Assessment of clonal relationships in ipsilateral and bilateral multiple breast carcinomas by comparative genomic hybridisation and hierarchical clustering analysis

MR Teixeira*1,2, FR Ribeiro1, L Torres1, N Pandis3, JA Andersen4, RA Lothe5,6 and S Heim7

1Department of Genetics, Portuguese Oncology Institute, Porto, Portugal; 2Institute of Biomedical Sciences, University of Porto, Portugal; 3Department of Genetics, Saint Savas Hospital, Athens, Greece; 4Department of Pathology, Odense University Hospital, Odense, Denmark; 5Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway; 6Department of Molecular Biosciences, University of Oslo, Oslo, Norway; 7Department of Cancer Genetics, The Norwegian Radium Hospital-University Clinic, Oslo, Norway

The issue of whether multiple, ipsilateral or bilateral, breast carcinomas represent multiple primary tumours or dissemination of a single carcinomatous process has been difficult to resolve, especially for individual patients. We have addressed the problem by comparative genomic hybridisation analysis of 26 tumours from 12 breast cancer patients with multiple ipsilateral and/or bilateral carcinoma lesions. Genomic imbalances were detected in 25 of the 26 (96%) tumours. Using the genomic imbalances detected in these 26 lesions as well as those previously found by us in an independent series of 35 unifocal breast carcinomas, we compared a probabilistic model for likelihood of independence with unsupervised hierarchical clustering methodologies to determine the clonal relatedness of multiple tumours in breast cancer patients. We conclude that CGH analysis of multiple breast carcinomas followed by unsupervised hierarchical clustering of the genomic imbalances is more reliable than previous criteria to determine the tumours’ clonal relationship in individual patients, that most ipsilateral breast carcinomas arise through intramammary spreading of a single breast cancer; and that most patients with bilateral breast carcinomas have two different diseases.

British Journal of Cancer (2004) 91, 775–782. doi:10.1038/sj.bjc.6602021 www.bjcancer.com

Published online 13 July 2004
© 2004 Cancer Research UK

Keywords: multiple breast carcinomas; clonality; CGH; chromosome aberrations; hierarchical clustering

Breast cancer patients not rarely present multiple tumours (Lesser et al., 1982; Anastassiades et al., 1993; Vaidya et al., 1996). The increased emphasis in recent years on breast-sparing surgical treatment (Holland et al., 1985; Fisher, 1992; Connolly et al., 1995) has made it more clinically important to know whether multiple, ipsilateral breast carcinomas represent several primary tumours (multicentricity) or intramammary dissemination of a single carcinomatous process (multifocality). The fundamental biological interest of the issue is also profound, as it is also for bilateral breast carcinomas (Dawson et al., 1991; Sterns and Fletcher, 1991): are they two separate primary tumours or the metastasis from one breast to the other? Several criteria have been suggested to discriminate multiple primary tumours from metastatic lesions, including different histological growth patterns, the presence of in situ components in both tumours (for both ipsilateral and bilateral lesions), and localisation in different quadrants or more than 5 cm apart (for ipsilateral tumours; Dawson et al., 1991, 1995; Sterns and Fletcher, 1991; Dawson, 1993). However, the unspecific and quantitative nature of these criteria makes a reliable assessment of the competing hypotheses difficult in most cases.

A comparison of more fundamental biological characteristics of the different tumours could conceivably provide more reliable information as to whether they are clonally related or not. Among the principles used have been X-inactivation analysis (Noguchi et al., 1994a,b; Shibata et al., 1996), comparisons of allelic imbalance patterns (Tsuda and Hirohashi, 1995; Imyanitov et al., 2002), or the distribution of TP53 point mutations (Ackerman et al., 1995; Kinoshita et al., 1995; Shibata et al., 1996), but these studies depend heavily on statistical group comparisons and therefore are of limited value in individual patients. We have in the past successfully assessed the pathogenetic and clonal relationships among multiple breast tumours by means of chromosome banding analysis with karyotyping (Pandis et al., 1995; Teixeira et al., 1994, 1997). Since this technique provides screening information about all chromosome-level abnormalities that exist in individual cells, with each karyotypic change serving as a separate clonality marker, it is in principle highly efficient for the present purpose. However, because karyotyping is critically dependent on the successful in vitro culture and cytogenetic analysis of the separate cancer lesions, inadequate cell attachment, low in vitro mitotic index, and poor technical quality of chromosome preparations of the cells of the neoplastic parenchyma (even if for only one of the lesions) may render the case uninformative.

To circumvent the above-mentioned methodological limitations, we have used comparative genomic hybridisation (CGH), a
Probabilistic model of clonal relatedness

In order to evaluate the clonal relatedness between any two given tumours present in the same breast cancer patient, we used the probabilistic model described by Kuukasjärvi et al. (1997). A common clonal origin of two tumours can be inferred if they share a set of copy number gains and losses, not likely to be shared at random. In short, if \( a1, a2, a3, \) and so forth are specific genomic gains or losses, the probability of a particular abnormality occurring is obtained by dividing the number of occurrences of that change by the number of tumours analysed. If \( T1 \) is the set copy number changes of one tumour and \( T2 \) the set of copy number changes of another tumour, the events common to both tumours are defined by \( T1 \cap T2 \). Thus, a conservative estimate of the probability of a particular set of shared copy number changes occurring independently in two tumours at random is \( p(1) \times p(2) \times p(3) \times \ldots \times p(ck) = X \). If two tumours share no genetic changes, \( X \) is by definition 1. A measure of clonal relatedness between two tumours is \( 1 - X \), which equals 0 if no changes are shared and approaches 1 if \( X \) is small.

According to this model, a clonal relationship between two tumours can be inferred whenever \( X < 0.05 \) (clonal relatedness > 0.95). The probabilities of each copy number change occurring in each chromosome arm used in this probabilistic model were those previously reported by us in a series of 37 unifocal breast carcinomas (Teixeira et al., 2001), 35 of which were shown to harbour genomic imbalances by CGH. Copy number gains or losses at the same chromosome arm in different tumours of the same patient were considered the same only if the respective breakpoints did not differ by more than two bands.

Hierarchical clustering of CGH data

In order to code CGH data for use with clustering software, the copy number findings were registered in an Excel spreadsheet by each of two methods: by chromosome arm (41 data points) and by chromosome band (302 data points; in both strategies excluding the \( p \) arms of the acrocentric chromosomes). The codification of copy number changes by band was carried out as follows: no change, 0; gain, +1; loss, −1; amplification (ratio > 2), +2. The codification of copy number changes by chromosome arm was carried out as follows: no change, 0; gain, +1; loss, −1; amplification, +2; gain and loss from the same chromosome arm, −2. Each spreadsheet was then saved as a tab-delimited text file and loaded into the clustering software J-Express Pro (Molmine, Bergen, Norway).

Unsupervised hierarchical clustering of genomic imbalances was first performed with three pairs of primary breast carcinoma and its respective lymph node metastasis displaying different degrees of clonal relatedness (varying from \( 0.24 \times 10^{-7} \) to \( 0.7 \times 10^{-3} \), calculated as described above). Several software settings (including Euclidean distance vs Pearson correlation as distance metric and single vs complete linkage as cluster method) were tested to determine those that gave the best hierarchical clustering of each pair of primary tumour and lymph node metastasis (data not shown). The settings selected for the rest of the analyses were complete linkage (cluster method) and Pearson correlation (distance metric).

Besides coding the genomic imbalances separately by chromosome arm and by band, unsupervised hierarchical clustering analysis was tested as an alternative way to measure clonal relatedness among multiple breast tumours in two different ways. First, the genomic imbalances of the tumours of each breast cancer patient were compared with those of a previously reported series of 35 breast carcinomas with copy number changes (Teixeira et al., 2001). Second, the genomic changes of all 26 tumours of the 12 patients here reported were compared among themselves. The dendograms thus obtained group the tumours with the highest degree of genetic similarity in short-branched clusters, with longer...
branches indicating increasing genomic disparity. The clustering apart of tumours from the same patient, indicating that they were more similar with carcinomas of other patients than with one another, was taken as evidence of clonal unrelatedness. The clustering together of two tumours from the same woman indicated clonal relatedness among them, whereas the clustering of breast carcinomas from different women signified that they had followed similar pathogenetic pathways.

RESULTS

Comparative genomic hybridisation

All but one tumour presented copy number changes by CGH (Table 1). The number of imbalances per tumour ranged from 0 to 28 with an average of 8.2 per tumour. The number of copy number gains varied from 0 to 15 (average: 3.5) and the losses

| Case no. | Lab no. | Tumour | Diagnosis | CGH | Xb | Hierarchical clusteringc | Clonal relatednessd |
|----------|---------|--------|-----------|-----|----|---------------------|---------------------|
| 1        | 326A/92 | A      | COM       | rev sh enh(1p22pter,1q22q24,2q23q,12q3,7p14p21, 8q21pter,9pterq33,18p11q12,18q22q32,20p1pter), dim(Xp11p21,Xp22q24.4p14.3q34q35,5q12q14.6p24, 6p25,10p14p15,10q23,12q24pter,11q13pter,13q14q21, 13q33q34,15q24,16q22pter) | 2 × 10⁻⁴ | C | R |
| 2        | 368A/92 | A      | DC        | rev sh dim(1p34pter,2p22pter,15q3pter,16q22q14pter) | 0.26 | UC | UR |
| 3        | 426A/92 | B      | DC+CGH    | rev sh enh(1q7p11p21,16p10pter),dim(1p21pter,7q21pter,1q21q23,11q22qter,16q23qter,22qq) | 8 × 10⁻⁵ | C | R |
| 4        | 522A/92 | A      | DC        | rev sh enh(1q5.8,16p1lpter),dim(15q21q22,15q4q25,16q22q) | 0.002 | C | R |
| 5        | 197/92  | Right  | DC        | rev sh enh(1q25q32,1q41) | 1 | UC | UR |
| 6        | 307/92  | Right  | DC+CGH    | rev sh enh(1p31,1q23q42,3q26q27,6p21,6p23q12pter,12q15q21), dim(8p21pter,11q21pter,13q15q23,15q4q26, 16q17pter,17q24q25,18p11) | 0.09 | UC | UR |
| 7        | 308/92  | Right  | LC+LCS    | rev sh enh(8p12pter,16p),dim(9q15q31,10p12p14,11p14pter,12q22qter, 13q21q16) | 0.31 | UC | UR |
| 8        | 376/92  | Right  | LC+LCS    | rev sh enh(1q21q23,15q21qter,16p11pter),dim(3q21q25,5q14,6q14qter,8q12q21, 10q24qter,11p1q14pter,12p11pter,14q24,22q2pter) | 0.24 | UC | UR |
| 9        | 50/93   | Right  | DC+CGH    | rev sh enh(1q22q13),dim(6p22q6q21q27,11q21qter,16q24q22q, 17q12q21q3,13q33q33,16q23qter,18q22pter) | 1 | UC | UR |
| 10       | 588/93  | Right  | DC+CGH    | rev sh enh(8q12pter,16p),dim(8q21pter,9q22,16q) | 1 | UC | UR |
| 11       | 257/94  | Left   | LC        | rev sh enh(1q),dim(6q4q4ter,16q17pter,22q) | 0.001 | C | R |

*COM = comedo carcinoma; DC = ductal carcinoma; DCIS = ductal carcinoma in situ; LC = lobular carcinoma; LCIS = lobular carcinoma in situ; EH = epithelial hyperplasia. **X, probability of the shared genomic imbalances occurring just by chance (see Materials and methods). *Unsupervised hierarchical clustering analysis based on chromosome-arm coding of CGH copy number changes (see Materials and methods). C = clustered; UC = unclustered. **Clonal relatedness based on unsupervised hierarchical clustering analysis of CGH data coded by chromosome arm; R = related; UR = unrelated. *Calculation for comparing the two left and the two right tumours (both tumours of each side taken together).
from 0 to 14 (average: 4.7). The chromosome arms from which material was most frequently gained were 1q (62%), 16p (46%), 8q (42%), proximal 8p (38%), and 7p (27%), whereas the ones most often displaying losses were 16q (65%), 11q (38%), 22q (35%), 6q (31%), and distal 8p and 13q (23% each). One tumour showed genomic amplification (here defined as ratio above 2.0) at 8p12.

Clonal relatedness by the probabilistic model

The probability that the separate tumours of the 12 breast cancer patients share genomic imbalances just by chance is presented in Table 1. Whenever this probability was less than 0.05, a clonal relatedness between two tumours could be inferred with a probability greater than 0.95. With this method, a clonal relationship between the two tumours was found in three of four patients with ipsilateral breast tumours (75%) and in one of seven patients with bilateral breast carcinomas (14%). In the patient with both ipsilateral and bilateral multiple breast tumours (case 12), clonal relatedness was inferred between the two tumours in the left breast, a finding that was not confirmed by hierarchical clustering (see below).

Hierarchical clustering of CGH data

We first used unsupervised hierarchical clustering analysis of genomic imbalances coded by chromosome arm to measure clonal relatedness among multiple breast tumours. When the copy number changes of the tumours of each breast cancer patient were compared with those of a previously reported series of 35 breast carcinomas, a clonal relationship between the two tumours was found in three of the four pairs of ipsilateral breast tumours (Figure 1) and in one of seven pairs of bilateral breast carcinomas (Figure 2; Table 1). The findings for these 11 patients were completely concordant with those obtained with the probabilistic model described above. This was not so for the patient with both ipsilateral and bilateral multiple breast tumours, however, as hierarchical clustering indicated that all four tumours were clonally unrelated. The clonal relationships determined by hierarchical clustering of the chromosome arm

Figure 1 Comparative genomic hybridisation (A, B) and hierarchical clustering (C) of two ipsilateral breast tumours (case 1) together with 35 other breast carcinomas with copy number changes (coded by chromosome arm). Green bars to the right and red bars to the left of the chromosome ideograms indicate copy number gains and losses, respectively. See Table 1 for case numbers and a detailed description of the genomic imbalances. The two tumours cluster together (asterisks), showing that they are clonally related (multifocal breast carcinomas).
imbalances of all 26 tumours of the 12 patients here reported (Figure 3) were in all cases similar to that obtained after comparison of each patient with the separate series of breast carcinomas.

When we used unsupervised hierarchical clustering analysis of the genomic imbalances coded by chromosome band of all 26 tumours from the 12 breast cancer patients, the clonal relationships thus determined were in all cases identical to those described above with chromosome arm coding. The same was true for all but one case when the copy number changes coded by chromosome band present in the tumours of each of the 12 patients were compared against the separate series of 35 breast carcinomas; the exception was the apparent clonal relatedness between the two left tumours of case 12.

**DISCUSSION**

Comparative genomic hybridisation revealed genomic imbalances in 25 of the 26 (96%) breast tumours present in the 12 breast cancer patients, underscoring the considerable potential of the present methodology when it comes to determining the clonal relatedness among two or more tumours in the same patient. The overall pattern of copy number changes detected in this series of 26 tumours does not differ significantly from that previously found by us in an independent series of 37 breast cancer patients (Teixeira et al, 2001), the most frequent genomic imbalances being gain of 1q, 16p, 8q, proximal 8p, and 7p and loss of 16q, 11q, 22q, 6q, distal 8p, and 13q. Some discrepancies, such as the apparently increased frequency of 16q and 22q loss in the present series, are probably due to the fact that some of the 26 tumours were not truly independent and the differences become insignificant when only pathogenetically independent carcinomas are considered.

One way to find out whether two ipsilateral or bilateral breast tumours are two primary carcinomas or represent metastatic spreading of a single neoplastic process is to use a probabilistic model based on the frequency of genomic imbalances in an unselected series of cases (Kuukasjärvi et al, 1997). This model estimates the probability that shared genomic imbalances in two
predictions are counterintuitive when one looks at the strikingly low likelihood (0.09). These findings leave little doubt that some ipsilateral tumours are pathogenetically more similar to carcinomas in other women than to their contralateral tumours. Of these six patients with two tumours compared, the proper clustering of primary tumours and their respective metastases took place irrespective of whether the CGH imbalances were coded by chromosome arm or by chromosome band. The only discrepancy in the results obtained by the two coding methods occurred for case 12, in which clustering of the 26 tumours by chromosome band revealed an apparent clonal relatedness between the two tumours, whereas even if both result from single events.

Unsupervised hierarchical clustering of CGH data revealed that multiple, ipsilateral breast tumours were clonally related in three of four patients, both when the copy number changes of the tumours of each breast cancer patient were compared with those of a separate series of 35 breast carcinomas and when comparison was with the 26 tumours of this series. This finding confirms our previous conclusion based on chromosome banding that most ipsilateral breast tumours arise through intramammary spreading of a single breast cancer (multifocal carcinomas; Teixeira et al, 1994, 1997). The fact that two of the three multifocal carcinomas had carcinoma in situ lesions in both tumours, and that these were separated by 10–40 mm of macroscopically normal-looking breast tissue, indicated that breast cancer can spread extensively via the ductal system (Ohtake et al, 1995; Middleton et al, 2002) and illustrates the limitations of these morphological criteria when it comes to discriminating multifocal from multicentric breast carcinomas (Dawson, 1993; Dawson et al, 1995; Tsuda and Hirohashi, 1995; Middleton et al, 2002). One case was classified as multicentric carcinoma by all the strategies here employed in spite of the fact that the tumours shared an 8p loss. The relatively high frequency of 8p loss in breast cancer, which makes it a marker of only poor information value, and the presence of another nine genomic changes, none of which were shared between the two tumours, leave little doubt that some ipsilateral tumours are pathogenetically independent.

The bilateral breast tumours of six of seven patients segregated apart after hierarchical clustering, indicating that they were clonally unrelated. This shows that these breast carcinomas were pathogenetically more similar to carcinomas in other women than to their contralateral tumours. Of these six patients with two genetically disparate breast carcinomas, different histological types were found in two (cases 6 and 9 had a lobular and a ductal carcinoma; three patients had two ductal carcinomas and the remaining presented two lobular carcinomas) and carcinoma in situ components were present in both tumours of five cases, thus
supporting the conclusion that most patients with bilateral breast carcinomas have two different diseases (Dawson et al., 1991; Sterns and Fletcher, 1991; Pandis et al., 1995; Shibata et al., 1996; Imyanitov et al., 2002). This notwithstanding, one of the patients presented two bilateral lobular carcinomas sharing all five genomic imbalances by CGH, strongly arguing that one tumour must have metastasised to the contralateral breast, as we have also seen occasionally before (Pandis et al., 1995). We favour this interpretation over the hypothesis that a common milieu may have given rise to two clonally independent, but pathogenetically identical, breast carcinomas (Imyanitov et al., 2002); if this was a relevant mechanism in breast carcinogenesis, pathogenetically similar bilateral breast carcinomas should be the rule, not the exception.

We show that assessment of the clonal relationship among multiple breast tumours is significantly more informative by CGH (92% of the individual patients were successfully examined) than by chromosome banding analysis after short-term culturing (success rates of between two-thirds and three-fourths of the cases; Pandis et al., 1995; Teixeira et al., 1997). When the same cases were analysed by both techniques, discordant conclusions were obtained in a few of them (Teixeira et al., 1994, 1996, 1997; Pandis et al., 1995). Clonal selection during in vitro culture impairing detailed cytogenetic analysis of all tumour cell subpopulations and CGH failing to detect small clones or balanced genomic aberrations (Persson et al., 1999; Teixeira et al., 2001; Kleivi et al., 2002) may contribute to these discrepancies.

We conclude that CGH analysis of multiple breast carcinomas followed by unsupervised hierarchical clustering of the genomic imbalances coded by chromosome arm is more effective than previously used criteria to determine their clonal relationship in individual patients. To know precisely which ipsilateral or bilateral breast carcinomas are multiple primaries could help delineate optimal treatment strategies for these patients as well as identify families with increased susceptibility to breast cancer.

ACKNOWLEDGEMENTS

The financial support of The Norwegian Cancer Society is gratefully acknowledged (grants to RAL and SH). FRR is a fellow of Fundação para a Ciência e a Tecnologia. LT is a fellow of Liga Portuguesa Contra o Cancro, Centro Regional do Norte.
Clonal origin of multiple breast tumours

MR Teixeira et al

Teixeira MR, Pandis N, Heim S (2002) Cytogenetic clues to breast carcinogenesis. *Genes Chromosomes Cancer* 33: 1–16

Teixeira MR, Tsarouha H, Kraggerud SM, Pandis N, Dimitriadis E, Andersen JA, Lothe RA, Heim S (2001) Evaluation of breast cancer polyclonality by combined chromosome banding and comparative genomic hybridization analysis. *Neoplasia* 3: 204–214

Tsuda H, Hirohashi S (1995) Identification of multiple breast cancers of multicentric origin by histological observations and distribution of allele loss on chromosome 16q. *Cancer Res* 55: 3395–3398

Vaidya JS, Vyas JJ, Chinoy RF, Merchant N, Sharma OP, Mittra I (1996) Multicentricity of breast cancer: whole organ analysis and clinical implications. *Br J Cancer* 74: 820–824