Comparative genomic analysis of primary tumors and metastases in breast cancer

François Bertucci¹,²,³, Pascal Finetti¹,*, Arnaud Guille¹,*, José Adélaïde¹,*, Séverine Garnier¹,*, Nadine Carbuccia¹, Audrey Monneur¹,², Emmanuelle Charafe-Jauffret¹,³,⁴, Anthony Goncalves²,³, Patrice Viens²,³, Daniel Birnbaum¹, Max Chaffanet¹

¹Département d’Oncologie Moléculaire, Centre de Recherche en Cancérologie de Marseille (CRCM), Institut Paoli-Calmettes, INSERM UMR1068, CNRS UMR725, Marseille, France
²Département d’Oncologie Médicale, CRCM, Institut Paoli-Calmettes, Marseille, France
³Faculté de Médecine, Aix-Marseille Université, Marseille, France
⁴Département de Biopathologie, CRCM, Institut Paoli-Calmettes, Marseille, France
*These authors contributed equally to this work

Correspondence to: François Bertucci, e-mail: bertuccif@ipc.unicancer.fr

Keywords: array-CGH, breast cancer, genomics, metastasis, sequencing

Received: January 08, 2016    Accepted: March 14, 2016    Published: March 25, 2016

ABSTRACT

Personalized medicine uses genomic information for selecting therapy in patients with metastatic cancer. An issue is the optimal tissue source (primary tumor or metastasis) for testing. We compared the DNA copy number and mutational profiles of primary breast cancers and paired metastases from 23 patients using whole-genome array-comparative genomic hybridization and next-generation sequencing of 365 “cancer-associated” genes. Primary tumors and metastases harbored copy number alterations (CNAs) and mutations common in breast cancer and showed concordant profiles. The global concordance regarding CNAs was shown by clustering and correlation matrix, which showed that each metastasis correlated more strongly with its paired tumor than with other samples. Genes with recurrent amplifications in breast cancer showed 100% (ERBB2, FGFR1), 96% (CCND1), and 88% (MYC) concordance for the amplified/non-amplified status. Among all samples, 499 mutations were identified, including 39 recurrent (AKT1, ERBB2, PIK3CA, TP53) and 460 non-recurrent variants. The tumors/metastases concordance of variants was 75%, higher for recurrent (92%) than for non-recurrent (73%) variants. Further mutational discordance came from very different variant allele frequencies for some variants. We showed that the chosen targeted therapy in two clinical trials of personalized medicine would be concordant in all but one patient (96%) when based on the molecular profiling of tumor and paired metastasis. Our results suggest that the genotyping of primary tumor may be acceptable to guide systemic treatment if the metastatic sample is not obtainable. However, given the rare but potentially relevant divergences for some actionable driver genes, the profiling of metastatic sample is recommended.

INTRODUCTION

During the last decades, the progression-free survival of patients with metastatic breast cancer has improved [1] thanks to the introduction of targeted therapies such as ERBB2 inhibitors, bevacizumab [2], and mTOR inhibitors [3]. Other promising drugs such as PARP inhibitors [4] or CDK4 inhibitors [5] are in development. However, the benefit is limited to patients’ subgroups that remain to be defined. In this context, the identification of biomarkers of response is fundamental to better tailor these expensive and potentially toxic treatments.

Current personalized medicine is based on the accurate identification of actionable molecular alterations present at the time of relapse. One key issue is whether these must be searched in the metastasis or may be...
identified in the primary tumor removed some years earlier, but generally available. Metastatic progression is a complex multistep phenomenon characterized by the accumulation of multiple molecular alterations, notably genetic, within cell clones of a primary tumor, oligoclonal and genetically instable [6]. It has thus been suggested that the genomic profile of metastases should be discordant from that of the primary tumor, with the presence of mutations conferring selective growth or invasive advantage to the metastatic cells. New generation sequencing has revealed the genetic heterogeneity between primary and metastases, between metastases, but also within different regions of the same tumor. The analysis of rare subclones within tumors has evidenced the concept of branched evolution where both divergence and independent convergence may happen synchronously in multiple subclonal populations [7]. Such heterogeneity partly explains the emergence of therapeutic resistance, a predominant cause of cancer-related death, and the discrepancies between primary tumor and metastases reported for some clinical biomarkers [8]. In breast cancer, the tumor-metastasis discrepancy rates are 9–18% for estrogen receptor (ER), 24–31% for progesterone receptor (PR), and ~10% for ERBB2 [9]. That led ASCO to recommend, in patients with accessible metastases, the biopsy for diagnostic confirmation and retesting of ER, PR, and ERBB2 status [9].

With the increasing development of targeted therapies and the advent of personalized medicine, the genotyping of metastatic samples, mainly based on array-based comparative genomic hybridization (aCGH) and next-generation sequencing (NGS), is being used in research as molecular screening before enrollment in clinical trials [10, 11], and might enter the routine clinical practice in the coming years. However, obtaining good quality biopsies of metastatic lesions is often challenging: core biopsy specimens tend to be small in size that will not always allow molecular analyses aside from immunohistochemistry (IHC); they tend to be relatively impure because of stromal contamination; rebiopsies are not always possible according to metastatic sites and may be associated with morbidity. Whether the archival specimen of the operated primary tumor accurately contains already the critical genomic alterations present on metastasis is thus a clinically relevant issue, which has been addressed in certain cancers such as colon cancer with discordant results [12, 13].

Here, we have compared the genomic profiles of primary tumors and matched metastasis from 23 patients with breast cancer by using whole-genome high-resolution aCGH and targeted NGS of 365 “cancer-associated” genes. Our aim was to determine the concordance degree of molecular alterations between tumors and their paired metastases, and the possibility of targeted therapy based on the respective genomic profiles.

**RESULTS**

**Patients’ characteristics**

Twenty-three women with breast cancer were included in the study. Clinico-pathological characteristics are summarized in Table 1. Median age at time of breast cancer diagnosis was 41 years (range, 33 to 72). All cases were carcinomas, including different pathological subtypes. As expected, primary tumors showed poor-prognosis features with a high percentage of grade 2–3 (84%), ER-negative (32%), and ERBB2-positive (33%) cases. One tumor/metastasis pair corresponded to synchronous metastasis, and other pairs corresponded to metachronous metastasis. Median delay between the diagnosis of primary cancer and the metastasis profiled was 36 months (range, 0 to 149). The metastasis corresponded either to the first metastatic relapse (N = 16) or to the 2nd to 6th metastatic progression (N = 7). None of the primary tumors had been exposed to systemic therapy before removal. Regarding the metastatic samples, two patients had not received any systemic therapy before removal or biopsy, whereas 21 had received systemic therapy, mainly chemotherapy and hormone therapy, but also chemotherapy and targeted therapies (anti-ERBB2, and a PI3K inhibitor, BKM120). The main metastatic sites profiled were liver, skin and lymph nodes.

**Copy number profiles**

We first compared the aCGH genomic profiles of the 23 primary tumors and 23 metastases. Figure 1A (left panel) shows the frequency plots of 23 primary tumors: as expected [14–16], the most frequently gained regions were on 1q, 8q, 11q, 17q and 20q chromosomal arms, whereas the regions frequently lost were on 8p, 11q and 16q. Globally, metastases and primary tumors showed similar altered regions with similar frequencies of alterations, and no region showed a different alteration frequency (Figure 1A). GISTIC analysis confirmed that most of altered regions were similar between primaries and metastases, but a few regions were different such as the ATM-including region, which was deleted in metastases (Supplementary Figure 1). The median percentage of probes displaying a CNA per sample was not different between primary tumors (5.18%, range 0.92–33.7%) and metastases (5.83%, range 0.06–23.7%; p = 0.846, paired Mann-Whitney test), even if a great variability existed for both types of samples (Supplementary Table 1). As shown by the correlation matrix generated with all probes (Figure 1B), each metastatic sample correlated more strongly with its paired primary tumor than with other samples. Hierarchical clustering of whole DNA copy number data showed that most of paired primary and metastatic samples (22 pairs out of 23) clustered together (Figure 1C),
Table 1: Clinicopathological characteristics of the 23 patients

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Age at diagnosis</th>
<th>Pathological type</th>
<th>Pathological grade</th>
<th>ER status*</th>
<th>ERBB2 status*</th>
<th>Systemic treatment between primary tumor and profiled metastasis</th>
<th>Metastatic location</th>
<th>Delay between primary tumor and metastasis (months)</th>
<th>N° of metastatic relapse or progression</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63</td>
<td>Mixed</td>
<td>2</td>
<td>Positive</td>
<td>Negative</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Lymph nodes</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>52</td>
<td>Ductal</td>
<td>2</td>
<td>Positive</td>
<td>Negative</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Skin</td>
<td>110</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>33</td>
<td>Ductal</td>
<td>3</td>
<td>Positive</td>
<td>Negative</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Liver</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>56</td>
<td>Ductal</td>
<td>3</td>
<td>Negative</td>
<td>Negative</td>
<td>Yes (chemoT)</td>
<td>Lymph nodes</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>41</td>
<td>Mixed</td>
<td>2</td>
<td>Positive</td>
<td>Positive</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Skin</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>43</td>
<td>Medullary</td>
<td>3</td>
<td>Negative</td>
<td>Negative</td>
<td>Yes (chemoT)</td>
<td>Muscle</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>61</td>
<td>Ductal</td>
<td>1</td>
<td>Positive</td>
<td>Negative</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Liver</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>72</td>
<td>Lobular</td>
<td>2</td>
<td>Positive</td>
<td>Negative</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Uterus</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>41</td>
<td>Ductal</td>
<td>1</td>
<td>Positive</td>
<td>Negative</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Ovary</td>
<td>32</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>34</td>
<td>Ductal</td>
<td>1</td>
<td>Positive</td>
<td>Negative</td>
<td>Yes (chemoT)</td>
<td>Ovary</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>50</td>
<td>Ductal</td>
<td>1</td>
<td>Positive</td>
<td>Negative</td>
<td>No</td>
<td>Lymph nodes</td>
<td>81</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>51</td>
<td>Ductal</td>
<td>3</td>
<td>Negative</td>
<td>Positive</td>
<td>Yes (hormoneT)</td>
<td>Lung</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>34</td>
<td>Ductal</td>
<td>2</td>
<td>Negative</td>
<td>Positive</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Bladder</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>35</td>
<td>Ductal</td>
<td>1</td>
<td>Positive</td>
<td>Negative</td>
<td>No</td>
<td>Lymph nodes</td>
<td>88</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>38</td>
<td>Ductal</td>
<td>3</td>
<td>Positive</td>
<td>Positive</td>
<td>Yes (chemoT)</td>
<td>Skin</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>70</td>
<td>Lobular</td>
<td>1</td>
<td>Negative</td>
<td>Positive</td>
<td>Yes (chemoT)</td>
<td>Skin</td>
<td>42</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>62</td>
<td>Ductal</td>
<td>3</td>
<td>Negative</td>
<td>Positive</td>
<td>Yes (chemoT, trastuzumab, lapatinib, BKM120, T-DM1)</td>
<td>Skin</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>60</td>
<td>Ductal</td>
<td>2</td>
<td>Positive</td>
<td>Positive</td>
<td>Yes (chemoT, hormoneT, trastuzumab, lapatinib, T-DM1)</td>
<td>Peritoneum</td>
<td>149</td>
<td>4</td>
</tr>
<tr>
<td>21</td>
<td>37</td>
<td>Ductal</td>
<td>Positive</td>
<td>Negative</td>
<td>Negative</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Lymph nodes</td>
<td>73</td>
<td>6</td>
</tr>
<tr>
<td>22</td>
<td>33</td>
<td>Ductal</td>
<td>3</td>
<td>Positive</td>
<td>Negative</td>
<td>Yes (chemoT, hormoneT, trastuzumab)</td>
<td>Liver</td>
<td>52</td>
<td>4</td>
</tr>
<tr>
<td>23</td>
<td>38</td>
<td>Metaplastic</td>
<td>3</td>
<td>Negative</td>
<td>Negative</td>
<td>Yes (chemoT)</td>
<td>Liver</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>34</td>
<td>Ductal</td>
<td>Positive</td>
<td>Negative</td>
<td>Positive</td>
<td>Yes (chemoT, hormoneT)</td>
<td>Liver</td>
<td>63</td>
<td>4</td>
</tr>
<tr>
<td>26</td>
<td>38</td>
<td>Ductal</td>
<td>3</td>
<td>Positive</td>
<td>Positive</td>
<td>Yes (chemoT, hormoneT, trastuzumab)</td>
<td>Ovary</td>
<td>78</td>
<td>1</td>
</tr>
</tbody>
</table>

*IHC status: ER (10% positivity cut-off) and ERBB2 (0–3 + score, DAKO HercepTest, with > 1 + defined as positive).
suggesting genetic similarity and potential clonal relationship. For only one pair (patient N°9), samples were distantly related, suggesting distinct genetic relationship.

We then focused the comparison of CNAs on known driver oncogenes located within regions frequently amplified in breast cancer: ERBB2 (17q12), CCND1 (11q13.3), FGFR1 (8p11.23), MYC (8q24), and PAK1 (11q14.1). As expected, ERBB2 was in our series the most frequently amplified gene, and showed 100% concordance between the aCGH status and the IHC status for the 41 informative samples. The concordance rate of amplified/non-amplified status between primary tumors and paired metastases was 100% for ERBB2 and FGFR1, 96% for CCND1 and PAK1, and 88% for MYC (Figure 2), suggesting possible differences regarding some driver genes.

**Mutational profiles**

Among the 365 sequenced genes, 499 mutations, including 414 SNVs (non-synonymous, stop/gain) and 85 indels, were retained as putative somatic alterations within the 46 samples. They corresponded to 298 different mutations (see Supplementary Table 2 for the details of alterations). All samples exhibited at least one mutation. As expected for breast cancers [17], mutational profiles of primary tumors included AKTI, CDH1, ERBB2, GATA3, MLL3/KMT2C, PIK3CA, and TP53 mutations (Figure 3). A total of 247 mutations (205 SNVs, 42 indels) were identified among the primary tumors, and 252 (209 SNVs, 43 indels) among the metastases. The median number of mutations per sample was similar between primary
tumors (9, range 4 to 24) and metastases (9, range 4 to 29; p = 0.709, paired Mann-Whitney test). We measured the similarity between each metastasis and all primary tumors by measuring the correlation of variant allele frequencies (VAF) of all detected variants: each metastasis correlated more strongly with its paired primary tumor than with other samples, suggesting strong similarity (Figure 4). However, the VAF of some variants showed strong differences between primary tumor and metastasis in some cases, such as PIK3CA in the pairs N°9 and 24, and TP53 in the pair N°23 (Supplementary Table 2). This global similarity was also observed in the concordance analysis, as shown by the correlation matrix in Supplementary Figure 2. The global rate of concordance for the detected variants between primary tumor and paired metastases was 75%, with 374 shared variants and 125 unshared variants (Table 2).

The 499 variants included 39 recurrent variants (8%; 14 unique) and 460 non-recurrent variants (92%; 284 unique). The 39 recurrent variants (Table 3) concerned four driver genes of breast cancer: AKT1, ERBB2, PIK3CA, and TP53. They were equally distributed with 19 variants in primary tumors and 20 in metastases. Eighteen out of 23 primary tumors harbored at least one of these recurrent variants (two variants in one case) and 17 out of 23 metastases harbored at least one recurrent variant (two variants in three cases) (Supplementary Table 2). The concordance rate (Table 2) was higher for the recurrent variants (92%; 36/39) than for the non-recurrent variants (73%; 338/460; p = 0.076, Fisher’s exact test, Odds Ratio = 3.1 [CI95 0.93–16.4]). PIK3CA variants were the most frequently detected recurrent mutations (23 samples from 12 patients) with one discordant mutation observed in patient N°8, who harbored the E545K variant in the primary tumor but not in the metastasis. TP53 variants were observed in 11 samples from 6 patients with one discordant mutation observed in patient N°5, who harbored the R273C variant in the metastasis but not in the primary tumor. AKT1 variants were observed in 2 patients with no discordant mutation. The only ERBB2 variant (L755S) was observed in patient N°20 in the metastasis but not in the primary tumor. Before the biopsy of the profiled metastasis (4th metastatic progression), this patient had received several lines of anti-ERBB2 drugs (trastuzumab, lapatinib, T-DM1). The L755S ERBB2 variant induces resistance to lapatinib via restriction of

---

**Figure 2:** Genomic profiles within four regions frequently amplified in breast cancer. The copy number profiles of each region (log2 ratios) were plotted for each of the 46 samples (23 pairs). Different colors correspond to different pairs, and the full line corresponds to the primary tumor and the dashed line to the metastasis. Four regions frequently amplified in breast cancer and one oncogene driver per region are shown: 17q12 and ERBB2 (A) 8q24 and MYC (B) 8p11.23 and FGFR1 (C) and 11q14.1 and PAK1 (D).
kinase conformational flexibility that blocks lapatinib binding [18]. It is likely that the mutation present in this metastasis had been induced by the preceding treatment. Thus, 20 out of 23 tumor/metastasis pairs (87%) were concordant with respect to the detected recurrent variants, whereas 13% were discordant.

**Choice of targeted therapy**

Because personalized medicine theoretically relies on the molecular profile of the relapse, we compared for each patient the therapeutic choices guided by the profiling of primary tumor versus metastasis using two examples of clinical trials (Supplementary Table 3). Here, the actionable molecular alterations retained for analysis concerned genes coding for molecular targets of drugs proposed in the trials or involved in pathways targeted by these drugs. Furthermore, they had to be activating for oncogenes and biallelic inactivating for tumor suppressor genes such as PTEN. First, we focused on the seven drugs proposed in the SHIVA trial [11]: imatinib, everolimus, vemurafenib, sorafenib, erlotinib, dasatinib, and lapatinib combined with trastuzumab. Based on the profiling of metastasis, 15 patients (65%)...
would have been candidate to at least one targeted therapy (everolimus: 13; lapatinib plus trastuzumab: 2), whereas 16 (70%) would have been candidate (everolimus: 14; lapatinib plus trastuzumab: 2) according to the profiling of primary tumor. The concordance degree with respect to therapeutic selection was 96% (22 out of 23). Then, we focused on the five drugs proposed in the ongoing MOST trial (NCT02029001): everolimus, nilotinib, sorafenib, pazopanib, and lapatinib. The results were similar, with a concordance rate of therapeutic selection of 96%, with 17 patients (74%) candidate to at least one targeted therapy (everolimus: 14, lapatinib: 2, nilotinib: 1) according to the molecular profile of the metastasis, versus 18 (78%; 15 for everolimus, 2 for lapatinib, 1 for nilotinib) according to the profile of primary tumor. Thus, in both trials, the metastasis profiling did not reveal additional actionable therapeutic target as compared with the profiling of primary tumor.

## DISCUSSION

Our objective was to assess the concordance of high-throughput CNA and mutational profiles between primary breast cancers and paired metastases. We showed in our series of 23 patients that overall those profiles are concordant and the resulting selection of targeted therapy would be the same in all but one patient. However, some differences that might be extremely relevant were identified for some genes recurrently altered in breast cancer.

Our analysis was based on whole-genome high-resolution aCGH and targeted NGS of 365 genes chosen for their relevance in oncology by biologists and medical oncologists of our institution. We are currently using these technologies in the PERMED trial (NCT02342158) that aims to establish the genomic profile of advanced tumors as pre-therapeutic screening tool. All samples; primary tumors and metastases, harbored many CNAs.

![Figure 4: Correlation between each metastatic sample and all primary tumors with respect to mutational profiles.](image)

Correlation matrix based on the variant allele frequency (VAF) for all detected variants generated between all primary tumors and all metastases: the Pearson coefficient is color-coded according to the scale shown below the matrix.
and mutations commonly found in breast cancer [17]. The profiled metastatic samples corresponded to different metastatic locations and timings. Different combinations of chemotherapy, hormone therapy and anti-ERBB2 therapies had been delivered between the removals of the primary tumor and the paired metastasis. In term of genes tested, genomic data of paired samples may be compared in several ways. Our analysis was both global, integrating all genes together, but also individual, concerning genes with known oncogenic and recurrent alterations in breast cancer and for which targeted therapies are available.

Despite the above-quoted divergences and the known genetic instability of cancer cells, we found a very high level of global concordance between primary and secondary tumors. The concordance with respect to whole CNAs was first suggested by similar frequency plots and numbers of alterations per sample, then more importantly by clustering and correlation matrix which showed that each metastasis correlated more strongly with its paired primary tumor than with other samples. At the gene levels, genes with recurrent amplifications in breast cancer showed different degrees of concordance: 100% for \(ERBB2\) and \(FGFR1\), 96% for \(CCND1\), but 88% for \(MYC\), suggesting possible differences for driver genes. The assessment of global similarity of mutational profiles was based on different criteria: number of mutations per sample, correlation matrix, and concordance analysis of detected variants. This later was 75% when we considered all variants, but higher for recurrent variants (92%), which concern driver genes involved in disease progression, than for non-recurrent variants (73%), which are generally random/passenger and the consequence of genomic instability. Of course, our findings need to be interpreted in the context of clinically relevant clonality: here we chose a 2% cut-off, but the optimal clonal frequency is currently unknown with different values used ranging between 2 and 10%. Reanalysis of our data using 5% and 10% cut-offs decreased the degree of concordance with respective rates of 70 and 60% for all variants, 82 and 77% for recurrent variants, and 69 and 59% for non-recurrent variants (data not shown). The three recurrent variants for which we found rare cases of divergent mutation (\(TP53\), \(PIK3CA\), and \(ERBB2\)) are part of a list of 16 genes already reported with frequent clonal divergence in a list of 46 genes tested [19]. Their discordance between primary tumor and metastasis may have several explanations: false-negativity due to low cellularity (but cellularity in our cohort was not different between the paired samples) or intra-sample heterogeneity, or true negativity: the two cases with novel mutation in the metastasis (\(TP53\), \(ERBB2\)) may indicate accumulation of mutations over time associated with the therapeutic resistance (due to previous lapatinib treatment for \(ERBB2\)), whereas the case with novel mutation in the primary tumor (\(PIK3CA\)) may suggest that the metastasis branched off before the acquisition of this mutation within the primary tumor. Further degree of mutational discordance came from very different VAF for some variants between primaries and metastases, differences that might have important clinical implications, notably for treatment.

Some comparative studies based on high-throughput molecular analyses have been published and reported similar high degree of global concordance.

### Table 3: List of 39 recurrent somatic variants detected in the 46 samples

<table>
<thead>
<tr>
<th>Gene</th>
<th>cDNA mutation</th>
<th>Impact on protein synthesis</th>
<th>Primary tumors occurrence (N)</th>
<th>Metastases occurrence (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIK3CA</td>
<td>C1616G</td>
<td>P539R</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>G1624A</td>
<td>E542K</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>G1633A</td>
<td>E545K</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A3140G</td>
<td>H1047R</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>A3140T</td>
<td>H1047L</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>G353A</td>
<td>G118D</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TP53</td>
<td>C152G</td>
<td>S51X*</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>G128A</td>
<td>R43H</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>G317A</td>
<td>C106Y</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>T304C</td>
<td>Y102H</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>C421T</td>
<td>R141C</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>G422T</td>
<td>R141L</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AKT1</td>
<td>G49A</td>
<td>E17K</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>ERBB2</td>
<td>T2264C</td>
<td>L755S</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td><strong>19</strong></td>
<td><strong>20</strong></td>
</tr>
</tbody>
</table>

*Stop codon
In breast cancer, at least five studies are available. In a series of 22 matched primary-recurrences sequenced using targeted NGS (196-gene panel), the concordance rate was 85% for the known driver gene mutations [20]. Similarly, high-resolution aCGH profiling of 20 pairs showed strong concordance between the primary tumors and paired lymph node metastases, suggesting high clonal relationship [21]. In a series including 79 primary/metastasis pairs sequenced using targeted NGS (46-gene panel), the concordance rate for all detected variants was 84% [19]. Very good concordance (100% for CNAs, 80% for mutations) was observed for the two paired primary-metastases profiled using whole-genome aCGH and whole-exome, whereas the four bilateral breast cancers pairs showed discordant profiles [22]. Finally, no significant difference was observed between primary tumor and brain metastasis in a series of 15 pairs profiled using targeted NGS of 50 genes [23]. Strong concordance was also observed between the transcriptional profiles of primary tumors and paired metastases [24, 25]. High mutational concordance (94%) between tumors and metastases profiled using targeted NGS has been reported for known recurrent genomic alterations in a series of 15 pairs of non-small cell lung cancers [26]. In colon cancer profiled using targeted NGS, the concordance rates were 78% (90% for recurrent variants), 79%, and 85% in respective series of 13 pairs [27], 69 pairs [28], and 86 pairs [19]. By contrast, discordance in CNA of potential clinical relevance was reported in urothelial carcinoma [29].

Finally, using two examples of clinical trials of personalized medicine, we showed that the chosen targeted therapy would have been strongly concordant (96%) between two choice based on the profiling of primary tumor and paired metastasis respectively. Of note, the proportion of patients candidate to one of the proposed targeted therapies was very similar between the two trials, ranging from 65 to 74%, as previously reported [10, 11, 20, 30]. Of course, our result remains dependent on the list of drugs tested. For example, the NOTCH4 mutation observed in the metastasis from patient N°9 but not her primary tumor might impact in a near future the therapeutic decision, but no NOTCH4 inhibitor was included in the list of drugs tested here.

To our knowledge, this study is the first one, which combines both whole-genome aCGH and targeted NGS of a panel of more than 300 genes to breast cancer pairs, and which compares the therapeutic selection based on the profiles of primary tumors and metastases. Of course, it displays some limitations: i) the small number of cases - even if it is the second largest study after that recently reported by the MD Anderson team [19]-, that should enlarge in the future thanks to the recently launched PERMED trial, provided that the frozen primary tumor is available for profiling; ii) the heterogeneity of the population in terms of molecular subtypes, locations and timings of metastases, even if the concordance does not seem different according to these parameters, but the number of pairs precluded any statistical analysis; iii) the delivery of different systemic treatments before the metastatic progression; iv) the relatively small number of genes analyzed by NGS, even if more comprehensive sequencing (whole exome, whole genome) did not identify any additional metastasis-specific actionable alterations in small recent series of breast [22] and colon [28] cancers, when compared to targeted NGS; v) the relatively small number of drugs available in the two tested clinical trials of personalized medicine, when compared to the much higher number of therapeutic targets tested. Whether the concordance rate in the therapeutic decision would be as high as 96% with drugs targeting all screened genes remains unknown, even if we showed strong concordance of recurrent alterations between primaries and metastases. Finally, the comparison between primaries and metastases should not be limited to CNA and mutational profiles, but could include notably proteomics and phosphoproteomics analyses, as well as preclinical models comparing the predictive effect for drug sensitivity of molecular alterations found in primaries and metastases. But today, aCGH and NGS represent the backbone of personalized medicine.

In conclusion, we have evidenced a high level of global concordance, but also a small but actual degree of quali- and/or quantitative divergence for some actionable driver genes. This is an important information for future studies of personalized medicine in metastatic patients. Indeed, because the addition or loss of one single mutation may be extremely relevant by profoundly affecting the signal transduction machinery, such studies should be designed on the basis of genomic profiling of contemporary sample; this is in agreement with the current ASCO guidelines, which recommend the biopsy for retesting ER, PR, and ERBB2 in patients with accessible metastases [9]. However, since the genotyping of the primary tumor seems sufficient to guide systemic treatment in the vast majority of cases, it is acceptable in cases where metastatic location, patient or doctor preference, comorbidity or cost make procurement of a more contemporary specimen untenable. Finally, the profiling of metastatic samples will be crucial not only to help understand the metastatic process and the resistance mechanisms by identifying the molecular alterations found in concordant versus divergent primary and metastatic tumor pairs [7], but also to collect precious data for future research.

MATERIALS AND METHODS

Breast cancer samples

Potential patients were retrospectively searched in our institutional breast cancer database. Inclusion criteria were: women, invasive breast carcinoma treated...
at the Institut Paoli-Calmettes, available frozen samples of both primary tumor and paired metastasis, tumor cellularity of at least 50% as assessed by one pathologist (ECJ) on tumor sections before DNA extraction, available clinicopathological data, and written informed patient’s consent. The study was approved by our institutional “Comité d’Orientation Stratégique” (N°13-002). Forty-six tissue samples from 23 patients were identified. Samples had been collected by surgery or imaging-guided biopsies and macrodissected and frozen in liquid nitrogen within 30 minutes of removal. Tumor DNA was extracted as previously described [31]. Quality was controlled on polyacrylamide gel electrophoresis, and concentration assessed by using Qubit dsDNA BR Assay.

Array-comparative genomic hybridization

DNA copy number alterations (CNA) were determined by using high-resolution CGH microarrays (SurePrint G3 Human 4 × 180, Agilent, France) as previously described [31]. Tumor DNA was cohybridized with a pool of 13 normal male DNA as reference. Scanning was done with Agilent Autofocus Dynamic Scanner (G2565BA, Agilent). Data analysis and visualization were done with CGH Analytics 3.4 software (Agilent). Data extraction (log₂ ratio) was done from CGH analytics, while normalized and filtered log ratio were obtained from “Feature extraction” software (Agilent). We eliminated data generated by probes mapped to X and Y chromosomes. The final dataset included 161,068 unique probes covering 16,684 genes and intergenic regions according to the hg19/NCBI human genome mapping database (build 37).

Data were analyzed using circular binary segmentation as implemented in the DNA copy R/Bioconductor package [32] with default parameters to translate intensity measurements in regions of equal copy number, each region being defined by at least five consecutive probes. Thus, each probe was assigned a segment value referred to as its “smoothed” value. We used two different threshold values (log₂ ratio > 0.5, and |1|) to distinguish low (gain/loss) from high (amplification/deletion) level CNAs respectively [31]. To identify altered regions, we used the GISTIC 2.0 (v2.0.21) algorithm [33], which computes for each segment through the genome a score based on the frequency of CNA combined with its amplitude, with bootstrapping to calculate the significance level (q < 0.25).

Next-generation sequencing

Targeted NGS was applied to a custom-made panel of 365 “cancer-associated” genes selected for their involvement in cancers (CCH-V6 panel, Supplementary Table 4). For each sample, we prepared the DNA libraries of all coding exons and intron-exon boundaries of all genes using the HaloPlex Target Enrichment System (Agilent, Santa Clara, CA, USA) as described [34]. Sequencing was done using the 2 × 150-bp paired-end technology on the Illumina MiSeq platform according to the manufacturer’s instructions (Illumina, San Diego, CA, USA).

Sequence data were aligned to the human reference genome (UCSC hg19) using Burrows-Wheeler Aligner [35]. Samples were sequenced at an average depth of 300× for the targeted regions. Bam files were processed as described [34]. Then, the single nucleotide variants (SNVs) calling was done with FreeBayes version 0.9.9 [36] with a minimal alternate variant frequency and coverage set at 0.02 and 10. Insertions/deletions (indels) calling was done using GATK haplotype caller version 2.5.2-gf57256b [37] with default parameters. The variants, i.e SNVs and indels, were annotated with the Annotate Variation Software (ANNOVAR, version 2013-11-12). Known variants found in dbsnp129 and dbsnp137 with a variant allele frequency (VAF) superior to 1% (1000 g or ESP6500) were removed. Finally, low frequency SNVs and indels that were suspected to be false positive were systematically inspected with IGV version 2.3.32 [38, 39].

Statistical analysis

The frequency of CNAs, computed for each probe locus, was compared between tumors and metastases using the Fisher’s exact test. The percentage of probes displaying a CNA per sample was calculated as the total number of probes with CNAs divided by the total number of probes. We analyzed the correlation (Pearson coefficient) of CNA profiles (log₂ ratio of all probes) of each metastasis with all primary tumors. Hierarchical clustering of whole-genome copy number data was also applied to assess the global genetic similarity between the primaries and metastases: we used the R-package pvclust [40] with the following parameters: Ward’s agglomerative method, Pearson correlation and 100 bootstrap replications to assess the robustness of clusters.

Regarding the variants, the similarity of samples was measured using the Pearson correlation based on the VAF for all detected variants of each metastasis with all primary tumors. Concordance analysis was done as described [22]. Recurrent variants were defined as alterations present in 10 or more samples in COSMIC V68; other variants were defined as non-recurrent. Correlations between sample groups and variables were calculated with the Fisher’s exact test (qualitative variables), and the Mann-Whitney test (continuous variables). We also compared for each patient the selection of targeted therapies guided by the profiling of primary tumor versus metastasis by focusing our analysis on genes coding for molecular targets of drugs proposed in the SHIVA [11] and MOST (NCT02092001) trials. The choice of therapy was based on the guidelines proposed in each respective protocol. All statistical tests were two-sided at the 5% level of significance. Analyses were done in the R software (version 2.15.2).
ACKNOWLEDGMENTS AND FUNDING

Our work is supported by the Association La Marie-Do (FB), Institut National du Cancer (INCa) Translational Research Grant 2009 (DB) and Biologic Research Grant 2010 «IVOIRES » (FB), the Ligue Nationale Contre le Cancer (label DB), SIRIC (INCa-DGOS-Inserm 6038 grant). and Institut Paoli-Calmettes.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES


ACKNOWLEDGMENTS AND FUNDING

Our work is supported by the Association La Marie-Do (FB), Institut National du Cancer (INCa) Translational Research Grant 2009 (DB) and Biologic Research Grant 2010 «IVOIRES » (FB), the Ligue Nationale Contre le Cancer (label DB), SIRIC (INCa-DGOS-Inserm 6038 grant). and Institut Paoli-Calmettes.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES


