Breast-conserving therapy is the preferred treatment for patients with early-stage breast cancer (1). It offers equal local control and overall survival (2) and superior psychosocial outcomes (3,4) compared with modified radical mastectomy. However, an ipsilateral breast cancer recurrence can be traumatizing and can lead to death (2).

When an ipsilateral breast cancer develops, the new tumor can either be a true recurrence—that is, a regrowth of clonogenic cells that were not removed by surgery or killed by radiotherapy—or a new primary tumor that arises from the remaining breast tissue (5). Several definitions have been used to distinguish true recurrences from new primary tumors. Initially, these distinctions were based

**Background**

To distinguish new primary breast cancers from true recurrences, pangenomic analyses of DNA copy number alterations (CNAs) using single-nucleotide polymorphism arrays have proven useful.

**Methods**

The pangenomic profiles of 22 pairs of primary breast carcinoma (ductal or lobular) and ipsilateral breast cancers from the same patients were analyzed. Hierarchical clustering was performed using CNAs and DNA breakpoint information. A partial identity score developed using DNA breakpoint information was used to quantify partial identities between two tumors. The nature of ipsilateral breast cancers (true recurrence vs new primary tumor) as defined using the clustering methods and the partial identity score was compared with that based on clinical characteristics. Metastasis-free survival was compared among patients with primary tumors and true recurrences as defined using the partial identity score and by clinical characteristics. All statistical tests were two-sided.

**Results**

All methods agreed on the nature of ipsilateral breast cancers for 14 pairs of samples. For five pairs, the clinical definition disagreed with both clustering methods. For three pairs, the two clustering methods were discordant and the one using DNA breakpoints agreed with the clinical definition. The partial identity score confirmed the nature of ipsilateral breast cancers as defined by clustering of DNA breakpoints in 21 of 22 pairs. The difference in metastasis-free survival of patients with new primary tumors and those with true recurrences was not statistically significant when tumors were defined based on clinical and histologic characteristics (5-year metastasis-free survival: 76%, 95% confidence interval [CI] = 52% to 100% for new primary tumors and 38%, 95% CI = 17% to 83% for true recurrences; \( P = .18 \); new primary tumor vs true recurrence, hazard ratio = 2.8, 95% CI = 0.6 to 13.7), but the difference was statistically significant when tumors were defined using the partial identity score (5-year metastasis-free survival: 100% for new primary tumors and 29%, 95% CI = 11% to 78% for true recurrences; \( P = .01 \)).

**Conclusions**

DNA breakpoint information more often agreed with the clinical determination than CNAs in this population. The partial identity score, which was calculated based on DNA breakpoints, allows statistical discrimination between new primary tumors and true recurrences that could outperform the clinical determination in terms of prognosis.

J Natl Cancer Inst 2008;100:48–58

**Affiliations of authors:** Département d’oncologie radiothérapie (MAB, AF), Service de Bio-informatique (NS, PN, GR, PH, EB), Département de Transfert (CD, JPM, JPT), Département de Biologie des tumeurs (IL, BSZ), Service de Biostatistiques (YDR, AS), and Centre National de la Recherche Scientifique, Unité Mixtes de Recherche 144 (CD, PH), Institut Curie, Paris, France; Institute of Molecular and Cell Biology Biopolis A*STAR, Singapore (JFT).

**Correspondence to:** Marc A. Bollet, MD, Département d’oncologie radiothérapie, Institut Curie, 26, rue d’Ulm, 75248 Paris cedex 05, France (e-mail: marc.bollet@curie.net).

See “Funding” and “Notes” following “References.”

DOI: 10.1093/jnci/djm266

© The Author 2007. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oxfordjournals.org.
on the location of the ipsilateral breast cancer (ie, the farther from the initial primary tumor, the more likely it is to be a new primary tumor) and on shared common histopathologic criteria (eg, type, grade, and hormone receptor status) (6–10). In the quest for additional ways to distinguish new primary breast tumors from true breast cancer recurrences, biologic studies of clonal relationships between the new and original tumor have also been performed. These studies have relied on ploidy (5,11), loss of heterozygosity (12–14), p53 analysis (15), or X chromosome inactivation (16) or have been based on DNA copy number alterations (CNAs) (17–19). CNA data can be obtained by high-resolution techniques, such as array-based comparative genomic hybridization or single-nucleotide polymorphism (SNP) arrays (20). One of the most commonly used ways to look at clonal relatedness using pangenomic data is to perform an unsupervised hierarchical clustering that organizes primary breast tumors and ipsilateral breast cancers on the basis of their overall genomic similarity (18,19). These measures of similarity are summarized in a dendrogram, in which the pattern and length of the branches reflect the relatedness of the samples in terms of DNA CNAs.

Changes in DNA copy numbers occur at chromosomal locations called breakpoints. We hypothesized that the precise locations of these breakpoints could serve as markers for clonal relatedness and that we could distinguish true recurrences from new primary tumors by the number of common breakpoints in the ipsilateral breast cancer and the primary tumor. In this study, we first aimed to test the added value of examining the clustering of breakpoints (over CNAs) when determining the nature of the ipsilateral breast cancer. Second, we aimed to develop a score to quantify the partial identity between two tumors according to their clonal relatedness (determination of the partial identity score). Third, we examined prognosis in terms of metastasis-free survival. In each case, these methods were compared with the clinical determination of the nature of the ipsilateral breast cancer.

Subjects and Methods
Selection of Patients
Specimens from patients with primary breast cancers and ipsilateral breast cancers were selected from freshly frozen samples of the Institut Curie tissue bank according to the following criteria: the primary tumor was either ductal or lobular invasive breast carcinoma; the patient was 49 years or younger at diagnosis of the initial tumor; all patients were premenopausal; and there was no previous history of cancer, except for one nonmelanoma skin cancer. All tumor; all patients were premenopausal; and there was no previous history of cancer, except for one nonmelanoma skin cancer. All patients had been treated at the Institut Curie by breast-conserving surgery, including dissection of the axillary lymph nodes in most patients, followed by radiotherapy to the breast with or without a boost to the tumor bed (external beam radiotherapy or brachytherapy) and/or to the regional lymph node–bearing areas if indicated and, when required, systemic treatment as part of their initial management. For all tumors, histopathologic characteristics were reviewed by one pathologist (B. Sigal-Zafrani).

To ensure that the data would be informative, we restricted genomic analyses to tumors (primary and recurrences) in which at least 50% of cancer cells had been assessed by hematoxylin, eosin, and saffron staining of sections from snap-frozen samples. This study reports a series of 22 patients with assessable pairs of primary breast tumors and ipsilateral breast cancers.

Prior knowledge
Detecting changes in DNA copy number using single nucleotide polymorphism arrays has been a useful tool in distinguishing new primary breast tumors from recurrences.

Study design
Comparison of hierarchical clustering of DNA copy number and DNA breakpoints, an identity score based on the DNA breakpoint information, and clinical characteristics to accurately designate ipsilateral breast tumors as new primary tumors or true recurrences in breast tumor pairs from 22 patients.

Contributions
For 14 of the pairs, all methods agreed on the designation of the ipsilateral breast cancer as a new primary tumor or a true recurrence; however, for five pairs and three pairs, both clustering methods and clustering by DNA breakpoints, respectively, agreed with the clinical definition. For 21 pairs, the partial identity score confirmed the designation of the tumor as defined by both clustering methods. Patients with recurrences had poorer metastasis-free survival than patients with new primary tumors, according to the partial identity score, but this difference was not statistically significant using the clinical definition.

Implications
The partial identity score may outperform clinical determination for the prognosis of ipsilateral breast cancers.

Limitations
Freshly frozen tissue samples that contain a large number of cells from both the initial primary tumor and the ipsilateral tumor are needed to perform the DNA breakpoint analyses.

Clinical and Histologic Studies
The histologic/biologic properties of the breast cancers were determined by subjecting tissue sections to immunohistochemical analysis for the estrogen receptor (clone 6F11, 1:200 dilution; Novocastra, Newcastle Upon Tyne, England) and progesterone receptor (clone 1A6, 1:200 dilution; Novocastra) antibodies. Tumors were considered to be positive for these receptors if at least 10% of the invasive tumor cells in a section showed nuclear staining (21).

In accordance with theories of the clonal evolution of tumor cell populations, ipsilateral breast cancers were clinically defined as true recurrences if they had the same histologic subtype (ductal or lobular).
Genomic Studies

Total genomic DNA was extracted from tissue samples using a variation of the standard phenol:chloroform protocol (23). Genomic DNA was quantified by spectrophotometry using a ND-1000 Spectrophotometer (NanoDrop, Wilmington, DE), and quality was assessed by 0.8% agarose gel electrophoresis.

Genomic DNA from each sample was prepared for microarray hybridization using the GeneChips Mapping 50K Xba Assay Kit (Affymetrix Inc., Santa Clara, CA). Briefly, 250 ng of total genomic DNA was digested with the restriction enzyme XbaI and ligated to an adaptor sequence (XbaI adaptator: 5′-ATTATGAGCACGAGAAGCGCTGATCT-3′ and 5′-CTAGAGATCGGCGCTCTGTCGTGCTCATAA-3′) that recognizes the cohesive four base pair (bp) region (5′-GATC-3′). A generic primer (5′-ATT ATG AGC ACG ACA GAC GCC TGA TCT-3′) that recognizes the adaptor sequence was used to preferentially amplify adaptor-ligated DNA fragments 250–2000 bp in size by the optimized polymerase chain reaction (PCR) conditions, according to the manufacturer’s instructions. The amplified DNA was then fragmented by DNase treatment and hybridized to the Affymetrix GeneChips Human Mapping 50K array Xba 240 (Affymetrix), according to the manufacturer’s instructions. Washing, staining, and scanning of chips were performed using materials and methods provided by the manufacturer. The pangenomic profiles of the 22 pairs of primary tumors/ipsilateral breast cancers are available on ACTuDB (24) (http://bioinfo.curie.fr/actudb/). Human mapping 50K array Xba 240 annotations and sequence files are available on the Affymetrix website (http://www.affymetrix.com/support/technical/byproduct.affx?product=100k).

Metastasis-Free Survival

Metastasis-free survival was estimated by the Kaplan–Meier method (25) and compared between the groups of patients defined as having been diagnosed with either a true recurrence or a new primary tumor using the log-rank test. The confidence interval (CI) of the hazard ratio was obtained using a semiparametric Cox model (26).

Statistical Methods

Copy Number Alteration Determination. SNP data were gathered from the pangenomic profile and analyzed using the iterative and alternative normalization of copy number SNP array (ITALICS) algorithm with default parameters, which simultaneously normalizes the genomic profile and detects the biologic signal. Briefly, ITALICS alternatively estimates the biologic signal (i.e., the DNA copy number at each SNP locus) with the gain and loss analysis of DNA algorithm (27) and normalizes the data to correct the nonrelevant effects (CG content and fragment length of PCR products, oligonucleotide CG content, and SNP effect). These two steps are repeated iteratively to improve the biologic signal estimation until no more improvement is seen. ITALICS outperforms other methods of normalization. The result of this process is a segmented genomic profile that consists of regions of homogeneous DNA and information on their corresponding copy numbers. Each region is given a smoothing value (i.e., the median of the SNP copy numbers within the region) and a status (i.e., gain, normal, or loss).

We defined a breakpoint as 1) a SNP locus located at a change of status (e.g., normal/gain or gain/loss) or as 2) a SNP locus located at a change of smoothing value that occurred within a region of gain or loss, thus defining different levels of gain or loss among these regions. Additional breakpoints were also added at the extremities of the chromosome to take into account their gain or loss whenever applicable. Because some breakpoints could be due to copy number variations that occur in healthy individuals, breakpoints arising in the copy number variable regions in the HapMap collection (28) were excluded. The visualization and further analysis of the data was performed through a graphic user interface, Visualization and analysis of array CGH, transcriptome and other molecular profiles (29).

Hierarchical Clustering. Similarity between genomic profiles. We considered two measures of similarity among the genomic profiles of a primary tumor and ipsilateral breast cancer. First, we used the Pearson correlation between their CNA profiles. Second, we used a measure $M$ that is derived from the percent concordance proposed by Waldman et al. (18) and adapted from Dice’s formula (30) and corresponds to the number of common breakpoints divided by the mean number of breakpoints in either a primary tumor or an ipsilateral breast cancer. $M$ is computed as follows, for a $(i,j)$ pair:

$$M_{ij} = \frac{\#(S_i \cap S_j)}{1/2 \times (\#S_i + \#S_j)},$$

in which $S_i$ and $S_j$ are the subsets of breakpoints present in the SNP arrays of the primary tumor, $i$, and ipsilateral breast cancer, $j$. An example of $M$ is given in Supplementary Fig. 1 (available online).

Two tumors had common breakpoints if the following conditions were fulfilled: 1) the changes in copy number occurred at the exact same locus and 2) the changes in copy number were of the same nature (i.e., either an increase or a decrease in numbers) in the two tumors.

Assessing clonal relatedness from a dendrogram. We assumed that clonal unrelatedness was revealed by the clustering apart of the two tumors (primary tumor and ipsilateral breast tumor) from the same patient, reflecting that they were more similar to carcinomas of other patients than to each other. In contrast, the clustering together of two tumors from the same patient indicated clonal relatedness among them. For both measures of similarity (Pearson coefficient and $M$ measure), we used Ward’s criteria (31) as an agglomerative method in the hierarchical clustering.

Partial Identity Score. Score definition. To distinguish true recurrences from new primary tumors, we developed a partial identity score
Patients were treated with tamoxifen for 5 years. Chemotherapy consisted of 5-fluorouracil, anthracyclines, and cyclophosphamide.

Table 1. Patient and tumor characteristics of the 22 patients whose tumors (both PT and IBC) had exploitable SNP arrays

<table>
<thead>
<tr>
<th>Pair</th>
<th>Age, y</th>
<th>Family</th>
<th>Prob</th>
<th>BRCA1</th>
<th>BRCA2</th>
<th>pT</th>
<th>pN</th>
<th>Surgical margin, mm</th>
<th>Radiotherapy dose, GY</th>
<th>No. of cycles of chemotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>23.1</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>≥4</td>
<td>54</td>
<td>4</td>
</tr>
<tr>
<td>P2</td>
<td>42.1</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>≥4</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>P3</td>
<td>42.6</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>≥4</td>
<td>54</td>
<td>4</td>
</tr>
<tr>
<td>P4</td>
<td>48.2</td>
<td>1</td>
<td>44</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>≥4</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>P5</td>
<td>45.5</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>1</td>
<td>≥4</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>P6</td>
<td>35.7</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0−3</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>P10</td>
<td>46.2</td>
<td>1</td>
<td>95</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>≥4</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>P11</td>
<td>49.0</td>
<td>1</td>
<td>95</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>≥4</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>P12</td>
<td>48.9</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>≥4</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>P13</td>
<td>45.0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
<td>0</td>
<td>≥4</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>P14</td>
<td>43.6</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>0−3</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>P15</td>
<td>46.1</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>≥4</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>P16</td>
<td>48.4</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>≥4</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>P17</td>
<td>27.9</td>
<td>1</td>
<td>82</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0−3</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>P19</td>
<td>49.1</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
<td>0</td>
<td>0−3</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>P20</td>
<td>47.1</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
<td>1</td>
<td>0−3</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>P21</td>
<td>46.3</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>DCIS</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>P22</td>
<td>35.0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
<td>2</td>
<td>≥4</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>P23</td>
<td>30.8</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
<td>0</td>
<td>≥4</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>P24</td>
<td>47.7</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>1</td>
<td>≥4</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td>P25</td>
<td>43.0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>0</td>
<td>0−3</td>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>P26</td>
<td>30.5</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>≥4</td>
<td>52</td>
<td>4</td>
</tr>
</tbody>
</table>

* PT = primary tumor; IBC = ipsilateral breast cancer; SNP = single nucleotide polymorphism; Family = family history of breast cancer in the first two degrees (0 = no, 1 = yes); Prob = age-specific risk estimates of breast cancer according to the Claus Model (32); BRCA1 and BRCA2 = mutation found in BRCA1 and BRCA2 (0 = not found, 1 = deleterious, 2 = possibly deleterious, NA = not available); pT = histologic tumor classification according to Union Internationale Contre le Cancer (UICC) (33); pN = histologic lymph node classification according to UICC; DCIS = ductal carcinoma in situ.

† Chemotherapy consisted of 5-fluorouracil, anthracyclines, and cyclophosphamide.

‡ Patients were treated with tamoxifen for 5 years.

that is based on the $M$ measure of similarity described above. The score reflects the number of common breakpoints among the ipsilateral breast cancer and the primary tumor. In addition, because very frequent breakpoints may be less informative than frequent ones in estimating the clonal relatedness between two tumors, the added value of each breakpoint was weighted according to its frequency among the samples of 44 control patients. The partial identity score (PS) was thus

$$PS_{ij} = \frac{1}{2k} \times \left( \frac{1}{k} \sum_{k=1}^{k_x} (1 - F_k) \right) \times \left( \frac{1}{k} \sum_{k=1}^{k_y} (1 - F_k) \right),$$

in which $F_k$ represents the frequency of appearance of the breakpoint $k$ calculated in the series of the 44 control breast cancers. An example of a partial identity score is given in Supplementary Fig. 1 (available online).

**Statistical testing for partial identity.** The partial identity score was calculated for all 462 possible “artificial pairs” ($462 = 22 \times 21$, because each of the 22 primary tumors could be artificially paired with the ipsilateral breast cancer of the 21 other patients, see Table 3 notes). The distribution under the null hypothesis, $H_0$, of no partial identity between the two tumors was estimated using all 462 possible artificial pairs. We rejected $H_0$ with a type I error fixed at 5%, that is, we considered that a local recurrence shared partial identity with a primary tumor when the score was higher than the upper 5th percentile in the distribution of artificial pairs. The score was then calculated for the “natural pairs,” that is, a primary tumor and its ipsilateral breast cancer occurring in the same patients (see Table 3 notes). Ipsilateral breast cancers from pairs with scores higher than this cutoff, that is, with shared partial identity, were considered to be true recurrences.

**Robustness of the score.** The robustness of the partial identity score was assessed by randomly selecting two subgroups of 15 and 7 patients from the population of 22 breast cancer patients. The first subgroup of 15 patients was used to compute the scores of the artificial pairs and to record the cutoff score corresponding to the 95th percentile. This score was then used to determine the status of each of the natural pairs in the seven patients of the other subgroup. To make the comparison statistically sound, each process was repeated 1000 times. The variation of the cutoff scores was assessed by box plot representation. The consistency of the ipsilateral breast cancer status was calculated as the percentage of extractions when the status of this pair was respectively a true recurrence or a new primary tumor.

All statistical tests were two-sided. $P$ values less than .05 were considered to be statistically significant.

**Results**

**Clinical and Histologic Features**

The clinical and tumor characteristics of 22 patients whose tumors had exploitable SNP arrays were analyzed (Tables 1 and 2). According to clinical and histologic criteria (Table 2), nine of the 22 ipsilateral breast cancers were new primary tumors and the other...
13 were true recurrences. Ipsilateral breast cancers occurred at a median time of 3.1 years after the initial breast cancer diagnosis (range = 0.8–6.5 years). In three of 22 (14%) patients, ipsilateral breast cancers occurred in a different quadrant than the initial tumor; all of these were defined clinically as new primary tumors.

**Genomic Studies**

The pangenomic profiles of a primary tumor and its ipsilateral breast cancer revealed common breakpoints, with a precision within a SNP that can be used as markers of their clonal relatedness. Pair 5 is given as an illustration (Fig. 1).

The median number of breakpoints per array was statistically significantly higher for ipsilateral breast cancers (median = 71, range = 21–433) than for primary tumors (median = 52, range = 4–646) ($P = .001$) (Table 3). The mean number of common breakpoints per pair was also statistically significantly higher for natural pairs (mean = 18.8, SD = 18.8) than for artificial pairs (mean = 4.1, SD = 3.1) ($P = 0.5 \times 10^{-9}$).

**Clustering by Copy Number Alterations or Breakpoints**

According to hierarchical clustering by DNA CNAs (Fig. 2) and by breakpoints (Fig. 3), five and six ipsilateral breast cancers, respectively, were new primary tumors. The two clustering methods and the clinical definition agreed for 14 pairs (Table 2). However, for five pairs (P6, P12, P16, P20, P22), the clinical definition disagreed with both clustering methods and, for three others (P1, P2, P15), the clustering by breakpoints disagreed with that by CNAs but agreed with the clinical definition. The recurrences in pairs 1 and 2 were identified as true recurrences by the CNA clustering but as new primary tumors by the clinical definitions because of the reappearance of estrogen receptors in the pair 1 ipsilateral breast cancer and different histologic type (ductal instead of lobular carcinoma) in pair 2. In pair 15, CNA clustering did not find a true recurrence, whereas the clinical definition did. No statistically significant differences in clinical and histologic characteristics between the patients diagnosed with new primary tumors or true recurrences were observed by breakpoint information, apart from a suggestion for patients with new primary tumors to be younger and to have a more frequent family history of breast cancer (Supplementary Table 1, available online).

**Partial Identity Score**

According to the partial identity score reported for each pair in Table 2, 15 ipsilateral breast cancers were true recurrences and seven were new primary tumors (Fig. 4). With a type I error set at 5%, the partial identity score disagreed with clustering by breakpoints in pair 12 only; the clinical definition was new primary tumor because of a change in tumor location. When the score was determined according to Waldman’s percent of concordance without either weighing the influence of the coexistence of breakpoints according to their frequency in a similar population or excluding
the breakpoints that occur in the copy number variable regions in the HapMap collection, the attribution of the status of three pairs (20 changed from a true recurrence to a new primary, whereas 6 and 12 became true recurrences) and two pairs (10 and 12 changed from new primaries to true recurrences) changed, respectively.

The status of all pairs was confirmed by the 1000 random extractions (Supplementary Table 2, available online). The mean cutoff value was 0.1203 (SD = 0.0102) (Supplementary Fig. 2, available online). The cutoff used to determine the status of the 22 ipsilateral breast cancers, which was defined using all 462 artificial pairs, was 0.1212.

Prognostic Value of the Determination of the Nature of the Ipsilateral Breast Cancer

Patients who were diagnosed with true recurrences had lower metastasis-free survival than those diagnosed with new primary tumors (Supplementary Fig. 3, available online). The difference in metastasis-free survival in the two groups was not statistically significant when they were defined based on clinical and histologic characteristics (5-year metastasis-free survival: 76%, 95% CI = 52% to 100% for new primary tumors and 38%, 95% CI = 17% to 83% for true recurrences; $P = .18$; primary tumors vs true recurrences, hazard ratio = 2.8, 95% CI = 0.6 to 13.7). However, metastasis-free survival was different when the groups were defined according to the partial identity score (5-year metastasis-free survival: 100% for new primary tumors and 29%, 95% CI = 11% to 78% for true recurrences; $P = .01$).

Discussion

DNA breakpoint information was more often in agreement with the clinical definition than that from CNAs to define true recurrences among ipsilateral breast cancers in this population. We developed a partial identity score that is based on DNA breakpoints, which allowed statistical discrimination between new primary tumors and true recurrences. This score outperformed the clinical prognosis determination in terms of metastasis-free survival.

We chose to base our study on a series of young (<50 years old) premenopausal women not only because young age is recognized as one of the most important independent prognostic factors for ipsilateral breast recurrence (34–40) but also to ensure a very high level of homogeneity. In addition, all patients had undergone breast-conserving surgery followed by whole-breast radiotherapy for their initial breast cancers, which were selected as either ductal or lobular invasive carcinomas, and were treated at the same cancer center.

Our results show that some ipsilateral breast cancers share with their primary tumors many DNA CNAs breakpoints at the same locations (precision to within a SNP, as illustrated in Fig. 1). From these observations, we produced a method of determining true recurrences that relies on a number of assumptions. The first and most obvious is that the vast majority of breast cancers are of clonal origin. The second is that a tumor retains a substantial number of genomic alterations throughout its evolution. The third assumption, which is key to the method that we have developed, is that the exact locations of the breakpoints that are on the edge of a given change in DNA copy numbers are better hallmarks of a given tumor than the magnitude or width of the genomic alteration itself. For example, because the deletion that causes the loss of Phosphatase and TENsin homolog (PTEN) alters regulatory pathways that lead to precocious development and neoplasia in the mammary gland (41), it can be found in many breast cancers (42–44); however, the exact location of the breakpoints bordering this deletion can be specific to a given tumor. We provide as an
example (Supplementary Fig. 4, available online) the prototype case of PTEN deletion in which the breakpoints are identical between the primary tumor and ipsilateral breast cancer of pair 5 and yet differ in all the other tumors that also harbor a loss of PTEN.

Because clustering is commonly used to determine whether two tumors are clonally related and because it performs better than CNAs by comparing clustering by CNAs and by breakpoints to determine the nature of the ipsilateral breast cancer. We concluded from the comparison of clusterings of CNAs of breakpoints to determine the nature of the ipsilateral breast cancer rather than at CNAs by comparing clustering by CNAs and by breakpoints to determine the nature of the ipsilateral breast cancer.

In addition, we chose to weigh the influence of a common breakpoint in a population of similar tumors. This weighting changed the determination of three of 22 pairs.

The clinical definition considered an ipsilateral breast cancer as a new primary tumor when the partial identity score did not include in the study, the ipsilateral breast cancer from pair 6 would have been considered as a true recurrence rather than a new primary tumor. Conversely, the assessment of the partial identity score robustness was satisfactory with a narrow range of the cutoff (Supplementary Table 2, available online) and with the consistency of the ipsilateral breast cancer status (Supplementary Table 2, available online). Moreover, a score allows one to choose the cutoff that best distinguishes new primary tumors from true recurrences. In this study, we chose a type I error rate at 5% to favor sensitivity for diagnosing true recurrences over the specificity. Further studies will be needed to verify the biologic validity of this choice (Supplementary Fig. 3, available online).

In addition, we chose to weigh the influence of a common breakpoint between the ipsilateral breast cancer and its primary tumor by a factor that takes into account the frequency of this given breakpoint in a population of similar tumors. This weighting changed the determination of three of 22 pairs.

The clinical definition considered an ipsilateral breast cancer as a new primary tumor when the partial identity score did not in three instances. In the first because of a change in location for pairs 12, 13 and 20, in the second because of a lesser degree of differentiation for pair 12 and 20, in the third because of a change in histology for pairs 12 and 20.
A criticism that can be made of the clinical definition is that it assumes that a true recurrence is derived from its primary tumor instead of only being related to it. A true recurrence, according to some clinical definitions (5,6,11), cannot be more differentiated than its primary tumor. Usual classifications define differentiation according to histologic grading, DNA ploidy, or the presence of ductal carcinoma in situ. They are based on the assumption that tumors accumulate genetic alterations with time (22,45,46) and that the chronologic order of these alterations reflects the development of a tumor clone. This assumption is, however, challenged by the fact that the ipsilateral breast cancers are neither more aggressive nor more undifferentiated than their primary tumors (47).

The situation with pair 22 illustrates another possible limitation of histologic determination. Here, the clinical status of the ipsilateral breast cancer was of a new primary tumor because its histologic type was a micropapillary carcinoma, whereas the initial tumor was a ductal carcinoma. However, after further histologic analysis, a minor component of micropapillary carcinoma was revealed in the initial carcinoma that otherwise would have been overlooked (Supplementary Fig. 5, available online). This finding implies that, in some instances, the current histologic taxonomy, which is based more on architectural features than on biologic ones, could become obsolete and that some ipsilateral breast cancers could qualify as true recurrences without sharing the same histologic type as their primary tumors.

We observed that patients with true recurrences had lower metastasis-free survival than patients with new primary tumors and that this difference became statistically significant when the partial identity score, instead of clinical definition, was used to define ipsilateral breast cancer types. This observation has been shared by many authors (5,6,10,12). Possible explanations are,
first, that a true recurrence is the expression of clones that are resistant to adjuvant treatment and therefore could be more difficult to eradicate and, second, that it could be the tip of the iceberg, that is, distant metastases. Conversely, new primary tumors have a prognosis similar to de novo primary cancers but can also reflect a genetic predisposition to develop breast cancer, in the contralateral breast in particular. The clinical implication should therefore be to advocate the use of a systemic treatment in the case of true recurrences and the use of either chemoprevention, such as hormone therapies (48–50) or screening with magnetic resonance imaging (51–53), for patients who are diagnosed with new primary tumors. Here, using breakpoint information led to a better discrimination between new primary tumors and true recurrences in terms of metastasis-free prognosis than the clinical definition.

We also hope that a better distinction among ipsilateral breast cancers of tumors that are genetically related to their primary tumors, that is, true recurrences, will help reveal genetic differences that would provide new information on radioresistance and tumor aggressiveness. To date, little is known about the differential or similarity of the pangenomic expression or the nature of both new primary tumors and ipsilateral breast cancers. Kreike et al. (54) performed a gene expression analysis of 18000 CDNAs in nine pairs of primary breast cancer with their ipsilateral breast recurrences among women who were younger than 51 years at the time of their initial breast-conserving therapy. Paired data analysis showed no set of genes that had consistently different levels of expression in primary tumors and local recurrences. Another route that has still scarcely been explored is the search for a biologic signature to predict the risk of local recurrence, especially after breast-conserving treatment (54–56). A better distinction between new primary tumors and true recurrences is needed to perform a supervised study based on the occurrence of true recurrences only and not of all ipsilateral breast cancers.

However, our scoring method, which is based on the DNA breakpoint partial identity, has two shortcomings. First, it suffers from the need to conserve unaltered, freshly frozen tissue samples of both the primary tumor and the ipsilateral breast recurrence. This problem should, however, be resolved in time with the possibility of performing the same genomic studies on formalin-fixed paraffin-embedded tissue samples (57–61) or when cryoconservation of either biopsies or fine-needle aspirations (because only 250 ng of DNA is needed, ie, less than 50 000 cells) become standard practice and will make it possible to perform SNP arrays on many more patients. Second, it requires selecting tumors with a cancer cellularity of more than 50%, discarding in the process a number of potentially analyzable tumors. This loss should be diminished in time with both a better selection of frozen tissue material due to the increased experience of the pathologist and the possibility of performing laser capture microdissection.

References


Funding
Institut Curie, the “Courir pour la vie, Courir pour Curie” association, the “Odyssée” association and the PHRC 2006 (AOM 06 149).

Downloaded from https://academic.oup.com/jnci/article-abstract/100/1/48/2567698 by guest on 24 November 2018
Notes

M. A. Bollet and N. Servant contributed equally to this work. The authors thank the members of the departments of Tumor Biology (Marital Caly, Blandine Massemia, Michèle Galut), Biostatistics (Eléonore Gravier, Chantal Gautier), Translational Research (David Gentien, Cécile Reyes, Audrey Rapinat, Benoît Albaud, Vincent Lepetit), and Bioinformatics (Philippe La Rosa, Séverine Lair) who participated in this study. The authors are also indebted to Anne Vincent-Salomon, Patricia de Crémoux, Dominique Stoppa-Lyonnet, and particularly Olivier Delattre for their very valuable comments on this work. Finally, they thank all the members of the Institut Curie Breast Cancer Group.

The sponsors had no role in the study design, data collection, interpretation of the results, preparation of the manuscript, or the decision to submit the manuscript for publication.

Manuscript received June 4, 2007; revised October 16, 2007; accepted November 13, 2007.