Progression potential of ductal carcinoma in situ assessed by genomic copy-number profiling

Mina Kitamura[1,2], Takahisa Nakayama[1], Ken-ichi Mukaisho[1], Tsuyoshi Mori[2], Tomoko Umeda[2], Suzuko Moritani[3], Ryoji Kushima[3], Masaji Tani[2], Hiroyuki Sugihara[1*]

[1]Division of Molecular Diagnostic Pathology, Department of Pathology, Shiga University of Medical Science; [2]Division of Digestive, Breast and General Surgery, Department of Surgery, Shiga University of Medical Science; [3]Division of Diagnostic Pathology, Shiga University of Medical Science Hospital, Otsu, 520-2192 Japan

[*]Correspondence: sugihara@belle.shiga-med.ac.jp

[1]Division of Molecular Diagnostic Pathology, Department of Pathology, Shiga University of Medical Science, Otsu, 520-2192 Japan

Footnote: The present study was supported in part by JSPS KAKENHI Grant Number JP16K08689.

Abstract

Background: Ductal carcinoma in situ (DCIS) of breast is heterogeneous in terms of the risk of progression to invasive ductal carcinoma (IDC). To treat DCIS appropriately for its progression risk, we classified individual DCIS by its profile of genomic changes into 2 groups and correlated them with clinicopathological progression factors.

Methods: We used surgically resected, formalin-fixed, paraffin-embedded tissues of 22 DCIS, and 30 IDC lesions. We performed immunohistochemical intrinsic subtyping, array-based comparative genomic hybridization and unsupervised clustering. ***Results:*** The samples were divided into 2 major clusters, A and B. Cluster A showed a greater number of gene and chromosome copy-number alterations, a larger IDC/DCIS ratio, a higher frequency of non-luminal subtype, a lower frequency of luminal subtype, and a higher nuclear grade, when compared with cluster B. But, there was no difference in the frequencies of lymph node metastasis between clusters A and B. We identified 9 breast-cancer related genes, including *TP53* and *GATA3*, that highly contributed to the discrimination of A and B clusters. ***Conclusion:*** Classification of breast tumors into rapidly progressive cluster A and the other (cluster B) may contribute to select the treatment appropriate for their progression risk.

Keywords

Copy-number alterations, array-based comparative genomic hybridization, breast cancer, ductal carcinoma in situ, papilloma, unsupervised hierarchical clustering

Introduction

Ductal carcinoma in situ (DCIS) is characterized by noninvasive spreading within mammary ducts. Whether DCIS inevitably becomes invasive and what factors determine the rate of progression remains unclear.

In 2011, breast cancer was newly detected in approximately 48,000 Japanese women, including 14.2% with DCIS, whereas the percentage of DCIS was only 8.2% in 2004. This increase in the proportion of DCIS may reflect an increase in the number of breast cancer patients who were detected by mammographic screening [1]. Mammographic screening is more prevalent in North America than in Japan, and the number of DCIS cases detected by screening accounts for nearly 20% of all breast cancers in recent years [2]. This increase in DCIS detection is likely to continue until it reaches the prevalence of latent DCIS, which is as high as 8.9% of autopsy cases [3].

If DCIS inevitably becomes invasive ductal carcinoma (IDC), the screening-based detection of DCIS can lead to a reduction in the mortality of breast cancer patients.

Many studies reported that screenings had actually reduced breast cancer mortality but that the reduction size was smaller than expected probably because of overdiagnosis of very dormant tumors as cancers [4, 5]. A retrospective cohort study demonstrated that there was no significant difference in the weighted hazard ratios of breast cancer–specific 10-year survival between surgery and non-surgery groups for low-grade DCIS [6]. Accordingly, a follow up study for up to 20 years after a biopsy diagnosis of DCIS without subsequent treatment reported that 28% of these patients recurred for IDC within approximately 15 years [7]. This study and other similar reports [3, 8] suggest that some DCIS lesions remain dormant and have a very slow progression rate to IDC. It is also possible that some DCIS lesions do not become invasive.

Molecular and epidemiological data indicate that breast cancer development is a multi-step process [9]. The progression from DCIS to IDC may involve stepwise genetic alterations [10, 11]. Some studies have suggested that copy number aberrations (CNAs) are associated with the progression from DCIS to IDC, including amplifications of *MYC*[12], *FGFR1* [13], and *CCND1* [14], which were more frequently observed in IDC than DCIS. However, other studies comparing the DCIS and invasive components from the same patient demonstrated that CNAs at the chromosome level were very similar between intraductal and invasive components [15]. Analogous approaches

searching for common transcriptomic and/or genomic differences between matched synchronous DCIS and IDC have been unsuccessful [16, 17]. Current studies have also failed to identify driver genes that play a significant role in the transition from DCIS to IDC [2]. Thus, the presence of progression-related genomic changes and overall genomic similarity between DCIS and IDS requires further study.

In the present study, we focused on the CNA profiles of genomic DNA using array-based comparative genomic hybridization (CGH). These profiles are unique for individual neoplasms because they include random gene alterations that have a neutral role for carcinogenesis, accumulate over time based on genetic instability, and are selected by the tissue microenvironment. Approximately three-quarters of the natural history of solid cancers has elapsed once tumors reach 1 cm in diameter [18], which leaves a limited opportunity for the accumulation of additional genomic changes. Thus, the CNA profile of clinically detectable breast cancer may already include information enough for outcome prediction. In our previous gastric cancer studies, we classified the samples based on their CNA profiles using unsupervised hierarchical cluster analyses, and demonstrated that nearly all early cancer of the undifferentiated type can become advanced [19], whereas approximately 80% of non-invasive neoplasm of the differentiated type showed a lineage distinct from advanced cancers [20].

Applying the similar approach to breast cancers, in the present study, we have found that breast tumors, including DCIS, IDC and papillomas, were classified into rapidly progressive group and slowly progressive group. This classification may contribute to select the treatment of individual breast tumors appropriate for their progression risk.

Methods

Patients

The study consisted of 50 patients who underwent a partial or total mastectomy for DCIS lesions (n = 22) or IDCs (n = 30; 15 T1, 14 T2, and 1 T4 tumors) from December 2009 to January 2014 (Supplementary file 1). Two patients had two tumors: 1 patient (#13) with bilateral IDC and DCIS and 1 patient (#17) with unilateral IDC and DCIS in areas A and C, respectively. All patients were female. The mean age was 55.2 years old (range, 35–84 years). None of the patients received any preoperative radio- and/or chemotherapy. The conduct in this study was approved by the Institutional Review Board at the Shiga University of Medical Science on the condition that the materials used remained anonymous (Permission number: 26–36 on July 24, 2014). Written informed consent was not required in this retrospective study because of the use of archival materials and detected acquired genomic copy-number changes.

Tissue samples

We used formalin-fixed, paraffin-embedded (FFPE) tissues. Tissues were fixed in buffered 10% formalin for 24 to 48 hours. In 20 of the 30 IDCs, DNA samples were taken from both ductal and invasive components. DNA samples of metastatic tumors in lymph nodes (LNs) were available in 7 cases (Supplementary file 1). Only patient #18 had distant metastasis, from which no sample was available. N2 or N3 nodal metastasis was not detected in any of the patients in this study. .

Immunohistochemistry and in situ hybridization

We used 3 μm-thick tissue sections for the immunohistochemical (IHC) analysis of the estrogen receptor (ER), progesterone receptor (PgR), HER2, basal and myoepithelial markers (cytokeratin (CK) 5/6 and p63), and Ki-67. We used the following antibodies: anti-ER (clone 1D5, DAKO, Santa Clara, CA, USA; dilution 1:50), anti-PgR (clone PgR636, DAKO; dilution 1:50), anti-Ki-67 (clone MM1, Leica Biosystems Newcastle Ltd, Newcastle Upon Tyne, UK; dilution 1:100), anti-HER2 (clone 4B5, Ventana, Tucson, AZ, USA; pre-diluted), and anti-CK5/6 and anti-p63 (based on a previous study) [21]. IHC was optimized by evaluating serial sections with multiple

antibody concentrations using a Discovery Automated Immunostainer (Ventana Medical Systems, Tucson, AZ, USA).

For Her2 testing, dual-color fluorescence in situ hybridization (FISH), using the PathVysion HER-2 DNA Probe Kit (PathVysion; Abbott Molecular, Des Plaines, IL, USA), and IHC were used in accordance with the guidelines of the American Society of Clinical Oncology (ASCO) [22]. After counting the signals of Her2 and those of centromeric enumeration probe 17 (CEP17) under a fluorescence microscope, a Her2/CEP17 ratio ≥ 2.0 was defined as amplification.

Genomic DNA extraction

Tumor and normal lymph node (reference) samples were obtained from 5 µm-thick tissue sections using a laser microdissection system (LMD6000; Leica Microsystems, Wetzlar, Germany) [20]. Briefly, each sample was dissected from an area ≥6 mm^2. In tumor samples, neoplastic cells comprised 90% of the total cell count. The cells were digested with a 200 mg/mL proteinase K solution (P2308, Sigma-Aldrich, St. Louis, MO, USA) for 70 ± 2 h at 37°C prior to a phenol/chloroform DNA extraction. DNA quality was assessed based on the A260/A280 ratio (cut-off >1.5), A260/A230 ratio (cut-off >1.0), and the presence or absence of double-stranded DNA.

Whole genome amplification (WGA)

Sample DNA was amplified using the GenomePlex Whole Genome Amplification Kit (WGA2 Kit; Sigma, St. Louis, MO, USA) according to the manufacturer's protocol [23].

Array CGH

For genomic DNA analysis, a 60-mer oligonucleotide CGH microarray (Agilent, Santa Clara, CA, USA) was used according to the manufacturer's instructions [24]. The genomic DNA enzymatic labelling and subsequent array CGH was performed as previously described [20]. The tumor-to-reference fluorescence intensity ratio (T/R) was calculated from the hybridized array images obtained. The UCSC Genome Browser was used with the latest resource content: hg19 assembly - Design ID 021429 (GRCh Build 37). CNAs were defined as a gain, loss, and amplification when the base 2 logarithm of the T/R ratio was >0.3219, <−0.3219, and >1.0, respectively. The microarray data were registered in the Gene Expression Omnibus (GEO) database (Accession number: GSE86988).

Validation of array CGH data by fluorescence in situ hybridization

Using the samples positive for *ERBB2* gene amplification, we compared FISH signal numbers and the T/R ratio of the array CGH. For dual-color FISH for *ERBB2* gene amplification, we used PathVysion Her2 DNA Probe Kit, (Abbot Molecular Inc.). We randomly selected 7 and 8 samples from the sample groups that showed strong (3+) and weak (1+ or 2+) Her2 immunoreactivities, respectively.

Clustering algorithm

To enhance the signal-to-noise ratio in the hybridization analysis, we averaged the T/R ratio of the probes within each gene prior to performing the cluster analyses. The noise-canceling effect of averaging depends on the gene size (probe number within the gene), whereas the clustering reproducibility becomes lower as the gene number becomes smaller. Notably, larger gene sizes correspond to smaller gene numbers. Thus, we repeated the clustering analysis to determine the optimal gene size and number.

To classify samples based solely on genome-wide similarities in gene copy-number gain/loss patterns, we performed an unsupervised hierarchical cluster analysis using a free software program (Cluster 3.0, version 1.52 and TreeView, version 1.1.6r2). The clustering results were assessed as previously described [20, 25, 26]. Briefly, we

repeated the clustering analysis using genes ranging from 370 genes containing ≥10 probes to 9,487 genes containing ≥2 probes. We selected 2 gene size conditions that showed the highest reproducibility in clustering dendrograms, and then selected the condition with the highest proportion of sample sets that derived from the same tumor and showed the neighboring in the clustering dendrogram.

The clustering condition was set to a complete linkage (maximum of distance metric on similarities) and the uncentered correlation distance (distance measures based on modified Pearson's correlation).

Statistical analysis

The CNA differences for each gene between clusters A and B were statistically assessed in an unequal sample-size t-test (Welch's t-test). A bilateral *p*-value of ≤0.05 was considered statistically significant. For multiple comparisons, the t test was subsequently adjusted using the Bonferroni correction [25] (Microsoft Office Excel 2013). To assess trend differences in either nuclear atypia or CNA accumulations between the 2 groups, a Fisher's exact test (2 × 2 contingency tables) was performed (BellCurve for Excel, Social Survey Research Information Co., Ltd., Tokyo, Japan).

Results

Immunohistochemistry

The intrinsic subtype was estimated immunohistochemically on the basis of the clinico-pathological surrogate definitions of subtypes based on the 2013 St Gallen Consensus [27]; immunohistochemistry-based definition of luminal A-like tumors is ER-positive, HER2-negative, Ki-67 index less than 14% and PR positivity more than 20% [28]. The 81 carcinoma samples were classified into 20 luminal A-like, 37 luminal B-like (HER2 negative), 6 luminal B-like (HER2 positive), 6 HER2 positive (non-luminal), and 12 triple negative (ductal).

Association of chromosomal CNAs with clinicopathological factors

IDC and DCIS commonly showed gains of 1q, 5p, 8q, 11q13, 16p, 17q, and 21q and losses of 4q, 8p, distal 11q, 13q, 14q, 16q, 17p, and 22q, but different in the frequencies of 6q−, 11q−, and 22q− (Fig. 1a-d, Table 1). The chromosomal CNAs different in frequency between NG1 and NG2/3 samples, were 4q−, 16q−, and 22q− (Fig. 1e,f, Table 2a). Those different between luminal-like (ER+) and non-luminal-like (ER−) subtypes were 4q−, 7p+, 8q+, 10p+, 16q−, and 21q+ (Fig.1g, h, Table 2b). Only 16q− was significantly different in frequency between the samples of N0 and N1 tumors (Fig.1i, j, Table 3).

Amplifications of 17q12 were found in 9/12 of the tumor samples that showed strong (3+) immunohistochemical expression of HER2 irrespective of DCIS or IDC.

Validation of array CGH data by fluorescence in situ hybridization

In the 7 samples with strong Her2 immunoreactivity, the T/R ratios of *ERBB2* gene were 1.81 to 3.12 (average 2.53) and the FISH signal ratio was 2.4 to 5.8 (average 4.2). In the 8 samples with weak Her2 immunoreactivity, the T/R ratios of *ERBB2* gene were -0.93 to 0.03 (average -0.24), and the FISH signal ratio was 0.9 to 1.3 (average 1.05), No gene amplification is detected in the samples with weak HER2 immunoreactivity, and vice versa.

Clustering of gene copy-number profiles

The individual probe T/R ratios within a specified gene were averaged. The average T/R ratios of 30,471 gene regions were calculated from 55,023 probes. Genes selected based on size were subjected to an unsupervised hierarchical cluster analysis. After repeated clustering using varying minimum gene sizes, we found that the gene sizes of ≥3 probes and ≥4 probes gave highest reproducibility of clustering results (Supplementary file 2). In the condition of a gene size ≥4 probes, all samples from the

same case were in the neighboring position in the clustering dendrogram, confirming the reproducibility of the CNA profile. Thus, we adopted this as the optimal condition for the unsupervised hierarchical clustering of all (cancer and papilloma) samples. This condition yielded 2 main clusters: A and B (Fig. 2). The cancer samples of cluster A showed a greater IDC/DCIS ratio, a higher ER−/ER+ (non-luminal/luminal) ratio, and higher nuclear grade than those in cluster B (Table 4). There were also tendencies for higher Ki-67 index, higher triple negative tumors in Cluster A. There was no difference in the frequency of lymph node metastasis between the clusters A and B.

Lineage-specific chromosomal CNA profile

The penetrance plots of clusters A and B revealed distinct chromosomal CNA profiles (Fig. 3). In addition to changes common to these clusters (a gain of 1q and stage-specific losses of 6q and 16q), cluster A was characterized by gains of 5p, 16p, and 21q and losses of 4p and 8p, whereas cluster B scarcely showed these changes (Table 2c).

Genes exhibiting significantly different CNAs between 2 major clusters

We extracted 728 genes that show significantly different mean copy numbers

between clusters A and B in a t-test with Bonferroni correction. Out of the top 45 genes shown in Supplementary file 3a, 42 were protein-coding, including 10 genes that showed concordance between the gene CNA and chromosome CNA in cluster A. In 7 of these 10 genes, gene function was relevant to the direction of CNA, i.e., gain and loss of protooncogene and tumor suppressor genes, respectively. These 7 genes included the proto-oncogene, *LMO3*, which reportedly plays a role in T cell leukemia and brain tumors. The other 3 genes, which included *PIK3R5*, showed a loss of gene and chromosome copy number. Notably, this loss is functionally opposite because the genes are protooncogenic. These genes may be passenger genes, and thus only useful as a lineage marker for the differentiation of clusters A and B.

Of the 93 genes that were reported to be important in a next generation sequencing analysis of breast cancers [29], 44 genes showed statistically significant differences in the mean CNA between clusters A and B (Supplementary file 3b), and 9 genes (*GATA3*, *TP53*, *TET2*, *NCOR1*, *NOTCH2*, *PIC3CA*, *CREBBP*, *MYC,* and *ERBB2* in decreasing order of significance) showed concordance between both the gene and chromosome CNAs and the function and gain/loss of genes. Of these genes, only *GATA3* and *TP53* remained significant after Bonferroni correction. *GATA3* gain and TP53 loss were common in cluster A.

Discussion

For validation of array CGH data, we utilized the samples with and without the overexpression (3+) of Her 2 protein and demonstrated almost complete concordance between array CGH and FISH results. In the absence of gene amplification, quantitative polymerase chain reaction (qPCR) as well as FISH is difficult to use for the validation of CNAs [20]. We attempted to cancel the noise inherent to FFPE tissues by averaging the T/R ratio of larger-sized genes. Using the internal standard mentioned in Methods, we optimized the gene size for the assessment of gene-level copy number alterations. Applying this method to the present breast cancer samples, we found 2,828 genes of ≥4 probes as the optimal condition for unsupervised clustering.

Next generation sequencing (NGS) showed that gene copy-number changes are more common than significant DNA sequence changes in cancers [29]. Additionally, recent NGS-based approaches to tumor heterogeneity demonstrated that chromosomal CNAs are the principal factor for tumor progression, whereas sequence changes of driver genes as well as chromosomal CNAs are important for the earlier phases of carcinogenesis [30]. In the present chromosome-level CNA analysis, our penetrance plot of IDC (invasive part) showed a gain/loss pattern very similar to recent

NGS data of 560 breast cancers [29], except for 9q loss, which was observed less frequently in our study.

The chromosomal CNA profiles were similar between DCIS and IDC, as previously reported [15, 31], except for the scarcity of losses in 6q, 11q, and 22q in DCIS (Table 1). Gains of 5p, 7, 11q, 16p, and 20q and a loss of 8p were previously reported to be more frequent in IDC than DCIS [32, 33]. In the present study, however, none of them showed statistically significant difference between DCIS and IDC. Other clinicopathological factors were correlated with characteristic chromosomal changes (Tables 2 and 3): 22q− correlated with high NG; 7p+, 8q+, and 10p+ with non-luminal-like (ER−) subtypes; the absence of 16q− with lymph node metastasis. The 8q+, 17q+, and 8p− reported to be common in high-grade breast cancers [31] were not significant in our results. CNAs, including a gain of 1q and loss of 16q, were detected in nearly 50% of IDCs and 25% of DCIS. All samples with both 1p+ and 16q− were luminal-like (ER+) type, as previously reported [34]. Problem is whether these changes are useful for the specification of progression-prone DCIS. Table 1 demonstrates that neither NG nor intrinsic subtype showed significant correlation with an IDC/DCIS ratio, which may be related to the risk of progression from DCIS to IDC. Thus, we analyzed the following gene-level CNAs to assess the outcome of individual tumors.

Under the above-mentioned optimal condition, the unsupervised hierarchical clustering gave 2 main clusters for all (cancer and papilloma) samples: A and B. The cancer samples of cluster A showed a greater frequency of chromosomal gain and loss, a greater IDC/DCIS ratio, a higher non-luminal-like/luminal-like (ER−/ER+) ratio and higher nuclear grade than cluster B (Table 4). Thus, the tumors of cluster A may be phenotypically less differentiated, and have accumulated a greater number of genomic changes (Figs. 2 and 3), reflecting higher level of genomic instability. The DCIS lesions in cluster A that correlates with high NG may be more progression-prone to IDC than those in cluster B. This is consistent with the previous notion that nuclear grade largely paralleled the total number of gene CNAs and the risk of progression from DCIS to IDC [31, 35].

However, there was no difference in the frequency of lymph node metastasis between clusters A and B (Table 4) as well as between low and high NG and between ER+ and ER− tumors (Table 3). Additionally, Table 3 demonstrated that the accumulation of chromosomal CNAs was scarcely different between N0 and N1 samples. The only chromosomal CNA significantly different between N0 and N1 was the absence of 16q loss in metastasis samples. This suggests that the metastasis risk might be determined in earlier phase of tumor development because the 16q loss is

common early event in breast carcinomas (Fig. 1). This point, whether 16q copy number is useful in the prediction of metastasis risk should be further studied using larger breast cancer cohorts.

It seems that the metastasis risk and the progression risk form DCIS to IDC reflect different genomic and epigenomic features. In stomach cancers, we similarly found that the copy-number profiling approach could stratified tumor samples into rapidly and slowly progressive 2 groups, but no difference in metastasis risk was shown between these groups, whereas, different from the present study, metastatic gastric samples accumulated later chromosomal CNAs more frequently than the samples without metastasis (unpublished data).

Of the 30,471 gene regions, 728 showed significant copy-number differences between clusters A and B. Among them, 9 genes (*TP53*, *GATA3*, *CDKN2A*, *ATR*, *ATRX*, *PHF6*, *SMARCA4*, *APC,* and *ASXL1*) were included in the breast cancer-related 93 genes [29]. Of these 9 genes, only *TP53* and *GATA3* [36] remained significant after the Bonferroni correction and showed concordance between gene and chromosome changes and between gain/loss and gene function. The other genes showed discordance between gene function and copy-number changes, and thus may be passenger genes (Supplrmentary file 3b). Recent meta-analysis of 10 articles reported

that GATA3 expression was associated with better prognosis [37]. This seems contradictory to our result that invasion-prone tumors often showed *GATA3* copy-number gain. In our unpublished data, copy-number changes of *GATA3* were not in parallel with its IHC results, which may reflect epigenetic regulations.

Still currently, screening-positive patients are often treated with additional radiotherapy after segmental surgical excision. Such postoperative therapies can be individualized based on the progression risk of each DCIS; the targets can be pinpointed to invasion-prone DCIS, as detected in the present study. The genes we specified could be useful for a construction of a simple system for pinpointing the invasion-prone DCIS.

References

1 Kurebayashi J, Miyoshi Y, Ishikawa T, Saji S, Sugie T, Suzuki T, Takahashi S, Nozaki M, Yamashita H, Tokuda Y, Nakamura S: Clinicopathological characteristics of breast cancer and trends in the management of breast cancer patients in Japan: Based on the Breast Cancer Registry of the Japanese Breast Cancer Society between 2004 and 2011. Breast Cancer 2015;22:235-244.

2 Cowell CF, Weigelt B, Sakr RA, Ng CK, Hicks J, King TA, Reis-Filho JS: Progression from ductal carcinoma in situ to invasive breast cancer: revisited. Mol Oncol 2013;7:859-869.

3 Welch HG, Black WC: Using autopsy series to estimate the disease

"reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find? Ann Intern Med 1997;127:1023-1028.

4 Screening IUPoBC: The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380:1778-1786.

5 Welch HG, Black WC: Overdiagnosis in cancer. J Natl Cancer Inst 2010;102:605-613.

6 Sagara Y, Mallory MA, Wong S, Aydogan F, DeSantis S, Barry WT, Golshan M: Survival Benefit of Breast Surgery for Low-Grade Ductal Carcinoma In Situ: A Population-Based Cohort Study. JAMA Surg 2015

7 Page DL, Dupont WD, Rogers LW, Jensen RA, Schuyler PA: Continued local recurrence of carcinoma 15-25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy. Cancer 1995;76:1197-1200.

8 Stuart KE, Houssami N, Taylor R, Hayen A, Boyages J: Long-term outcomes of ductal carcinoma in situ of the breast: a systematic review, meta-analysis and meta-regression analysis. BMC Cancer 2015;15:890.

9 Leonard GD, Swain SM: Ductal carcinoma in situ, complexities and challenges. J Natl Cancer Inst 2004;96:906-920.

10 Amari M, Moriya T, Ishida T, Harada Y, Ohnuki K, Takeda M, Sasano H, Horii A, Ohuchi N: Loss of heterozygosity analyses of asynchronous lesions of ductal carcinoma in situ and invasive ductal carcinoma of the human breast. Jpn J Clin Oncol 2003;33:556-562.

11 Yen MF, Tabár L, Vitak B, Smith RA, Chen HH, Duffy SW: Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. Eur J Cancer 2003;39:1746-1754.

12 Robanus-Maandag EC, Bosch CA, Kristel PM, Hart AA, Faneyte IF, Nederlof PM, Peterse JL, van de Vijver MJ: Association of C-MYC amplification with

progression from the in situ to the invasive stage in C-MYC-amplified breast carcinomas. J Pathol 2003;201:75-82.

13 Jang M, Kim E, Choi Y, Lee H, Kim Y, Kim J, Kang E, Kim SW, Kim I, Park S: FGFR1 is amplified during the progression of in situ to invasive breast carcinoma. Breast Cancer Res 2012;14:R115.

14 Burkhardt L, Grob TJ, Hermann I, Burandt E, Choschzick M, Jänicke F, Müller V, Bokemeyer C, Simon R, Sauter G, Wilczak W, Lebeau A: Gene amplification in ductal carcinoma in situ of the breast. Breast Cancer Res Treat 2010;123:757-765.

15 Hernandez L, Wilkerson PM, Lambros MB, Campion-Flora A, Rodrigues DN, Gauthier A, Cabral C, Pawar V, Mackay A, A'hern R, Marchiò C, Palacios J, Natrajan R, Weigelt B, Reis-Filho JS: Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive breast cancers reveals intra-tumour genetic heterogeneity and clonal selection. J Pathol 2012;227:42-52.

16 Vincent-Salomon A, Lucchesi C, Gruel N, Raynal V, Pierron G, Goudefroye R, Reyal F, Radvanyi F, Salmon R, Thiery JP, Sastre-Garau X, Sigal-Zafrani B, Fourquet A, Delattre O, Curie bcsgotI: Integrated genomic and transcriptomic analysis of ductal carcinoma in situ of the breast. Clin Cancer Res 2008;14:1956-1965.

17 Liao S, Desouki MM, Gaile DP, Shepherd L, Nowak NJ, Conroy J, Barry WT, Geradts J: Differential copy number aberrations in novel candidate genes associated with progression from in situ to invasive ductal carcinoma of the breast. Genes Chromosomes Cancer 2012;51:1067-1078.

18 Friberg S, Mattson S: On the growth rates of human malignant tumors: implications for medical decision making. J Surg Oncol 1997;65:284-297.

19 Sonoda A, Mukaisho K, Nakayama T, Diem VT, Hattori T, Andoh A, Fujiyama Y, Sugihara H: Genetic lineages of undifferentiated-type gastric carcinomas analysed by unsupervised clustering of genomic DNA microarray data. BMC

Med Genomics 2013;6:25.

20 Vo DT, Nakayama T, Yamamoto H, Mukaisho K, Hattori T, Sugihara H: Progression risk assessments of individual non-invasive gastric neoplasms by genomic copy-number profile and mucin phenotype. BMC Med Genomics 2015;8:6.

21 Furuya C, Kawano H, Yamanouchi T, Oga A, Ueda J, Takahashi M: Combined evaluation of CK5/6, ER, p63, and MUC3 for distinguishing breast intraductal papilloma from ductal carcinoma in situ. Pathol Int 2012;62:381-390.

22 Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, Allred DC, Bartlett JM, Bilous M, Fitzgibbons P, Hanna W, Jenkins RB, Mangu PB, Paik S, Perez EA, Press MF, Spears PA, Vance GH, Viale G, Hayes DF, Oncology ASoC, Pathologists CoA: Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Arch Pathol Lab Med 2014;138:241-256.

23 Little SE, Vuononvirta R, Reis-Filho JS, Natrajan R, Iravani M, Fenwick K, Mackay A, Ashworth A, Pritchard-Jones K, Jones C: Array CGH using whole genome amplification of fresh-frozen and formalin-fixed, paraffin-embedded tumor DNA. Genomics 2006;87:298-306.

24 Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R, Tsang P, Curry B, Baird K, Meltzer PS, Yakhini Z, Bruhn L, Laderman S: Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. Proc Natl Acad Sci U S A 2004;101:17765-17770.

25 Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863-14868.

26 Quackenbush J: Computational analysis of microarray data. Nat Rev Genet 2001;2:418-427.

27 Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, Senn HJ, members P: Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 2013;24:2206-2223.

28 Prat A, Cheang MC, Martín M, Parker JS, Carrasco E, Caballero R, Tyldesley S, Gelmon K, Bernard PS, Nielsen TO, Perou CM: Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. J Clin Oncol 2013;31:203-209.

29 Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC, Van Loo P, Ju YS, Smid M, Brinkman AB, Morganella S, Aure MR, Lingjærde OC, Langerød A, Ringnér M, Ahn SM, Boyault S, Brock JE, Broeks A, Butler A, Desmedt C, Dirix L, Dronov S, Fatima A, Foekens JA, Gerstung M, Hooijer GK, Jang SJ, Jones DR, Kim HY, King TA, Krishnamurthy S, Lee HJ, Lee JY, Li Y, McLaren S, Menzies A, Mustonen V, O'Meara S, Pauporté I, Pivot X, Purdie CA, Raine K, Ramakrishnan K, Rodríguez-González FG, Romieu G, Sieuwerts AM, Simpson PT, Shepherd R, Stebbings L, Stefansson OA, Teague J, Tommasi S, Treilleux I, Van den Eynden GG, Vermeulen P, Vincent-Salomon A, Yates L, Caldas C, van't Veer L, Tutt A, Knappskog S, Tan BK, Jonkers J, Borg Å, Ueno NT, Sotiriou C, Viari A, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Birney E, Stunnenberg HG, van de Vijver MJ, Martens JW, Børresen-Dale AL, Richardson AL, Kong G, Thomas G, Stratton MR: Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature 2016;534:47-54.

30 Uchi R, Takahashi Y, Niida A, Shimamura T, Hirata H, Sugimachi K, Sawada G, Iwaya T, Kurashige J, Shinden Y, Iguchi T, Eguchi H, Chiba K, Shiraishi Y, Nagae G, Yoshida K, Nagata Y, Haeno H, Yamamoto H, Ishii H, Doki Y, Iinuma H, Sasaki S, Nagayama S, Yamada K, Yachida S, Kato M, Shibata T, Oki E, Saeki H, Shirabe K, Oda Y, Maehara Y, Komune S, Mori M, Suzuki Y, Yamamoto K, Aburatani H, Ogawa S, Miyano S, Mimori K: Integrated Multiregional Analysis Proposing a New Model of Colorectal Cancer Evolution. PLoS Genet 2016;12:e1005778.

31 Lopez-Garcia MA, Geyer FC, Lacroix-Triki M, Marchió C, Reis-Filho JS: Breast cancer precursors revisited: molecular features and progression pathways. Histopathology 2010;57:171-192.

32 Yao J, Weremowicz S, Feng B, Gentleman RC, Marks JR, Gelman R, Brennan C, Polyak K: Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. Cancer Res 2006;66:4065-4078.

33 Anbazhagan R, Fujii H, Gabrielson E: Allelic loss of chromosomal arm 8p in breast cancer progression. Am J Pathol 1998;152:815-819.

34 Buerger H, Otterbach F, Simon R, Schäfer KL, Poremba C, Diallo R, Brinkschmidt C, Dockhorn-Dworniczak B, Boecker W: Different genetic pathways in the evolution of invasive breast cancer are associated with distinct morphological subtypes. J Pathol 1999;189:521-526.

35 Fidalgo F, Rodrigues TC, Pinilla M, Silva AG, Maciel MoS, Rosenberg C, de Andrade VP, Carraro DM, Krepischi AC: Lymphovascular invasion and histologic grade are associated with specific genomic profiles in invasive carcinomas of the breast. Tumour Biol 2015;36:1835-1848.

36 Takaku M, Grimm SA, Wade PA: GATA3 in Breast Cancer: Tumor Suppressor or Oncogene? Gene Expr 2015;16:163-168.

37 Guo Y, Yu P, Liu Z, Maimaiti Y, Chen C, Zhang Y, Yin X, Wang S, Liu C, Huang T: Prognostic and clinicopathological value of GATA binding protein 3 in breast cancer: A systematic review and meta-analysis. PLoS One 2017;12(4):e0174843.

Figures

Figure 1 Penetrance plots. a. ductal carcinoma in situ (DCIS, 24 samples); b. ductal part of invasive ductal carcinoma (IDC) (Cd, 22 samples); c. invasive part of IDC (Ci,

28 samples); d. nodal metastatic tumors of IDC (CL, 7 samples); e. low nuclear grade (NG1, 22 samples); f. high nuclear grade (NG2,3, 59 sample); g: luminal subtype (ER+, 63 samples); h. (non-luminal subtype (ER−, 18 samples); i. nodal metastasis absent (N−, 37 samples); j. nodal metastasis present (N+, 20 samples). Yellow squares mark statistically significant changes.

Figure 2 Unsupervised, hierarchical clustering using 81 samples and 2828 larger genes that contain ≥4 probes. The abbreviations of samples, D, Ci, Cd and CL indicate ductal carcinoma in situ (DCIS), ductal part of invasive ductal carcinoma (IDC), invasive part of IDC and lymph node metastasis, respectively. Copy-number gains and losses are indicated by green and red squares, respectively in the heat map. Beneath the dendrogram are 3 color bars. The top, the middle and the bottom indicate nuclear grade (dark green NG3; light green, NG2; yellow, NG1), intrinsic subtype (red, Her2; pink, luminal A; light blue, luminal B; gray, triple negative; green) and tumor stage (white, IDC; gray, DCIS), respectively. Blue squares indicate sample groups of the same case)

Figure 3 Cluster-specific penetrance plots of ductal carcinomas. A. cluster A (50

samples); B. cluster B (31 samples). Yellow squares mark statistically significant changes.

Supplementary files

Supplementary file 1 Clinicopathological data of samples (XLSX 16 kb)

Supplementary file 2 Clustering dendrograms with varying gene numbers and sizes. These ranged from 9487 genes containing ≥2 probes to 370 genes containing ≥10 probes. Red horizontal lines indicate the border of 2 major clusters. (PDF 159 kb)

Supplementary file 3 a. The top 45 genes with highly different mean copy numbers between clusters A and B. The gene are arranged in a decreasing order of significance. **b.** A list of 44 breast cancer-related genes [32], which showed significantly different mean copy numbers between clusters A and B. The 9 genes above the horizontal line remain significant after Bonferroni correction. In **a** and **b**, the order of samples is the same as that in Fig. 2. The copy number of each gene is shown as the tumor/reference fluorescence ratio. The orange and moss green indicate copy-number gain and loss, respectively. In the columns of chromosome arms, pink and deep green indicate gain and loss respectively. Light pink and light green indicate slight changes. The column of

concordance shows concordance or discordance of copy-number changes between gene and chromosome levels. The column of relevance shows whether the copy-number changes are concordant to gene function for the expression of malignant phenotype. (PDF 1062 kb)

Table 1. Comparisons of clinicopathological features and chromosomal changes between DCIS and IDC samples.

	DCIS	IDC	P
Low NG (1)	10	12	0.0986
High NG (2,3)	14	45	
ER+ (Luminal A/B (Her2+/−))	20	43	0.5637
ER− (Her2+/TN)	4	14	
6q−	0	10	**0.0290**
absent	24	47	
8p−	2	13	0.2094
absent	22	44	
11q−	1	16	**0.0167**
absent	23	41	
16p+	2	10	0.4943
absent	22	47	
20q+	1	5	0.6641
absent	23	52	
22q−	1	15	**0.0300**
absent	23	40	

Table 2: Differences in sample-based frequencies of chromosomal copy-number changes between high nad low nuclear grade (a), between ER+ and ER− (b), and between clusters A and B (c),

	a			**b**			**c**		
	Number of samples		P-value	Number of samples		P-value	Number of samples		P-value
	High NG	Low NG		ER+	ER−		Cluster A	Cluster B	
	59	22		63	18		50	31	
1q+	21	9	0.7966	25	3	0.0938	17	11	1.0000
4p−	10	1	0.2731	5	6	0.0110	11	0	**0.0055**
4q−	13	0	**0.0157**	6	7	**0.0067**	10	3	0.3511
5p+	8	1	0.432	6	3	0.4083	9	0	**0.0112**
7p+	6	0	0.1824	2	4	**0.0201**	6	0	0.0774
8p−	12	3	0.7486	9	6	0.3399	15	0	**0.0003**
8q+	13	1	0.0975	5	9	**0.0201**	11	3	0.1466
10p+	6	0	0.1824	1	5	**0.0017**	6	0	0.0783
13q−	7	3	1.0000	10	0	0.1067	6	4	1.0000
14q−	8	0	0.0999	6	2	1.0000	7	1	0.1453
16p+	11	1	0.1648	7	5	0.1262	11	1	**0.0267**
16q−	14	12	**0.0148**	25	1	**0.0083**	13	13	0.1501
17p−	13	1	0.0975	11	3	1.0000	12	2	0.0675
21q+	7	0	0.1809	2	5	**0.0052**	7	0	**0.0401**
22q−	16	0	**0.0042**	14	2	0.5027	10	6	1.0000

Table 3. Comparisons of clinicopathological features and chromosomal changes between N0 and N1 samples of 57 invasive ductal carcinomas

	N0	N1	P
Low NG (1)	9	3	0.5098
High NG (2,3)	28	17	
ER+ (Luminal A/B (Her2+/−))	31	12	0.0594
ER− (Her2+/TN)	6	8	
4p−	7	4	1.0000
absent	30	16	
5p+	3	3	0.6542
absent	34	17	
6q−	4	6	0.1410
absent	33	14	
14q−	5	3	1.0000
absent	32	17	
16q−	17	2	**0.0077**
absent	20	18	
17p−	8	6	0.7651
absent	29	18	
22q−	10	6	1.0000
absent	27	14	

Table 4. Comparisons of breast carcinomas between Lineages A and B

	A (n=50)	B (n=31)	P
Mean gain regions	8.7	1.9	**0.0023**
Mean loss regions	6.2	2.2	**0.0143**
DCIS*	10	14	**0.0239**
IDC*	40	17	
Low NG (1)*	7	15	**0.0016**
High NG (2,3)*	43	16	
ER+ (Luminal A/B (Her2+/−))*	34	29	**0.0117**
ER− (Her2+/TN)*	16	2	
Mean Ki-67 index of cancer cells	12.6	9.5	0.0672
N+ in IDC cases (30)	6	3	1.0000
N−	15	6	
N+ in IDC samples (57)*	14	6	1.0000
N−*	26	11	

* Described in sample numbers

Fig.1

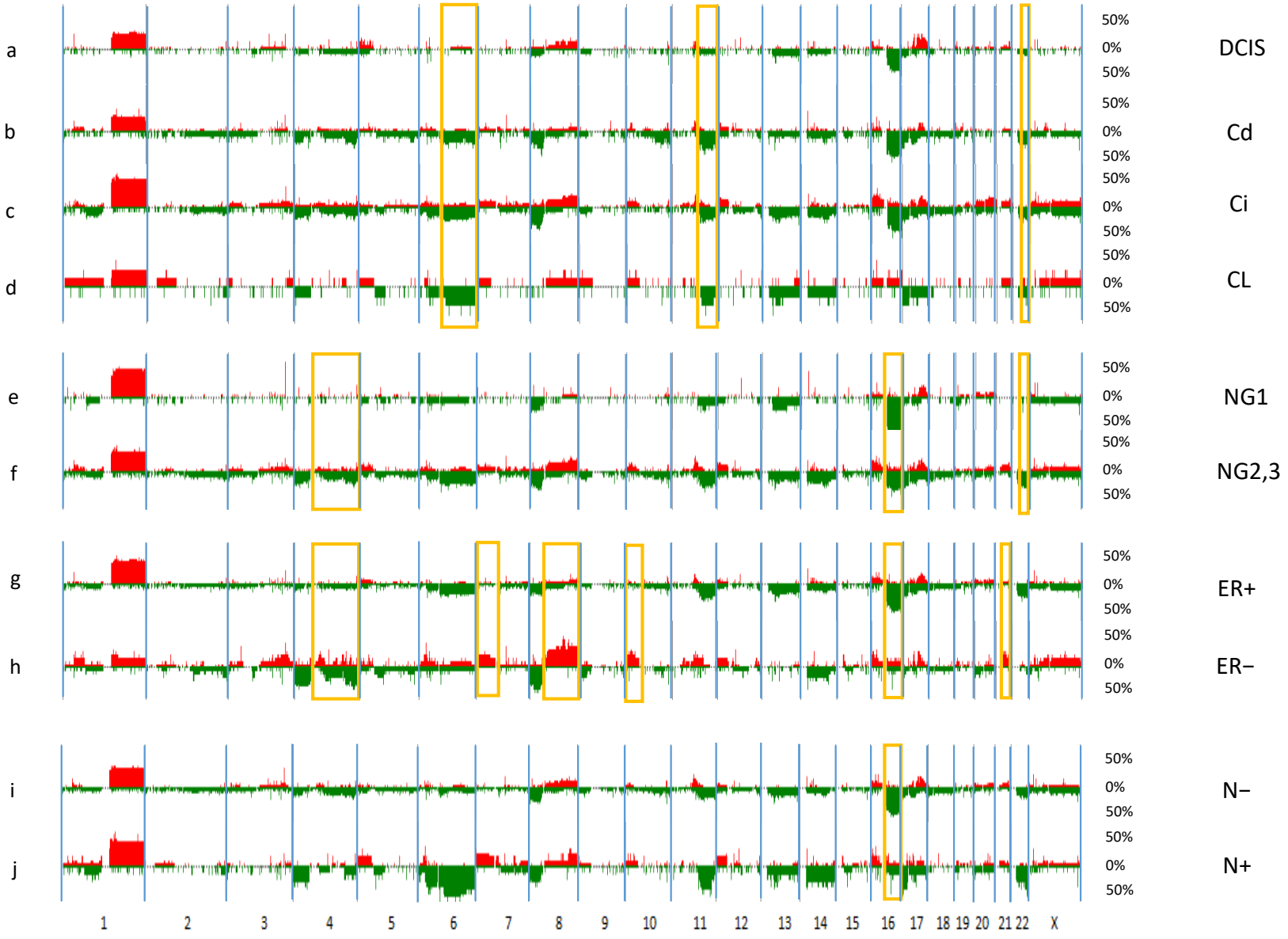

a
b
c
d
e
f
g
h
i
j
50%
0%
50%
DCIS
Cd
Ci
CL
NG1
NG2,3
ER+
ER−
N−
N+
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X

Fig.2

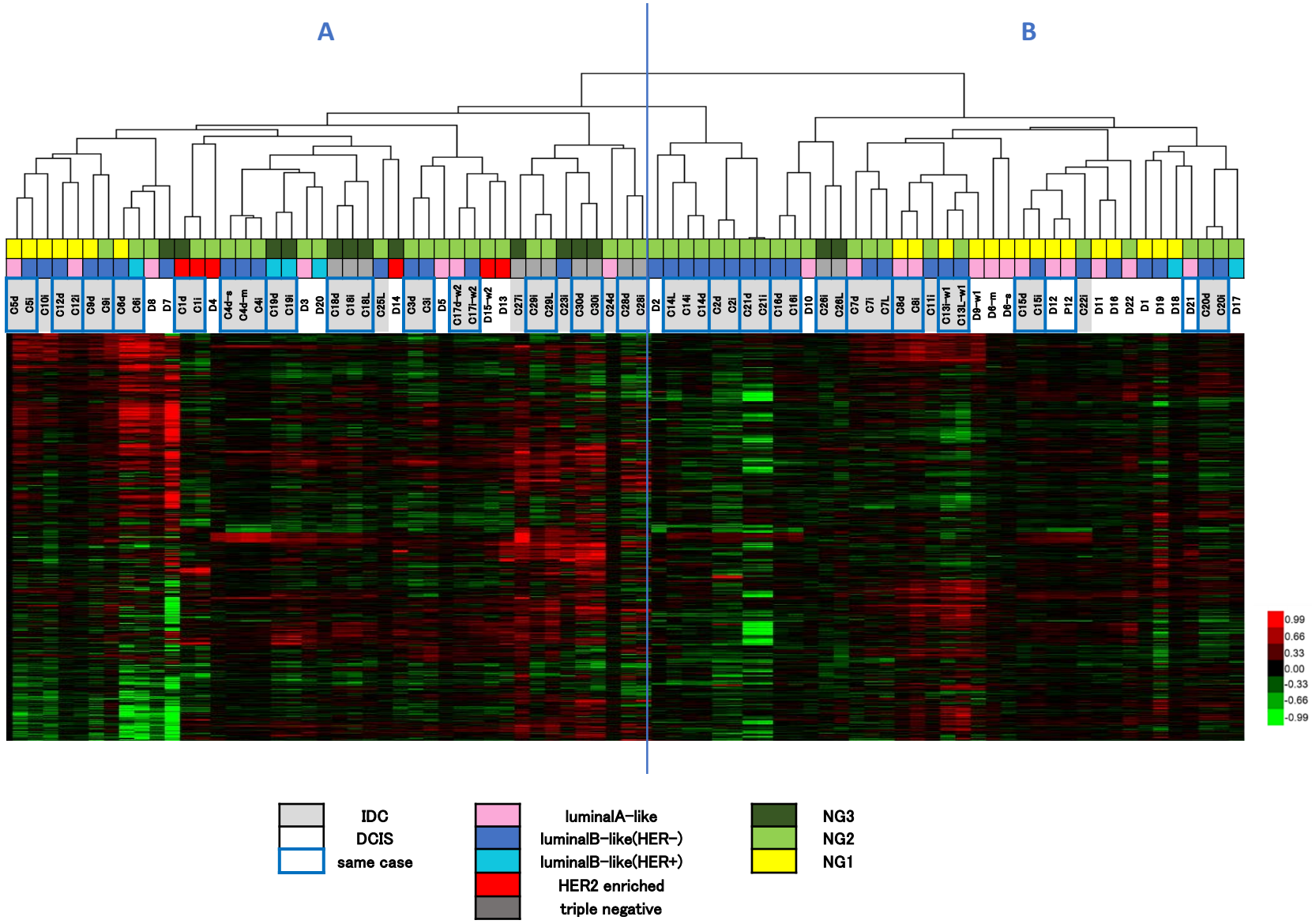
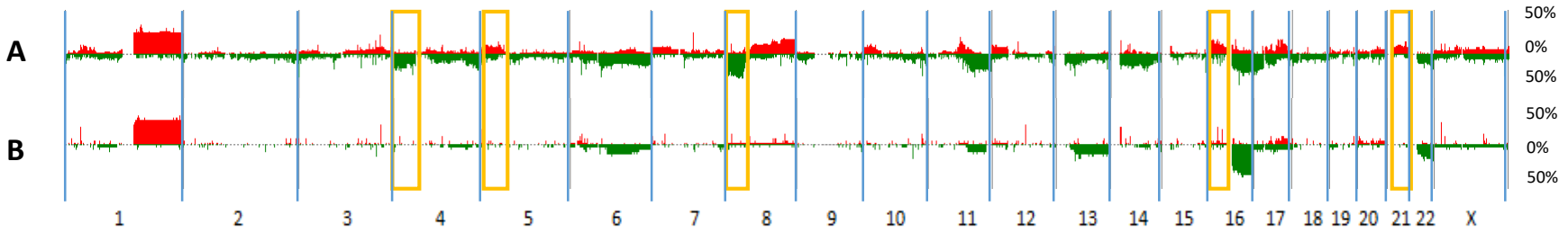
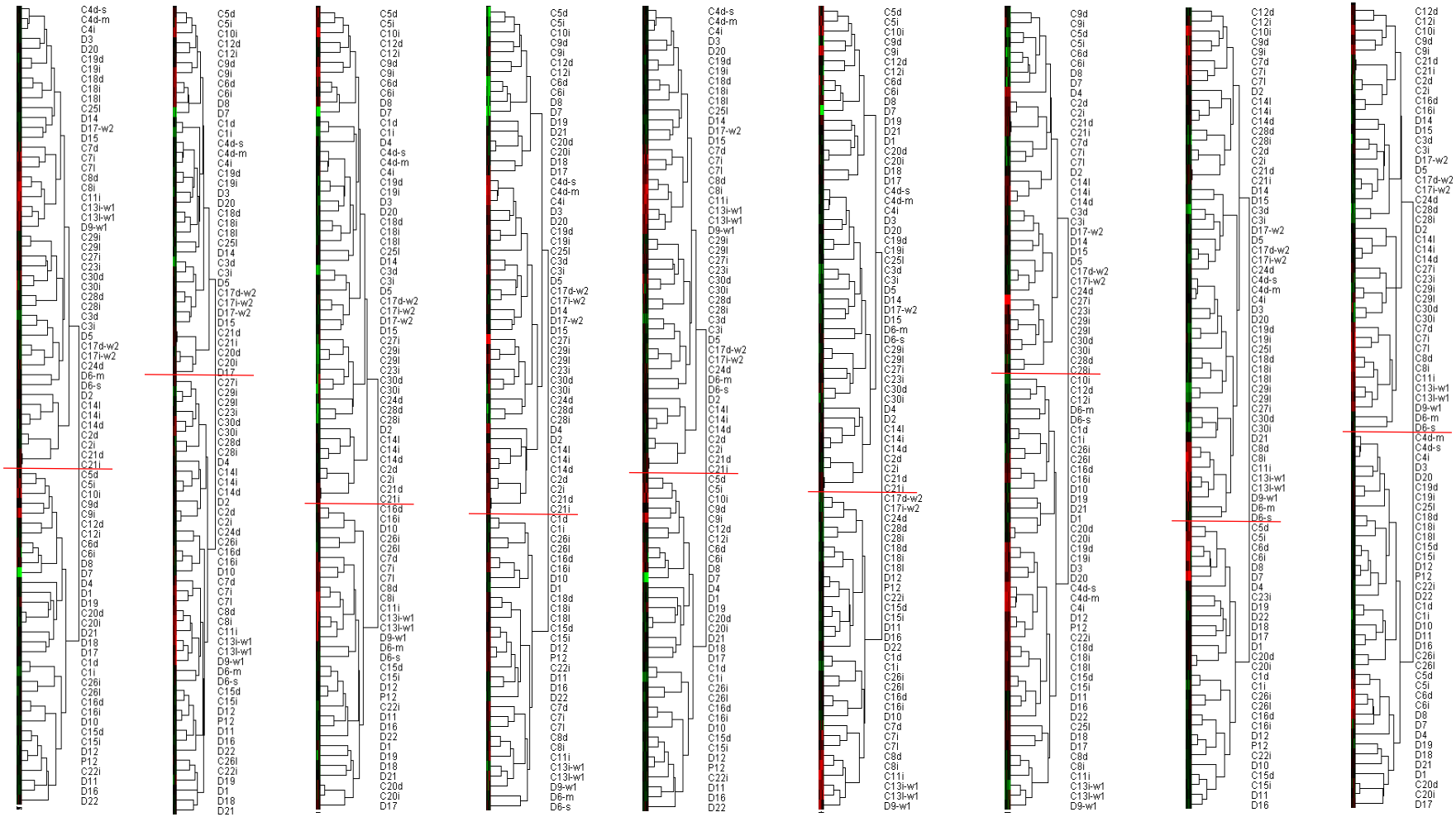

A
B
C5d C5i C10i C12d C12i C9d C9i C6d C6i D8 D7 C1d C1i D4 C4d-s C4d-m C4i C19d C19i D3 D20 C18d C18i C18L C25L D14 C3d C3i D5 C17d-w2 C17i-w2 D15-w2 D13 C27i C29i C29L C23i C30d C30i C24d C28d C28i D2 C14L C14i C14d C2d C2i C21d C21i C16d C16i D10 C26i C28L C7d C7i C7L C8d C8i C11i C13i-w1 C13L-w1 D9-w1 D6-m D6-s C15d C15i D12 P12 C22i D11 D16 D22 D1 D19 D18 D21 C20d C20i D17
0.99
0.66
0.33
0.00
-0.33
-0.66
-0.99
IDC
DCIS
same case
luminalA-like
luminalB-like(HER-)
luminalB-like(HER+)
HER2 enriched
triple negative
NG3
NG2
NG1

Fig.3



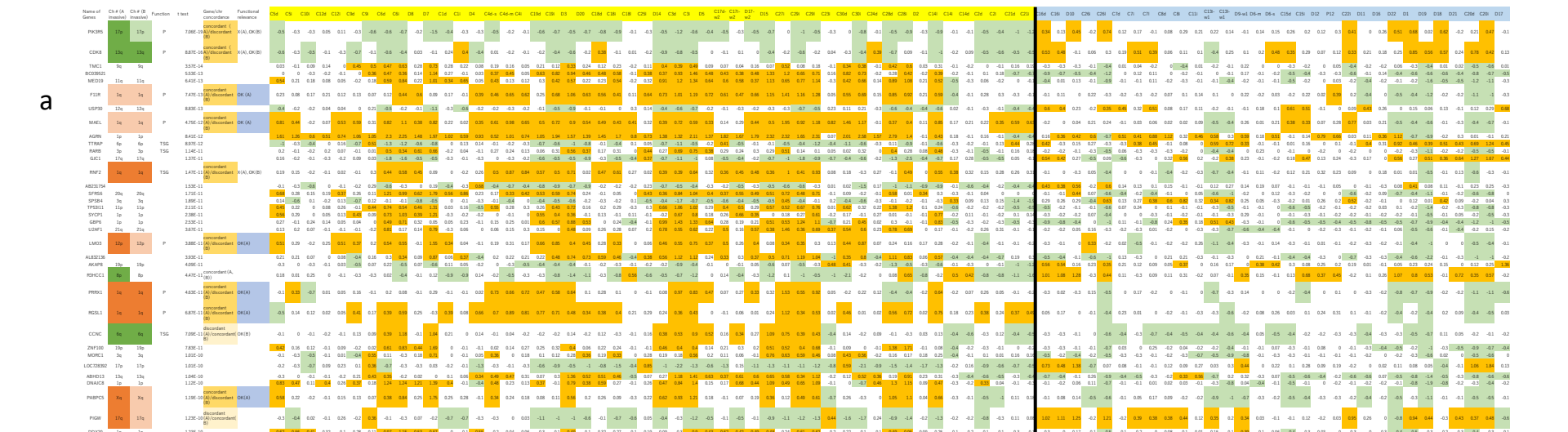
Patient number	No.	Age	TNM	Stage	Component	Nuclear atypia	ER(%)	PgR(%)	HER2	MIB-I(%)	Subtype
#1	C1i	49	1/0/0	I	Invasive	2	0	0	3	29	HER2 positive(non-luminal)
#1	C1d				Ductal	3	0	0	3	15	HER2 positive(non-luminal)
#2	C2i	53	2/0/0	IIA	Invasive	2	30	0	0	37	Luminal B-like(HER2 negative)
#2	C2d				Ductal	2	12	0	0	30	Luminal B-like(HER2 negative)
#3	C3i	62	1/0/0	I	Invasive	2	100	0	1	6	Luminal B-like(HER2 negative)
#3	C3d				Ductal	2	100	0	0	6	Luminal B-like(HER2 negative)
#4	C4i	62	2/0/0	IIA	Invasive	2	80	0	1	2	Luminal B-like(HER2 negative)
#4	C4d-m				Ductal	2	90	0	1	5	Luminal B-like(HER2 negative)
#4	C4d-s				Ductal	2	90	0	1	5	Luminal B-like(HER2 negative)
#5	C5i	62	1/1/0	IIA	Invasive	1	100	1	1	7	Luminal B-like(HER2 negative)
#5	C5d				Ductal	1	100	70	0	3	Luminal A-like
#6	C6i	83	1/0/0	I	Invasive	2	100	0	3	13	Luminal B-like(HER2 positive)
#6	C6d				Ductal	1	100	1	2	6	Luminal B-like(HER2 negative)
#7	C7i	37	1/1/0	IIA	Invasive	2	99	90	0	17	Luminal B-like(HER2 negative)
#7	C7d				Ductal	2	100	60	1	5	Luminal A-like
#7	C7L				LNs meta	2	100	40	0	27	Luminal B-like(HER2 negative)
#8	C8i	43	2/0/0	IIA	Invasive	1	90	90	1	13	Luminal A-like
#8	C8d				Ductal	1	99	100	1	7	Luminal A-like
#9	C9i	35	1/0/0	I	Invasive	2	98	0	0	25	Luminal B-like(HER2 negative)
#9	C9d				Ductal	1	100	0	0	8	Luminal B-like(HER2 negative)
#10	C10i	46	1/0/0	I	Invasive	1	98	90	1	33	Luminal B-like(HER2 negative)
#11	C11i	73	1/0/0	I	Invasive	2	100	50	0	15	Luminal B-like(HER2 negative)
#12	C12i	50	1/0/0	I	Invasive	1	98	40	1	9	Luminal A-like
#12	C12d				Ductal	1	98	10	0	4	Luminal B-like(HER2 negative)
#13	C13i-w1	57	1/1/0	IIA	Invasive	1	100	70	0	16	Luminal B-like(HER2 negative)
#13	C13L-w1				LNs meta	2	100	90	0	15	Luminal B-like(HER2 negative)
#14	C14i	47	2/1/0	IIB	Invasive	2	70	0	0	10	Luminal B-like(HER2 negative)
#14	C14d				Ductal	2	80	0	0	5	Luminal B-like(HER2 negative)
#14	C14L				LNs meta	2	90	10	0	10	Luminal B-like(HER2 negative)
#15	C15i	45	1/0/0	I	Invasive	1	70	10	1	10	Luminal B-like(HER2 negative)
#15	C15d				Ductal	1	80	20	1	10	Luminal A-like
#16	C16i	52	1/0/0	I	Invasive	2	50	90	1	18	Luminal B-like(HER2 negative)
#16	C16d				Ductal	2	50	80	1	20	Luminal B-like(HER2 negative)
#17	C17i-w2	84	1/0/0	I	Invasive	2	100	0	0	5	Luminal B-like(HER2 negative)
#17	C17d-w2				Ductal	2	100	30	0	3	Luminal A-like
#18	C18i	81	2/1/1	IV	Invasive	3	0	0	1	20	Triple negative
#18	C18d				Ductal	3	0	0	1	20	Triple negative
#18	C18L				LNs meta	3	0	0	1	18	Triple negative
#19	C19i	51	2/1/0	IIB	Invasive	3	70	5	3	15	Luminal B-like(HER2 positive)
#19	C19d				Ductal	3	70	5	3	14	Luminal B-like(HER2 positive)
#20	C20i	42	1/0/0	I	Invasive	2	90	5	0	10	Luminal B-like(HER2 negative)
#20	C20d				Ductal	2	90	5	0	5	Luminal B-like(HER2 negative)
#21	C21i	63	2/0/0	IIA	Invasive	2	100	0	0	15	Luminal B-like(HER2 negative)
#21	C21d				Ductal	2	90	0	0	10	Luminal B-like(HER2 negative)
#22	C22i	78	1/0/0	I	Invasive	2	100	5	1	20	Luminal B-like(HER2 negative)
#23	C23i	55	1/0/0	I	Invasive	3	100	0	0	25	Luminal B-like(HER2 negative)
#24	C24d	33	1/0/0	I	Ductal	2	100	60	0	5	Luminal A-like
#25	C25L	58	1/1/0	IIA	LNs meta	2	50	0	0	20	Luminal B-like(HER2 negative)
#26	C26i	58	1/1/0	IIA	Invasive	3	0	0	1	15	Triple negative
#26	C26L				LNs meta	3	0	0	0	15	Triple negative
#27	C27i	64	2/0/0	IIA	Invasive	3	0	0	0	10	Triple negative
#28	C28i	35	2/0/0	IIA	Invasive	2	0	0	0	1	Triple negative
#28	C28d				Ductal	2	0	0	0	1	Triple negative
#29	C29i	61	1/1/0	IIA	Invasive	2	0	0	0	3	Triple negative
#29	C29L				LNs meta	2	0	0	0	30	Triple negative
#30	C30i	59	1/0/0	I	Invasive	3	0	0	0	18	Triple negative
#30	C30d				Ductal	3	0	0	0	13	Triple negative
#31	D1	79		0	DCIS	1	100	15	0	1	Luminal B-like(HER2 negative)
#32	D2	51		0	DCIS	2	100	10	0	2	Luminal B-like(HER2 negative)
#33	D3	40		0	DCIS	2	60	60	1	4	Luminal A-like
#34	D4	52		0	DCIS	2	0	0	3	10	HER2 positive(non-luminal)
#35	D5	46		0	DCIS	2	90	90	0	13	Luminal A-like
#36	D6-m	49		0	DCIS	1	95	70	2	12	Luminal A-like
#36	D6-s				DCIS	1	99	30	2	10	Luminal A-like
#37	D7	53		0	DCIS	3	40	10	0	20	Luminal B-like(HER2 negative)
#38	D8	52		0	DCIS	2	90	95	2	7	Luminal A-like
#13	D9-w1	57		0	DCIS	1	100	50	1	7	Luminal A-like
#39	D10	64		0	DCIS	2	80	20	0	2	Luminal A-like
#40	D11	51		0	DCIS	1	90	60	1	1	Luminal A-like
#41	D12	44		0	DCIS	1	99	95	1	10	Luminal A-like
#41	P12			0	DCIS	1	80	80	0	3	Luminal A-like
#42	D13	63		0	DCIS	2	0	0	3	25	HER2 positive(non-luminal)
#43	D14	57		0	DCIS	3	0	0	3	10	HER2 positive(non-luminal)
#17	D15-w2	84		0	DCIS	2	0	0	3	10	HER2 positive(non-luminal)
#44	D16	68		0	DCIS	1	100	0	0	3	Luminal B-like(HER2 negative)
#45	D17	61		0	DCIS	2	100	0	3	10	Luminal B-like(HER2 positive)
#46	D18	52		0	DCIS	1	5	0	3	3	Luminal B-like(HER2 positive)
#47	D19	35		0	DCIS	1	100	0	1	5	Luminal B-like(HER2 negative)
#48	D20	56		0	DCIS	2	5	0	3	10	Luminal B-like(HER2 positive)
#49	D21	59		0	DCIS	2	100	30	0	5	Luminal A-like
#50	D22	49		0	DCIS	2	100	80	0	5	Luminal A-like

Supplementary file 2



Supplementary file 3

a



b