used TimeScape [24] by demonstrating the cellular prevalence changes of each CNV clone from the upper and outer breast tumors to lymph node.

Bulk Dna Extraction And Exome Sequencing

The left frozen biopsy samples were used for genomic DNA extraction and exome sequencing. Genomic DNAs were extracted by using QIAGEN QIAamp DNA Micro Kit following the manufacturer's instructions. We didn't get non-tumor tissue or blood samples from this patient, so in order to make a normal control for tumor samples to call somatic SNVs, we used the 18 single cells that were identified to be non-tumor cells by single cell CNV analysis (with no obvious CNV regions). We mixed the MALBAC amplified products of the 18 cells together as a control sample for exome sequencing.

For exome library preparation, DNAs were fragmented to around 180-280bp by Covaris S220 system, followed by exome region capture and exome library construction using Agilent SureSelect Human All Exon V6. First end repair, 3' dA tailing, half-length adapter ligation and PCR enrichment were performed. Then the DNAs were hybridized to the biotin labeled probes and magnetic beads with streptavidin were used to capture the exome regions. Finally, the captured DNAs were enriched by another PCR amplification. The quality of libraries was checked by qPCR and fragment analysis using LabChip NGS3K. The qualified libraries were sequenced by Illumina Hiseq X Ten 2×150bp.

Snv Analysis

Quality control of raw sequencing data was performed by using Trimmonatic-0.36 [21] to remove reads containing adapter contamination, low quality nucleotides and unrecognizable nucleotide. After that clean reads were mapped to hg19 human reference genome by BWA 0.1.22 [22]. The aligned reads were then sorted and merged with Samtools 1.0 [25] and the duplicated reads were marked by Sambamba v0.4.7. Somatic SNVs were called by using muTect 1.1.4 [26].

To predict the functional impacts of nonsynonymous SNVs, three different software were used: Mutation Assessor, SIFT and Polyphen-2. And protein annotations of representative SNVs were performed by MutationMapper with the link https://www.cbioportal.org/mutation_mapper.

The original phylogenetic tree was constructed by MEGA5 based on maximum parsimony algorithm [27]. The normal control sample was assigned to be outgroup. Then the tree was redrawn in Adobe Illustrator with trunk and branch lengths proportional to the number of nonsynonymous SNVs acquired on the corresponding trunk or branches. The trunk was colored in blue, the internal branch was colored in yellow

and the terminal branches were colored in red. To study the subclone architecture of SNVs, we used PyClone [28] to infer the cellular prevalence of SNVs and divide SNVs into different clusters. Then cellular prevalence changes of SNVs and SNV clusters from the outer breast tumor to the upper breast tumor or lymph node were demonstrated.

Results

SNV distribution among the two breast tumors and metastatic lymph node

One unanswered question for multifocal breast cancer is the clonal relationship of its multiple foci. For the patient enrolled in this study, the locations of the two separated breast tumors and one metastatic axillary lymph node were shown in Fig. 1a. Both bulk exome and single cell whole genome sequencing were performed. By bulk exome sequencing (with a DNA mixture from 18 non-tumor single cells as control), we detected 451, 525 and 441 somatic SNVs from the upper, outer breast tumors and lymph node respectively (Fig. 1b). Among them were 39, 38 and 40 nonsynonymous SNVs. Mutation types of these SNVs showed no significant differences among the two breast tumors and lymph node (Fig. 1c), indicating similar tumorigenesis mechanisms.

Then we focused on the nonsynonymous SNVs and drew a heatmap showing the distribution of 62 nonsynonymous SNVs (Fig. 1d). To find out the origin of the two breast tumors, we first compared nonsynonymous SNVs between the upper and outer breast tumors. For 54 nonsynonymous SNVs, 42.6% (23/54) were shared, and for 22 putative driver SNVs (also included in COSMIC database or KEGG: pathways in cancer), 86.4% (19/22) were shared. Such high percentages of shared SNVs indicated that the two breast tumors had a common origin. To study the clonal relationship of the metastatic lymph node with the two breast tumors, we compared nonsynonymous SNVs between the upper or outer breast tumors and lymph node separately. For the upper breast tumor and lymph node, among 58 nonsynonymous SNVs, 36.2% (21/58) were shared, and among 24 putative driver SNVs, 70.8% (17/24) were shared. For the outer breast tumor and lymph node, the percentages were much higher. Among 46 nonsynonymous SNVs, 69.6% (32/46) were shared, and among 21 putative driver SNVs, 81.0% (17/21) were shared. What's more, all of the nonsynonymous SNVs shared by the upper breast tumor and the lymph node also existed in the outer breast tumor. These results suggested that the lymph node was much more likely to come from the outer breast tumor, following the lymphatic drainage pathway.

To further understand how these nonsynonymous SNVs may affect the functions of protein, we performed the functional impact prediction by three methods: Mutation Assessor, SIFT and Polyphen-2 (Fig. 2a). Among all nonsynonymous SNVs, EPPK1 D2378H, which was shared by the two breast tumors and lymph node, were predicted highly deleterious by all of the three methods. Besides, PKDREJ C1134Y, LAGE3 L73F, VEPH1 A296D, ENG C182F, DIP2B N1321S, ZKSCAN2 G567D, etc. were also predicted deleterious to varying degrees.

For the 22 putative driver SNVs, NOTCH2 and TERT genes were also included in KEGG: pathways in cancer. Figure 2b showed their positions and upstream or downstream proteins in pathways in cancer. NOTCH2 gene encodes a member of Notch family, which plays a role in angiogenesis sustainment [29]. TERT gene encodes the protein component with reverse transcriptase activity of telomerase, which maintains telomere ends by addition of the telomere repeat [30]. Mutation of NOTCH2 were shared by the two breast tumors and lymph node, while mutation of TERT was specific to the upper breast tumor. 21/22 putative drivers SNVs (except for TERT gene) were also included in COSMIC database, and Fig. 2c showed protein annotations for 8 SNVs from 5 genes. For EPPK1 gene, R2239H, R2364Q and D2378H mutations affected its Plectin repeat domains, which are associated with intermediate filaments and cell adhesion. And for RFPL1 gene, I167V mutation affected its SPRY-associated domain, which is associated with innate immunity and inflammation. Mutations of EPPK1, NLRC5, NOTCH2 and RFPL1 genes were shared by the two breast tumors and lymph node, while mutation of PRPF38B gene was specific to the lymph node.

Phylogenic tree on SNV level confirms common origin and single focus metastasis

To clearly demonstrate the evolutionary relationships among the two breast tumors and lymph node, we constructed phylogenic tree based on nonsynonymous SNVs (Fig. 3a). In the phylogenic tree, the two breast tumors and lymph node shared a common ancestor, which confirmed common origin of the two breast tumors, just as indicated above. Besides, the outer breast tumor and lymph node shared a more recent ancestor, confirming single focus metastasis in this case. In other words, the lymph node metastasis came from the outer breast tumor instead of co-metastasis by both of the two breast tumors, the same as shown above. Majority (17/24) of the putative driver SNVs were trunk mutations, such as mutations of EPPK1, NLRC5, NOTCH2 and RFPL1 genes. While mutations of LRRC37A, TERT and PSG4 genes were branch mutations specific to the upper breast tumor, and mutations of PRB2 and PRPF38B genes were branch mutations specific to the lymph node.

The mutation frequency distribution in Fig. 3b further supported these conclusions. For the upper and outer breast tumors, more than half of the shared SNVs had close mutation frequencies, indicating their similarity. And compared with the upper breast tumor and lymph node, more of the shared SNVs between the outer breast tumor and lymph node had close mutation frequencies, meaning that they were more similar at SNV level.

To further investigate the subclone architecture of these nonsynonymous SNVs, we used PyClone to compare between the outer and upper breast tumors (Fig. 3c), and between the outer breast tumor and lymph node (Fig. 3d). For the two breast tumors, SNVs were divided into four clusters and the cellular prevalence of the four clusters kept stable from the outer to upper breast tumor. While for the outer breast tumor and lymph node, the subclone architecture of SNVs became more complex, with a total of five SNV clusters divided. And this complex subclone architecture (or the cellular prevalence of the five clusters) still kept stable from the outer breast tumor to the lymph node. In general, no matter during separation of

the two breast tumors or lymph node metastasis, the SNV subclone architecture still maintained unchanged, without presence of any advantageous or disadvantageous SNV subclones.

CNV patterns of single cells from the two breast tumors and lymph node

Besides bulk exome sequencing, in the meanwhile we performed low depth single cell whole genome sequencing of the two breast tumors and lymph node for CNV analysis. A total of 147 single cells were sequenced. Among them 4 cells didn't pass the quality control and 18 cells were identified as non-tumor cells. Sequencing data from 40, 53 and 28 single cells (a total of 125 cells) in the upper, outer breast tumors and lymph node were used in the following analysis. CNVs of the 125 single cells were shown in the heatmap in Fig. 4a. Clustering based on whole genome CNVs divided all single cells (without regard to origins) into three clones. Clone 1 and Clone 2 were major clones, accounting for 44% (55/125) and 51.2% (64/125) of all cells, and Clone 3 was a minor clone, accounting for only 4.8% (6/125). Majority of cells in Clone 1 were from the outer breast tumor, and the left few cells were from the lymph node. For Clone 2 cells were from the upper breast tumor and the lymph node. And Clone 3 included cells from the outer breast tumor and the lymph node. In this respect, compared with the upper breast tumor, the outer breast tumor may be more similar to the lymph node.

Then to find the characteristics of each clone, we draw the CNV patterns for the three clones in Fig. 4b. Each clone had its specific CNV characteristics. Except for Clone 3, which showed only a few CNV regions, Clone 1 and Clone 2 shared some CNV regions, for example, copy number gains at chr8q, 17q, 20q and 21q, and copy number losses at chr6q, 8p, 11q, 14q and 17p. Clone 1 and Clone 2 also had their specific CNV regions. Specific CNV regions of Clone 1 included copy number gains at chr1q and 16p, and copy number losses at chr10q, 16q, 17p and X. Specific CNV regions of Clone 2 included copy number losses at chr1p, 2q, 3p, 4p, 4q, 5q, 7p, 7q, 9p and 18q.

Clonal evolution on CNV level identifies a subclone with metastatic advantages

To reveal the CNV level clonal evolution from the breast tumors to lymph node, we used TimeScope to demonstrate this process. As shown in Fig. 4c, from the breast tumors to lymph node, the frequency of Clone 3 kept stable, the frequency of Clone 1 decreased, and the frequency of Clone 2 increased significantly. This indicated that compared with Clone 1 and Clone 3, Clone 2 had more metastatic advantages.

Then we analyzed the specific CNV regions of Clone 2 and putative driver genes located in these regions, as marked in Fig. 4b. In chr1p, with copy number losses in Clone 2, there were many famous cancer related genes, such as ARID1A, MTOR and JAK1. ARID1A encodes a member of SWI/SNF family, which is required for transcriptional activation of genes normally repressed by chromatin [31]. MTOR encoded protein belongs to a family of phosphatidylinositol kinase-related kinases, which mediates cellular responses to stresses such as DNA damage and nutrient deprivation [32]. JAK1 encoded protein is a member of protein-tyrosine kinases, playing a crucial role in effecting the expression of genes that mediate inflammation, epithelial remodeling and metastatic cancer progression [33]. Chr5q, with copy

number losses in Clone 2, also contained many important genes associated with cancer, such as APC, MAP3K1 and PIK3R1. APC encodes a tumor suppressor protein that acts as an antagonist of the Wnt signaling pathway [34]. MAP3K1 encoded protein is a serine/threonine kinase and is part of some signal transduction cascades, including the ERK and JNK kinase pathways as well as the NF-kappa-B pathway [35]. PIK3R1 encodes the regulatory subunit of phosphatidylinositol 3-kinase, which plays an important role in the metabolic actions of insulin [36]. Besides there were many other putative driver genes in Clone 2 specific CNV regions, for example ACVR2A, CXCR4 and LRP1B genes in chr2q with copy number losses, PHOX2B, RHOF and SLC34A2 genes in chr4p with copy number losses, CASP3, FAT1 and FBXW7 genes in chr4q with copy number losses, and BCL2, SETBP1, SMAD2 and SMAD4 genes in chr18q with copy number losses. The copy number gains or losses of these regions or genes had a high probability to be associated with lymph node metastasis.

Discussion

In our study, on SNV level the lymph node shared more SNVs with the outer breast tumor compared with the upper breast tumor. While on CNV level, the lymph node was more similar to the upper breast tumor, inconsistent with SNVs. Previous studies proved that in breast cancer CNVs and SNVs evolved in different patterns [37]. CNVs occurred early in tumor evolution and remained highly stable during tumor development. In contrast, SNVs evolved gradually. In our data we also observed distinct evolutionary patterns of SNVs and CNVs. On SNV level, no matter between the outer and upper breast tumor, or from the outer one to lymph node, the subclone architecture always kept stable. But on CNV level, from the breast tumors to lymph node, the frequency of Clone 1 significantly decreased and the frequency of Clone 2 significantly increased, indicating that Clone 2 may have metastatic advantages.

To explain this discordance between SNVs and CNVs, we combined the clonal evolution patterns of both SNVs and CNVs to build an integrative evolutionary model in this patient. As shown in Fig. 4d, CNVs occurred at a relative early time, and three different CNV patterns generated before the two breast tumors and lymph node separated. Then upon intramammary spread, part of CNV Clone 2 with metastatic advantages were separated to form the upper breast tumor. After that occurred the lymph node metastasis, the left CNV Clone 2 metastasized as the majority of the lymph node. And other CNV clones without metastatic advantages, especially Clone 1, were left to form the outer breast tumor. In the meanwhile, SNVs accumulated gradually, with trunk SNVs occurred before separation and branch SNVs occurred after intramammary spread or lymph node metastasis. In conclusion, similarity between the lymph node and outer breast tumor on SNV level was because they were evolutionarily closer and the lymph node may just come from the outer breast cancer. While similarity between the lymph node and upper breast tumor on CNV level was because this kind of CNV pattern had metastatic advantages and both of the lymph node and upper breast tumor were formed by spread of tumor cells with this CNV pattern.

The results above exhibited that from the breast tumors to lymph node, the SNV subclone architecture maintained unchanged without any SNV subclones showing metastatic advantages. While frequency of

CNV subclones changed obviously, with Clone 2 increased significantly, suggesting its metastatic advantages. In previous studies based on circulating tumor cells from many different cancer types, including lung cancer, breast cancer, gastric cancer, colon cancer and prostate cancer, the researchers also validated the importance of CNVs in cancer metastasis [38, 39]. They found a strong CNV selection toward metastasis, meaning that only cells with certain specific CNV patterns could be selected to form the metastasis of cancer. For genetic studies of cancer, most researchers mainly focused on SNVs to find molecular biomarkers for targeted therapy. These results above suggested us CNVs may also play important roles in cancer, especially in cancer metastasis.

Compared with bilateral breast cancer, in multifocal or multicentric breast cancer the distances among multiple tumor foci were much shorter. And multiple tumor foci in majority multifocal or multicentric breast cancer patients shared similar molecular subtypes, clinical pathologic staging and classification [40]. In our study, both of the two breast tumors were diagnosed as Luminal B subtype, IIIA stage and invasive ductal carcinoma. Our genetic analyses on SNV level further confirmed common origin of them. But our conclusion was limited by number of cases (only one patient was enrolled) and may not be suitable for all multifocal or multicentric breast cancer patients. Previous studies based on a larger cohort of multifocal or multicentric breast cancer patients (36 patients) exhibited that more than half (67%) of the patients had common origins and the left (33%) developed independently. While in bilateral breast cancer there may be a complex combination of different morphologic characteristics and hormonal receptor statuses [41]. And a genetic study based on four bilateral breast cancer patients found that they were all clonal independent [42]. These suggested the differences between multifocal or multicentric breast cancer and bilateral breast cancer, which deserved further investigation.

Conclusions

In conclusion, by bulk exome sequencing combined with single cell whole genome sequencing, we studied SNVs and single cell CNVs of two separated breast tumors and one metastatic lymph node and built an integrated evolutionary model in a multifocal breast cancer patient. In addition, we discovered distinct evolutionary patterns of SNVs and CNVs and CNV subclone with metastatic advantages.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital with approval number bc2020003.

Consent for publication

The patient gave consent for publication of this study.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (grant numbers 81974464, 81903124 and 81672743) and Beijing Tianjin Hebei Basic Research Cooperation Project (grant number 19JCZDJC64500(Z)).

Authors' contributions

The study was designed by BX and YG. The patient recruitment and clinical sample collection were performed by HW. The single cell sequencing experiments were performed by YG, NW and QD. The bioinformatics analyses were done by YG. All of the authors joined in the manuscript writing and editing.

Acknowledgements

We are grateful to Prof. Ping Zhu, Yinan Zhang and Chao Tang in Chinese Academy of Medical Sciences and Peking Union Medical College for assistance in the management of sequencing data. We thank all the members of Prof. Bo Xu's lab for their cooperation and help.

References

1. Salgado R, Aftimos P, Sotiriou C, Desmedt C. Evolving paradigms in multifocal breast cancer. *Semin Cancer Biol.*2015;31:111-118.
2. Lynch SP, Lei X, Chavez-MacGregor M, Hsu L, Meric-Bernstam F, Buchholz TA, et al. Multifocality and multicentricity in breast cancer and survival outcomes. *Ann Oncol.*2012;23:3063-3069.
3. Wolters R, Wockel A, Janni W, Novopashenny I, Ebner F, Kreienberg R, et al. Comparing the outcome between multicentric and multifocal breast cancer: what is the impact on survival, and is there a role for guideline-adherent adjuvant therapy? A retrospective multicenter cohort study of 8,935 patients. *Breast Cancer Res Treat.*2013;142:579-590.
4. Pekar G, Hofmeyer S, Tabar L, Tarjan M, Chen TH, Yen AM, et al. Multifocal breast cancer documented in large-format histology sections: long-term follow-up results by molecular phenotypes. *Cancer.*2013;119:1132-1139.
5. Weissenbacher TM, Zschage M, Janni W, Jeschke U, Dimpfl T, Mayr D, et al. Multicentric and multifocal versus unifocal breast cancer: is the tumor-node-metastasis classification justified? *Breast Cancer Res Treat.*2010;122:27-34.

6. Rezo A, Dahlstrom J, Shadbolt B, Rodins K, Zhang Y, Davis AJ, et al. Tumor size and survival in multicentric and multifocal breast cancer. *Breast*.2011;20:259-263.
7. Tot T, Gere M, Pekar G, Tarjan M, Hofmeyer S, Hellberg D, et al. Breast cancer multifocality, disease extent, and survival. *Hum Pathol*.2011;42:1761-1769.
8. Desmedt C, Fumagalli D, Pietri E, Zoppoli G, Brown D, Nik-Zainal S, et al. Uncovering the genomic heterogeneity of multifocal breast cancer. *J Pathol*.2015;236:457-466.
9. Schwartz CJ, Dolgalev I, Vasudevaraja V, Kelly S, Heguy A, Snuderl M, et al. Revisiting multifocal breast cancer: a clonality study of ductal carcinoma using whole exome sequencing. *Hum Pathol*.2019;94:71-77.
10. Chunder N, Roy A, Roychoudhury S, Panda CK. Molecular study of clonality in multifocal and bilateral breast tumors. *Pathol Res Pract*.2004;200:735-741.
11. 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;366:883-892.
12. Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat Genet*.2014;46:225-233.
13. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science*.2014;346:256-259.
14. Gundem G, Van Loo P, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, et al. The evolutionary history of lethal metastatic prostate cancer. *Nature*.2015;520:353-357.
15. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet*.2013;45:1127-1133.
16. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. *Nature*.2011;472:90-94.
17. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet*.2016;48:1119-1130.
18. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*.2000;406:747-752.
19. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*.2012;486:395-399.
20. Zong C, Lu S, Chapman AR, Xie XS. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science*.2012;338:1622-1626.
21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*.2014;30:2114-2120.
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*.2009;25:1754-1760.

23. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*.2012;28:423-425.
24. Smith MA, Nielsen CB, Chan FC, McPherson A, Roth A, Farahani H, et al. E-scape: interactive visualization of single-cell phylogenetics and cancer evolution. *Nat Methods*.2017;14:549-550.
25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*.2009;25:2078-2079.
26. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*.2013;31:213-219.
27. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*.2011;28:2731-2739.
28. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J, et al. PyClone: statistical inference of clonal population structure in cancer. *Nat Methods*.2014;11:396-398.
29. Zhao D, Xue C, Lin S, Shi S, Li Q, Liu M, et al. Notch Signaling Pathway Regulates Angiogenesis via Endothelial Cell in 3D Co-Culture Model. *J Cell Physiol*.2017;232:1548-1558.
30. Yuan X, Larsson C, Xu D. Mechanisms underlying the activation of TERT transcription and telomerase activity in human cancer: old actors and new players. *Oncogene*.2019;38:6172-6183.
31. Mathur R. ARID1A loss in cancer: Towards a mechanistic understanding. *Pharmacol Ther*.2018;190:15-23.
32. Ma Y, Vassetzky Y, Dokudovskaya S. mTORC1 pathway in DNA damage response. *Biochim Biophys Acta Mol Cell Res*.2018;1865:1293-1311.
33. Chang Q, Bournazou E, Sansone P, Berishaj M, Gao SP, Daly L, et al. The IL-6/JAK/Stat3 feed-forward loop drives tumorigenesis and metastasis. *Neoplasia*.2013;15:848-862.
34. Aoki K, Taketo MM. Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene. *J Cell Sci*.2007;120:3327-3335.
35. Xue Z, Vis DJ, Bruna A, Sustic T, van Wageningen S, Batra AS, et al. MAP3K1 and MAP2K4 mutations are associated with sensitivity to MEK inhibitors in multiple cancer models. *Cell Res*.2018;28:719-729.
36. Winnay JN, Solheim MH, Dirice E, Sakaguchi M, Noh HL, Kang HJ, et al. PI3-kinase mutation linked to insulin and growth factor resistance in vivo. *J Clin Invest*.2016;126:1401-1412.
37. Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*.2014;512:155-160.
38. Ni X, Zhuo M, Su Z, Duan J, Gao Y, Wang Z, et al. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. *Proc Natl Acad Sci U S A*.2013;110:21083-21088.

39. Gao Y, Ni X, Guo H, Su Z, Ba Y, Tong Z, et al. Single-cell sequencing deciphers a convergent evolution of copy number alterations from primary to circulating tumor cells. *Genome Res.*2017;27:1312-1322.

40. Tot T. Early and more advanced unifocal and multifocal breast carcinomas and their molecular phenotypes. *Clin Breast Cancer.*2011;11:258-263.

41. Kurian AW, McClure LA, John EM, Horn-Ross PL, Ford JM, Clarke CA. Second primary breast cancer occurrence according to hormone receptor status. *J Natl Cancer Inst.*2009;101:1058-1065.

42. Song F, Li X, Song F, Zhao Y, Li H, Zheng H, et al. Comparative genomic analysis reveals bilateral breast cancers are genetically independent. *Oncotarget.*2015;6:31820-31829.

Figures

Figure 1

SNV landscape of the two separated breast tumors and one metastatic axillary lymph node in a multifocal breast cancer patient. aA diagram shows the locations of samples collected in this study, including the right upper (12:00 direction, 6cm from the papilla), outer (9:00 direction, 4cm from the papilla) breast tumors and lymph node. b Numbers of somatic SNVs detected in the upper (Upper), outer (Outer) breast tumors and lymph node (LN). c Mutation types of somatic SNVs in the two breast tumors and lymph node. d A heatmap shows the distribution of 62 nonsynonymous SNVs in the two breast tumors and lymph node. Gene names are listed on the left and the putative driver SNVs (also included in COSMIC database or KEGG: pathways in cancer) are in red.

Figure 2

Functional analyses of nonsynonymous SNVs. a Functional impact prediction of 62 nonsynonymous SNVs by three methods: Mutation Assessor, SIFT and Polyphen-2. bThe positions of NOTCH2 and TERT genes in KEGG: pathways in cancer with mutation frequencies in the two breast tumors and lymph node (red: the upper breast tumor, yellow: the outer breast tumor, green: the lymph node). c Interpretation of representative putative driver SNVs for EPPK1, NLRC5, NOTCH2, RFPL1 and PRPF38B genes with protein annotations.

Figure 3

Inferred clonal evolution on SNV level among the two breast tumors and lymph node. a Phylogenetic tree of the upper, outer breast tumors and lymph node. Blue line represents trunk, green line represents internal branch, and red lines represent terminal branches. The lengths of trunk and branches are proportional to the numbers of acquired nonsynonymous SNVs. Representative putative driver SNVs are labeled beside the trunk and branches. b Mutation frequency distribution of nonsynonymous SNVs between the upper and outer breast tumors (left), the upper breast tumor and lymph node (middle), and the outer breast tumor and lymph node (right). c Inferred subclone architecture changes from the outer to upper breast tumors. The plot on the left shows cellular prevalence of each SNV cluster. The plot on the right shows cellular prevalence of each SNV included in different clusters with gene names labeled beside (genes in red existing in both c and d). d Inferred subclone architecture changes from the outer breast tumor to lymph node.

Figure 4

Clonal evolution on CNV level among the two breast tumors and lymph node. aA heatmap shows whole genome CNVs of 125 single cells from the two breast tumors and lymph node with CNV clustering on the left. All single cells can be divided into three subclones according to CNVs. bConsensus CNV pattern of each subclone. CNV regions specific to Clone 2 are squared, with red squares representing copy number gain regions and blue squares representing copy number loss regions. The representative driver genes included in the CNV regions are listed below. c Clonal evolution from the breast tumors (the upper and outer breast tumors) to lymph node on CNV level with clonal frequencies labeled. dAn integrative clonal evolution model on both SNV and CNV levels of the two breast tumors and lymph node in this multifocal breast cancer patient.