

Research article

Open Access

Genetic variability in MCF-7 sublines: evidence of rapid genomic and RNA expression profile modifications

Mélanie Nugoli¹, Paul Chuchana¹, Julie Vendrell¹, Béatrice Orsetti¹,
Lisa Ursule¹, Catherine Nguyen², Daniel Birnbaum³,
Emmanuel JP Douzery⁴, Pascale Cohen^{1,5} and Charles Theillet*¹

Address: ¹Equipe Génome et Cancer, UMR 5535 CNRS and EMI 0229 INSERM Centre de Recherche CRLC Val d'Aurelle, Montpellier, France, ²Laboratoire TAGC, CIML, Université d'Aix-Marseille II, Marseille, France, ³INSERM U119 and LBT, Institut Paoli Calmette, 232 blv Ste Marguerite, 13009 Marseille, France, ⁴Institut des Sciences de l'Evolution de Montpellier CNRS UMR 5554, Université des Sciences et Techniques du Languedoc Montpellier II, Montpellier, France and ⁵Institut de Biotechnologies et Pharmacologie CNRS UMR 5094, Faculté de Pharmacie Université Montpellier I, Montpellier, France

Email: Mélanie Nugoli - nugoli@valdorel.fncclcc.fr; Paul Chuchana - chuchana@valdorel.fncclcc.fr; Julie Vendrell - vendrell@valdorel.fncclcc.fr; Béatrice Orsetti - orsetti@valdorel.fncclcc.fr; Lisa Ursule - ursule@valdorel.fncclcc.fr; Catherine Nguyen - nguyen@ciml.univ-mrs.fr; Daniel Birnbaum - birnbaum@marseille.inserm.fr; Emmanuel JP Douzery - douzery@isem.isem.univ-montp2.fr; Pascale Cohen - pascale.cohen@ibph.pharma.univ-montp1.fr; Charles Theillet* - theillet@valdorel.fncclcc.fr

* Corresponding author

Published: 24 April 2003

Received: 6 December 2002

BMC Cancer 2003, 3:13

Accepted: 24 April 2003

This article is available from: <http://www.biomedcentral.com/1471-2407/3/13>

© 2003 Nugoli et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Both phenotypic and cytogenetic variability have been reported for clones of breast carcinoma cell lines but have not been comprehensively studied. Despite this, cell lines such as MCF-7 cells are extensively used as model systems.

Methods: In this work we documented, using CGH and RNA expression profiles, the genetic variability at the genomic and RNA expression levels of MCF-7 cells of different origins. Eight MCF-7 sublines collected from different sources were studied as well as 3 subclones isolated from one of the sublines by limit dilution.

Results: MCF-7 sublines showed important differences in copy number alteration (CNA) profiles. Overall numbers of events ranged from 28 to 41. Involved chromosomal regions varied greatly from a subline to another. A total of 62 chromosomal regions were affected by either gains or losses in the 11 sublines studied. We performed a phylogenetic analysis of CGH profiles using maximum parsimony in order to reconstruct the putative filiation of the 11 MCF-7 sublines. The phylogenetic tree obtained showed that the MCF-7 clade was characterized by a restricted set of 8 CNAs and that the most divergent subline occupied the position closest to the common ancestor. Expression profiles of 8 MCF-7 sublines were analyzed along with those of 19 unrelated breast cancer cell lines using home made cDNA arrays comprising 720 genes. Hierarchical clustering analysis of the expression data showed that 7/8 MCF-7 sublines were grouped forming a cluster while the remaining subline clustered with unrelated breast cancer cell lines. These data thus showed that MCF-7 sublines differed at both the genomic and phenotypic levels.

Conclusions: The analysis of CGH profiles of the parent subline and its three subclones supported the heteroclonal nature of MCF-7 cells. This strongly suggested that the genetic plasticity of MCF-7 cells was related to their intrinsic capacity to generate clonal heterogeneity. We propose that

MCF-7, and possibly the breast tumor it was derived from, evolved in a node like pattern, rather than according to a linear progression model. Due to their capacity to undergo rapid genetic changes MCF-7 cells could represent an interesting model for genetic evolution of breast tumors.

Background

Primary breast tumors are known for their elevated level of inter-tumor heterogeneity, however, an important body of data has brought evidence of intra-tumoral heterogeneity as well. Such evidence stems from cytogenetic studies which have shown that cytogenetically unrelated clones can be found in breast tumors [1]. These findings have been interpreted either as the result of genetic instability following loss of proper mitotic controls [2], or as the expression of the admixture of multiple genetically non related cellular clones [3]. Flow cytometry has been another way to address the question of intratumoral heterogeneity, showing that breast tumors correspond to intricate admixture of tumor cells with different DNA contents (i.e. different ploidies) [4]. These findings were extended by Bonsing and coworkers [5], who showed that diploid and aneuploid cells, concurrently present in breast tumors, had a number of genetic anomalies in common. In fact, all the allelic imbalances observed in the diploid compartment were found in aneuploid cells. This was a strong indication of a direct filiation between diploid and aneuploid cells in breast tumors. Heterogeneity is thus a major problem in mammary carcinogenesis and has important clinical implications in terms of prognosis and therapy.

MCF-7 cells are the most commonly used model of estrogen positive breast cancer. This cell line has been originally established in 1973 at the Michigan Cancer Foundation from a pleural effusion taken from a woman with metastatic breast cancer [6] and since then MCF-7 cells have been widely distributed in laboratories throughout the world resulting in the production of different cellular stocks. Quite early in the history of MCF-7 cells reports on clonal variations were made in the literature. Most of the reported differences concerned phenotypic traits such as estrogen responsiveness or ability to form tumors in syngeneic mice, but karyotypic differences were observed as well [7–9]. MCF-7 cells presented extensive aneuploidy with important variations in chromosome numbers ranging from 60 to 140 according to the variant examined. Other cytogenetic differences concerned the presence or absence of specific marker chromosomes. While loss of marker chromosomes seemed a rare event, occurrence of new aberrations was more common [8]. However, some doubt remained on the true origin of these differences, as some MCF-7 sublines corresponded to other cancer cells of unknown origin [10].

The available data suggested an elevated level of genetic instability in MCF-7 cells. The observed karyotypic differences could reflect changes in selective pressure due to different culture conditions. Alternatively, work by Resnicoff and coworkers [11] showed that, upon fractionation of MCF-7 cells on a Percoll gradient, it was possible to isolate six different subpopulations, one of which bore the capacity to regenerate all other cellular populations. These data suggested that MCF-7 cells contain a fraction of stem cells able to generate clonal variability. This was proposed as an explanation for the heterogeneity of this cell line and as a model for breast tumor heterogeneity.

In a previous work [12] we analyzed by Comparative Genomic Hybridization (CGH) two sublines of MCF-7 cells which showed surprisingly different genomic profiles. Our data were concordant with that reported by Jones and coworkers [13]. We became interested in: (1) documenting the genetic variability, at both genomic and RNA expression levels, that exists among different MCF-7 sublines of different origins; (2) retracing their evolutionary history and unraveling their filiation; (3) addressing the issue of the cause of this diversity and whether it reflected their intrinsic capacity to generate clonal heterogeneity or resulted from local changes in culture conditions. The resulting information should help understand tumor heterogeneity.

To address these questions we collected 9 cell lines identified as MCF-7 variants. We also established 3 cell clones starting from one of the collected sublines. The different MCF-7 variants were compared at the genetic level using CGH as well as RNA expression profiling. CGH and RNA expression profiles were subjected to phylogenetic analyses to determine the degree of filiation between the different cell lines studied.

Methods

Cell lines

Eleven variants or sublines of MCF-7 cells were tested in this study: MCF-7-ATCC, MCF-7-R, MCF-7-O, MCF-7-MF and MCF-7-MG, MCF-7-MVLN, MCF-7-MVLN-6ms7, MCF-7-MVLN-6ms8, MCF-7-R-F3, MCF-7-R-D4, MCF-7-R-G1. MCF-7-ATCC were obtained from Dr A. Pèlerin (Cancer Center, Montpellier, France) who purchased it from ATCC in 1996. Cells were at passage 143 when we analyzed them. MCF-7-R, MCF-7-O, MCF-7-MF and MCF-7-MG were all obtained from Dr F. Vignon (INSERM, Montpellier, France). MCF-7-R were originally obtained

from Dr Rich (Michigan Cancer Foundation, USA) by Dr F. Vignon in 1985, analyzed cells were at passage 61 (passage 0 at the time of arrival of the cells in Dr Vignon's laboratory). MCF-7-O were obtained from Dr Osborne (Texas Health Science Center, San Antonio) in 1987, analyzed cells were at passage 62. MCF-7-MF were obtained from Dr Lippman (Lombardi Cancer Center, Washington DC, USA) in 1988, analyzed cells were at passage 5. MCF-7-MG were obtained from Dr Mc Guire (Texas Health Science Center, San Antonio) in 1994, analyzed cells were at passage 338. MCF-7-MVLN were obtained from Drs. J.C. Nicolas and M. Pons (INSERM, Montpellier, France). These cells correspond to MCF-7 cells transfected with a construct containing the Luciferase gene under control of an ERE sequence. MCF-7-MVLN-6ms7 and MCF-7-MVLN-6ms8, respectively correspond to two subclones of MCF-7-MVLN cells which have developed resistance to Tamoxifene. MCF-7-R-F3, MCF-7-R-D4 and MCF-7-R-G1 correspond to cell clones established in our laboratory by limit dilution from MCF-7-R cells.

Other breast cancer lines used in this study included Brca-MZ-01 and Brca-MZ-02, MDA-MB-436 (kindly provided by Dr. A. Puisieux), BT-20, CAMA-1, HCC 1187, HCC 1428, HCC 1569, HCC 1937, HCC 1954, HCC 2218, MCF10F, MDA-MB-175, UACC-812 (ATCC, Manassas, Va.), EFM-19, EFM-192A (DSMZ, Braunschweig, Germany), KPL-1, SUM 149, SUM-229 (kindly provided by Dr S. Ethier). The Doxorubicin resistant line was provided to us by Dr Frederic Pinguet (Canter Center Montpellier). All cell lines were maintained in DMEM containing 10% FBS supplemented with L-Glutamine (200 mM, 100X) and Antibiotic-Antimycotic (100X) GibcoBRL, Life Technologies, Cergy Pontoise.

DNA and RNA purification

Genomic DNA and total RNA were isolated as previously described [14]. RNA integrity was controlled by denaturing formaldehyde agarose electrophoresis and checked by Northern blot, hybridizing the RNA with an oligonucleotide probe specific to the 28S rRNA.

Genetic analysis of the different sublines

All the cell lines used in this study have been haplotyped with a combination of 9 CA repeat microsatellite markers from the Généthon collection, respectively localized on chromosomes 1, 6 and 17: D1S2615, D1S2811, D1S2624, D6S310, D6S401, D6S460, D17S1855, D17S1865, D17S1604. Primers are described on the Généthon web site http://www.genlink.wustl.edu/genethon_frame/. PCR conditions and size analysis of the products were as described [15].

Comparative Genomic Hybridization

Metaphase preparation, genomic DNA labeling, CGH reaction and image analysis were as described [16].

Phylogenetic analysis on CGH data

The evolutionary history of cell lineages was reconstructed in a cladistic framework. Chromosomal bands were considered as characters, existing under three possible discrete character states: gain, loss and normal (i.e. no mutation). Transformations from one character state to another were equally weighted, and the normal state (i.e. tumor/normal hybridization ratio = 1) was considered ancestral. Phylogenetic trees were reconstructed under the maximum parsimony (MP) criterion, using the following hypothesis [17]: (1) all characters were considered as independent: *i.e.* events occurring at one band did not affect events occurring at another band; (2) they were unordered: it was possible to directly change from one state (either normal, amplified, or deleted) to a second one, without invoking the third one; (3) they were equally weighted: each change from one state to another had the same probability of occurrence. Using the MP approach has two main advantages: (1) it considered chromosome bands one by one, and integrated CGH information available for each of them for all twelve cell lineages simultaneously; (2) it allowed to trace *a posteriori* chromosomal events that characterize the different groups of cell lineages evidenced on the most parsimonious trees and to identify diagnostic events. All analyses were conducted with PAUP* [18], version 4 beta 8, with heuristic MP searches based on 1000 random addition of cell lineages, with tree bisection-reconnection (TBR) branch swapping, and accelerated transformation (ACCTRAN) optimization of character-states. To trace the character-state changes along the phylogenetic trees, we used the program MacClade [19], version 3.04. In order to evaluate whether CGH data were suitable for reconstructing the phylogeny of the cell lineages, and whether phylogenetic trees adequately represented them, the robustness of the different nodes has been measured, and independently estimated from two different approaches. Bootstrap [20] was conducted with 1000 replicates of character resampling, and the highest bootstrap percentages (BP) defined the strongest nodes. The Bremer approach [21] measured the number of extra-mutation-events required to break the corresponding nodes, and the highest Bremer support indices (BSI) defined the most robust nodes.

Preparation and hybridization of cDNA arrays

Variations in gene expression levels were analyzed by large-scale measurement with home-made cDNA mini-arrays (7.5 × 9 cm; 720 human genes; 11 genes/cm²) produced in our facility (TAGC, University of Marseille Luminy). The spotted targets were PCR products amplified from control clones and IMAGE cDNA clones

(IMAGE consortium, Hinxton, UK). Selected cDNA clones corresponded to identified genes positioned on chromosomes 1q and 17q. Information was gathered and crosschecked from different web based data bases such as genemap <http://www.ncbi.nlm.nih.gov/genemap99/>, Genecards <http://genecards.weizmann.ac.il/>, Genelynx <http://www.genelynx.org/> or UCSC Genome <http://genome.ucsc.edu/>. PCR amplification and automatic spotting of PCR products to the arrays (nylon Hybond-N+ membranes, Amersham Pharmacia Biotech; Little Chalfont, UK) were performed according to Bertucci and colleagues (1999). Each array was hybridized with a ³³P-labeled probe synthesized by reverse transcribing 5 µg of total RNA for each sample [22]. Labeling of complex probes, hybridization and washing conditions were as described <http://tagc.univ-mrs.fr/pub/cancer>. Arrays were exposed to phosphor-imaging plates and then scanned with a FUJI BAS 5000 beta imager (Raytest, Asnieres, France). Hybridization signals were quantified with the HDG Analyzer software (Genomic Solution, Ann Arbor, MI, USA), by integrating all spot pixel intensities and removing a spot background value determined in the neighboring area.

Clustering analysis of gene expression data

Data display and analysis was performed using Excel software (Microsoft, Richmond, WA, USA). Intensity values were adjusted by a normalization step based on the DNA quantification of each spot and the sum of intensities detected in each experiment. Expression profiles were analyzed by hierarchical clustering using the Cluster program developed by Eisen and colleagues [23] and represented as a cladogram using the treview software.

Results

MCF-7 variants

Originally we collected 9 MCF-7 sublines, MCF-7-ATCC, MCF-7-R, MCF-7-O, MCF-7-MF and MCF-7-MG, MCF-7-MVLN, MCF-7-MVLN-6ms7, MCF-7-MVLN-6ms8, as well as a doxorubicin resistant cell line which was believed to be a MCF-7 variant. All these sublines except the three MCF-7-MVLN were of different origins with variable number of passages and culture conditions.

MCF-7-MVLN, MCF-7-MVLN-6ms7 and MCF-7-MVLN-6ms8 resulted from a selection process. MCF-7-MVLN were transfected with an ERE-Luciferase construct [24] and selected for Gentamycin resistance, while both MCF-7-MVLN-6ms7 and MCF-7-MVLN-6ms8 have been produced by a long term exposure of MCF-7-MVLN cells to 200 nM OH-TAM. We also isolated cell clones from MCF-7-R using limit dilution. Three clones MCF-7-R-F3, MCF-7-R-D4 and MCF-7-R-G1 were selected for further studies. This allowed us to verify that MCF-7 cells showed intrapopulation heterogeneity.

Common genetic origin of MCF-7 variants

Available information on the history of the different sublines was not sufficient to rebuild lineages. It was, therefore, important to ascertain that all the tested sublines bore a common genetic origin. To this end allelotypes at 9 polymorphic microsatellite markers located on 3 chromosomal arms were determined. Eight of the 9 sublines had identical haplotypes while the doxorubicin resistant variant presented divergent allelic profiles at all markers analyzed. This was therefore taken out of the study (data not shown).

CGH analysis

Patterns of gains and losses shown by the different MCF-7 variants were highly diverse (Table 1 and Figure 1). Number of events ranged from 28 (MCF-7-ATCC) to 41 (MCF-7-MG) and, on average, losses were more frequent (21) than gains (15). Only 9 events (6 losses, 3 gains) were shared by the 11 cell lines (Figure 1). This small number of common events could in part be attributed to MCF-7-ATCC which presented the most divergent CGH pattern. Out of the 28 gains or losses this subline displayed, 11 (6 losses, 5 gains) were specific to MCF-7-ATCC cells. It was noticeable that the sizes of regions of losses or gains varied according to the subline. This was particularly striking for losses on 16q or gains at 3q or 5q (Figure 1). Generally regions of gains tended to be more heterogeneous in size and occurrence than losses.

These data were strong indications of the elevated level of genetic heterogeneity shown by MCF-7 cells. It was, therefore, interesting to verify how cell lines of known filiation compared to each other. Among all the cell lines tested two such subsets were available to us, MCF-7-MVLN and its two Tamoxifene resistant offshoots MCF-7-MVLN-6ms7 and MCF-7-MVLN-6ms8, as well as MCF-7-R and the three subclones we had derived; MCF-7-R-G1, MCF-7-R-D4 and MCF-7-R-F3. MCF-7-MVLN cells and its variants MCF-7-6ms7 and MCF-7-6ms8 presented rather homogeneous CGH patterns. Most anomalies found in MVLN cells were also present in 6ms7 and 6ms8 to the exception of 8 events; 1 gain (3p14) and 1 loss (6p22-p23) only present in the Tamoxifene sensitive cells, 4 losses (3p21, 9q21-qter, 12q24-qter, 16pter-q11) and 2 gains (2p13-p14, 4q22-q28) only present in the Tamoxifene resistant variants. MCF-7-R and its descendants MCF-7-R-D4, G1 and F3 showed more heterogeneous profiles of gains and losses. It was apparent that the 3 subclones derived from MCF-7-R presented a larger number of events than the mother line (Table 1). Seven regions of losses (4p16, 4q33-qter, 6p22-pter, 6q24-qter, 11p, 16p, 20p) and 6 regions of gains (1q22-q25, 3p11-q23, 4q28-q32, 5q, 8q12-q13, 12q13), observed in at least one of the subclones, were absent in parent cells (Figure 2). Conversely, some events present in parent cells were absent in at least one

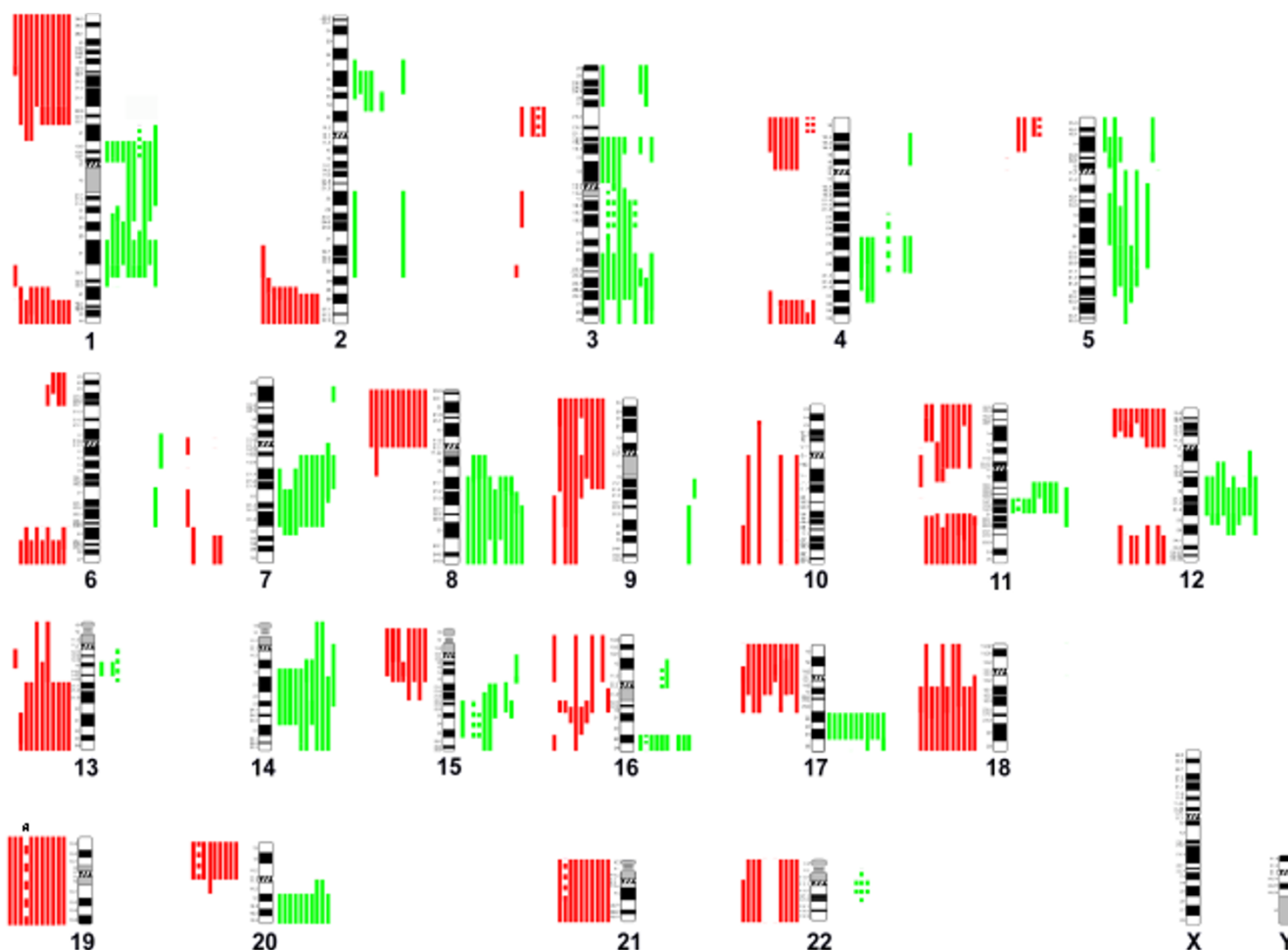


Figure 1

CGH profiles of 11 MCF-7 sublines. Copy number alterations are indicated as bars on each side of the chromosome ideograms, losses are shown by bars on the left, gains on the right. Each bar corresponds to an event observed in one subline. Events indicated by dotted lines corresponded to gains or losses reproducibly observed but whose fluorescence ratios did not reach the significance thresholds (1.3 or 0.75). Bars have been ordered from left to right for gains and from right to left for losses. The relative order was (1) MCF-7-R, (2) MCF-7-R-D4, (3) MCF-7-R-G1, (4) MCF-7-R-F3, (5) MCF-7-MVLN, (6) MCF-7-6ms7, (7) MCF-7-6ms8, (8) MCF-7-MF, (9) MCF-7-O, (10) MCF-7-MG, (11) MCF-7-ATCC.

subclone (gains at 2q22-q33, 3p22-pter, 5p15, 7q21, 13q12-q14, 15q22-q25 or losses at 10q, 12q23-q24 and 16q11-q13). These data suggested that MCF-7-R bear a higher level of clonal heterogeneity than MCF-7-MVLN cells. Further indications of clonal heterogeneity in MCF-7 cells could be found in the MCF-7-MF variant in which the loss of chromosome 19, shared by all the variants, was incomplete (i.e. the fluorescence ratio tumor DNA/normal DNA ranged between 1.0 and 0.75, whereas it was below 0.75 in the other variants) (Figure 1).

Phylogenetic analysis

The diversity of CGH patterns illustrates the genomic plasticity of MCF-7 cells and their capacity to acquire copy number aberrations. It was thus interesting to verify whether it was possible to reconstruct the phylogeny of the MCF-7 variants studied. This question is directly related to those classically addressed in evolutionary biology, where different species are ordered and hierarchized according to morphological and/or molecular characters. Hence, computational methods developed in systematics represented interesting tools to address the problem. We chose to apply a character based approach called cladistics under the maximum parsimony (MP) criterion, in which

Table 1: Number of copy number alterations found in MCF-7 variants.

Variants	Gains	Losses	Total number of events
MCF-7-R	18	16	34
MCF-7-R-D4	17	22	39
MCF-7-R-G1	18	20	38
MCF-7-R-F3	17	22	39
MCF7-MVLN-sens	15	22	37
MCF7-MVLN-6ms7	16	24	40
MCF7-MVLN-6ms8	15	21	36
MCF-7-MF	12	23	35
MCF-7-O	14	20	34
MCF-7-MG	18	23	41
MCF-7-ATCC	11	17	28

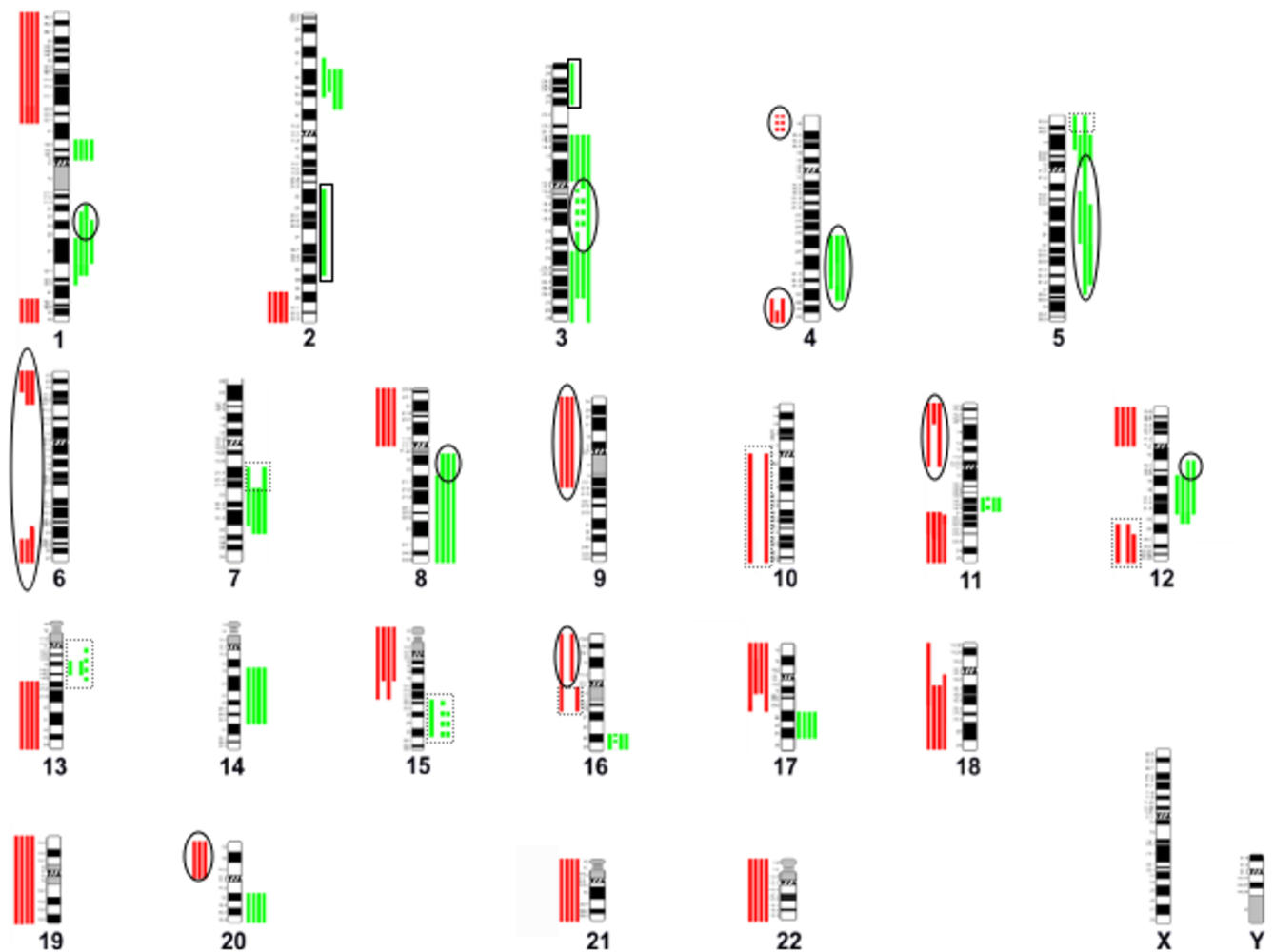


Figure 2
 CGH profiles of MCF-7-R cells and its three subclones. Events were ordered from left to right for gains and from right to left for losses. The relative order was (1) MCF-7-R, (2) MCF-7-R-D4, (3) MCF-7-R-G1, (4) MCF-7-R-F3. Circled events were specific to daughter clones. Boxed events correspond to gains or losses found only in the mother line (bold line) or in the mother and one or two subclones (dotted boxes).

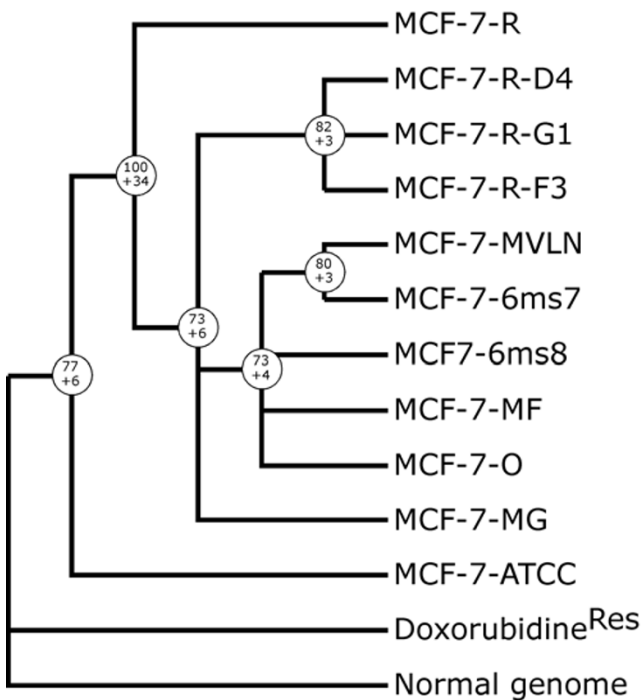


Figure 3

Phylogenetic tree describing the relationships between the MCF-7 sublines. The root was arbitrarily defined as corresponding to a genome devoid of any CNA (normal genome). The doxorubicin resistant line was also included in the analysis. Since it did not belong to the MCF-7 group it qualified as a potential outgroup and was indeed positioned as such by the analysis. This tree is a consensus tree corresponding to the 3 most parsimonious trees identified. It is 711 mutations long. Values represented at the nodes correspond to bootstrap percentages (top) and Bremer support indices (bottom). These values measure the robustness of the nodes.

different sublines were considered as taxa and copy number changes at every chromosomal band as characters. We favored maximum parsimony because it allows to order the different taxa and construct a phylogenetic tree requiring the fewest number of changes. In such a tree each cell line (or biological object) is represented as a leaf while nodes correspond to a collection of inferred characters encountered in hypothetical ancestors. A supplementary advantage of maximum parsimony is that it allows the identification of diagnostic events characterizing groups of cell lines. Each chromosomal band was considered to exist under three discrete states; normal, loss, gain. In the model we applied here, transformations from one character to another were equally weighted. Although this model did not perfectly match with CGH observations we adopted it as the most workable approximation. As a matter of fact, cytogenetic bands vary greatly in size and a

number of them are below the resolution limit of CGH. However, it is the only existing subdivision of chromosomal arms and it was not possible to base our analysis using chromosomal arms as a unit because of an insufficient number of characters.

The maximum parsimony analysis was done twice. In the first analysis we included only the 11 MCF-7 sublines and defined a normal genome as the origin or root of our putative tree. In the second we included the doxorubicin resistant cell line. Given its different genetic origin, it was interesting to check how this cell line positioned relative to *bona fide* MCF-7 variants in the phylogenetic tree. Furthermore, the order of MCF-7-R and its subclones MCF-7-R-G1, D4 and F3, as well as of MCF-7-MVLN and its Tamoxifene resistant offshoots MCF-7-6ms7 and 6ms8 were important indications of the reliability of the phylogenetic reconstruction method used. Figure 3 shows the phylogenetic tree corresponding to the analysis including the 11 MCF-7 variants and the doxorubicin resistant line. In this tree (or cladogram) doxorubicin resistant cells were positioned as an external group and MCF-7-ATCC occupied the position closest to the root, being the closest to a common ancestor. The next clade was formed by MCF-7-R which was identified as the ancestor of all other variants. The remaining sublines were ordered as three broad groups; one formed by MCF-7-MG, the second by the three MCF-7-R subclones which formed a discrete clade and the third with MCF-7-O, MCF-7-MF, MCF-7-MVLN and both Tamoxifene resistant clones. We noted that MCF-7-MVLN-6ms7 and MVLN-6ms8 were ordered at the same level as MCF-7-MF or MCF-7-O.

Diagnostic characters

Diagnostic characters were identified using the analysis in which only certified MCF-7 sublines had been included. Characters were considered as diagnostic for a given clade on the corresponding cladogram when they occurred once and only once during the evolution of the 11 MCF-7 variants. Events (losses or gains) selected as diagnostic characters corresponded to minimal consensus regions. Numbers of characters gradually added up when going down the tree. As shown in Table 2, 8 events (5 losses and 3 gains) were identified as diagnostic characters of the MCF-7 clade, since they were present in all the sublines, including MCF-7-ATCC. The number of diagnostic characters rose to 20 (9 corresponding to MCF-7-ATCC and 11 specific to R and its descendents) when MCF-7-R was taken as a starting point, 22 with MCF-7-MG and 25 with MCF-7-MVLN which is an endpoint on this tree.

Table 2: Diagnostic characters identified in the main nodes of the MCF-7 phylogenetic tree. Characters specific of each node (whose occurrence has been associated with the emergence of the corresponding branch) are presented in bold type sets. Events in italics correspond to characters passed on from ancestors.

MCF-7-ATCC	MCF-7-R	MCF-7-MG	MCF-7-MVLN
Loss 1p32-p36	<i>Loss 1p31-p36</i> Gain 1p13 Gain 1q31	<i>Loss 1p31-p36</i> <i>Gain 1p13</i> <i>Gain 1q31</i>	<i>Loss 1p31-p36</i> <i>Gain 1p13</i> <i>Gain 1q31</i>
Loss 2q36-q37	Loss 11q42-q44 <i>Loss 2q36-q37</i> Gain 3q26	<i>Loss 11q42-q44</i> <i>Loss 2q36-q37</i> <i>Gain 3q26</i>	<i>Loss 11q42-q44</i> <i>Loss 2q36-q37</i> <i>Gain 3q26</i> Loss 5q33 <i>Loss 6q25-q27</i> <i>Gain 7q22</i>
Loss 8p11-p23 Gain 8q22-q23	Gain 7q22 <i>Loss 8p11-p23</i> <i>Gain 8q22-q23</i> Loss 11q23-q25 Loss 12p13 Gain 12q15-q21 Loss 13q31-q34	<i>Loss 8p11-p23</i> <i>Gain 8q22-q23</i> <i>Loss 11q23-q25</i> <i>Loss 12p13</i> <i>Gain 12q15-q21</i> <i>Loss 13q31-q34</i>	<i>Loss 8p11-p23</i> <i>Gain 8q22-q23</i> <i>Loss 11q23-q25</i> <i>Loss 12p13</i> <i>Gain 12q15-q21</i> <i>Loss 13q31-q34</i>
Gain 14q21-q23	<i>Gain 14q21-q24</i>	<i>Gain 14q21-q24</i>	<i>Gain 14q21-q24</i> Gain 15q26 Gain 16p11
Gain 17q22-q24 Loss 18q12-q23 Loss 19p13-q13	Loss 17p11-p13 <i>Gain 17q22-q24</i> <i>Loss 18q12-q23</i> <i>Loss 19p13-q13</i> Gain 20q12-q13 Loss 21p13-q22	<i>Loss 17p11-p13</i> <i>Gain 17q22-q24</i> <i>Loss 18q12-q23</i> <i>Loss 19p13-q13</i> Loss 20p11-p13 <i>Gain 20q12-q13</i> <i>Loss 21p13-q22</i>	<i>Loss 17p11-p13</i> <i>Gain 17q22-q24</i> <i>Loss 18q12-q23</i> <i>Loss 19p13-q13</i> <i>Loss 20p11-p13</i> <i>Gain 20q12-q13</i> <i>Loss 21p13-q22</i>

RNA expression profiles

Because of the extent of the genomic changes shown by different MCF-7 sublines it was important to assess the consequences at the RNA expression level. We analyzed RNA expression profiles of 8 MCF-7 sublines (MCF-7-ATCC, MCF-7-O MCF-7-MF, MCF-7-MG, MCF-7-R, MCF-7-R-G1, MCF-7-R-F3 and MCF-7-R-D4) along with those of 19 unrelated breast cancer cell lines using home made cDNA arrays comprising 721 genes localized on chromosome 1q and 17q respectively. Expression data were analyzed by hierarchical clustering using the Cluster program. Seven of eight MCF-7 sublines were grouped within a cluster gathering 10 cell lines, whereas MCF-7-ATCC was grouped in an unrelated cluster (Figure 4A). Within the MCF-7 cluster it was noticeable that 6 sublines (MF, O, MG, R-F3, R-D4, R-G1) formed a tightly grouped subcluster together with KPL-1 cells, while MCF-7-R was ordered at a level equivalent to that of EFM-19 and EFM-192A cells. The position of MCF-7-ATCC, which coclustered with HCC 2218 and CAMA-1, was surprising. Although some differences were foreseen these results went beyond expectations and indicated the important distance this subline showed with other MCF-7 variants. A further question arose with the position of KPL-1 cells, which

grouped tightly with the MCF-7 subcluster. This raised some doubt on the true identity of this cell line. Upon verification of its haplotype KPL-1 turned out to be identical to MCF-7 cells. Altogether, these data show the elevated dispersion of MCF-7 expression profiles. However, because genomic differences observed between MCF-7 variants were not restricted to chromosomes 1q and 17q, we performed a complementary analysis on a set of 1000 genes selected for their proven or putative implication in cancer [25]. These genes were localized on all chromosomes. Seven MCF-7 variants (MCF-7-MVLN, MCF-7-ATCC, MCF-7-MF, MCF-7-MG, MCF-7-R, MCF-7-R-F3 and MCF-7-R-D4) were analyzed together with BT-474 breast cancer cells. Data were analyzed by hierarchical clustering and the dendrogram showed that while BT-474 behaved as an external group MCF-7-ATCC did not co-cluster with other MCF-7 variants (Figure 4B). Although the relative order found in this analysis is not identical to the one found in the first analysis, results were in accord confirming the divergence of MCF-7-ATCC cells.

Discussion

It is generally believed that divergence in cancer cell lines is the consequence of differences in culture conditions,

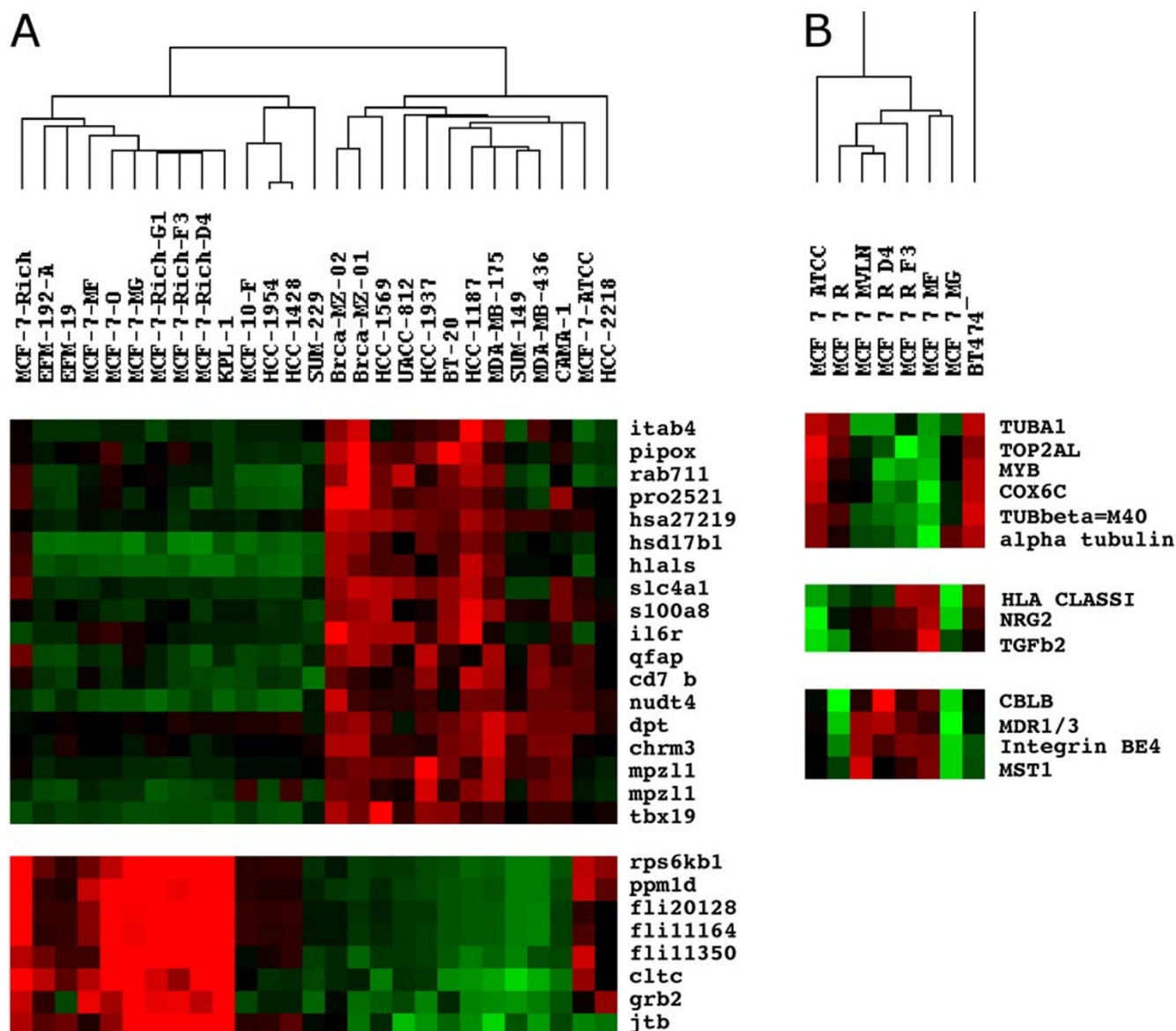


Figure 4
 Hierarchical clustering of RNA expression profiles. Panel A clustering analysis of expression profiles of 8 MCF-7 sublines along with those of 19 breast cancer cell lines. Expression profiling was done using home made Nylon arrays comprising 721 cDNAs corresponding to identified genes localized on either chromosome 1q or 17q. Panel B clustering analysis of profiles of 7 MCF-7 sublines and the BT-474 breast cancer cell line. Nylon arrays comprised 1034 genes selected on the basis of their involvement in cancer. Clustering analysis was done on raw quantification results, which were just subjected to a scaling step but not to ratio calculation. Parameters used in the analysis were Hierarchically Cluster Axes for Genes and Array: clusterand similarity metric correlation centered with average linkage clustering. The dendrogram on top of the diagram represents cell lines ordered according to their degree of similarity. Complete datasets can be found at <http://www.montp.inserm.fr/EMI0229/download>

which change the selective pressure and, thus, favor the selection of new genomic anomalies. If this situation is extended on a large number of cell passages it will lead to important differences between cellular stocks. The level of

divergence can be directly related to that of genetic instability and breast cancer cell lines seem particularly prone to it. Evidence for this can be found in recent work by Davidson and colleagues [12] and Kytola and colleagues [26],

who studied breast cancer cell lines using 24 color karyotyping or SKY. Seven cell lines were studied by both groups and, for 3/7, reported data presented extensive differences. Interestingly, MCF-7 cells were the most divergent in both studies adding further evidence to existing data on phenotypic or karyotypic variations in this cell line. MCF-7 cells of different origins are characterized by their variable chromosome numbers, which range from 55 to 90. Noticeably, some subsets present a bimodal distribution with a first peak at 70 chromosomes and a second one at 130 [8], indicating the coexistence of two cellular subpopulations, one of which had undergone endoreduplication.

Data presented here document that different MCF-7 variants underwent divergence at both the genomic and the RNA expression levels. Furthermore, they indicate that this can occur rapidly according to the MCF-7 variant considered. All the MCF-7 variants studied here showed extensive differences in their CGH profiles. These differences affected the number of regions of either losses or gains, which ranged from 28 in MCF-7-ATCC to 41 in MCF-7-MG, as well as the size of the regions involved. Remarkably, closely related sublines such as MCF-7-R and its 3 daughter clones MCF-7-R-D4, MCF-7-R-F3 and MCF-7-R-G1 presented variations in their CGH profiles as well. Daughter cells presented aberrations which were absent in the mother subline and, this was less expected, had lost anomalies present in the mother line. Furthermore, sister clones showed different sets of anomalies indicating that these cells bore the capacity to diverge over a limited number of cell generations, even kept in identical culture conditions. It is questionable whether this rapid upsurge of anomalies fits a linear progression model, where mutations are supposed to occur sequentially and be retained due to positive selection. We think more plausible that the differences shown by the 3 subclones be related to the oligoclonal nature of MCF-7-R parent cells. Anomalies found in the subclones in fact preexisted in MCF-7-R cells and were brought to light by cell cloning. In comparison MCF-7-MVLN and its two tamoxifene resistant derivatives MCF-7-MVLN-6ms7 and MCF-7-MVLN-6ms8 were less divergent. MCF-7-MVLN correspond to MCF-7 cells stably transfected with ERE-Luciferase construct and went through a gentamycin selection process. This could have lead to the loss of the preexisting genetic heterogeneity. We propose that MCF-7 cells contain an undetermined number of coexisting clones, out of which one (or several) possess stem clone potential and are responsible for the genetic oligoclonality.

The oligoclonal nature of MCF-7 cells can be related to aberrant or instable mitoses. Indeed, cells that have lost proper mitotic controls are prone to unequilibrated sister chromatid exchanges and tolerate the propagation of

damaged chromosomes. As such they rapidly become aneuploid and tend to accumulate physical aberrations. Such anomalies have been reported in cellular models in which the anaphase checkpoint gene MAD2 was disabled [27–29], as well as in human tumors [30]. Consequently instable mitoses will lead to rapid karyotypic changes. We, thus, verified the integrity of the M phase in MCF-7-R cells. MCF-7 cells did not show proper G2-M arrest when challenged with Nocodazole, a spindle inhibitor (data not shown). Our data are concordant with recently reported data by Yoon and coworkers [31] which showed that 7/9 breast cancer cell lines, among which MCF-7, presented important chromosome number variations.

The capacity to generate oligoclonality could be a strong selective advantage for cancer cells because it allows for rapid changes and as such confers an elevated genetic plasticity. Such tumor systems would evolve according to a nodal scheme (possibly through bursts) rather than following a linear selection model. Arguments in favor of a nodal evolution scheme stemmed from the phylogenetic analysis we have performed to reconstruct the history of MCF-7 sublines and identify diagnostic characters (CNAs). Because a number of analogies exist between evolution of species and that of tumor cells, classification methods developed for systematics have become increasingly employed to analyze genetic data in cancer. Approaches, based on hierarchical clustering or other distance-based models, have been applied to classify LOH [32] or CGH results [33,34]. We chose the maximum parsimony approach in a cladistic framework because it is a character based classification method and, as such, was considered to be best adapted to meet our goals [35]. We reconstructed the phylogeny of the MCF-7 clade and, interestingly, MCF-7-ATCC, which was the most divergent MCF-7 subline in our study, was positioned closest to the common ancestor. MCF-7-R came in second, positioned as the ancestor of all other MCF-7 sublines. Out of the total of 62 CNAs present in all the sublines tested, only 8 were selected as diagnostic of the MCF-7 clade. This means that this set of 8 events is shared by all the MCF-7 cells tested here and the original tumor possibly developed upon them. Thus, according to this phylogenetic tree MCF-7-ATCC and MCF-7-R, which bear respectively 28 and 34 CNAs, evolved from a common node. The robustness of these results was reinforced by bootstrap and Bremer analyses.

Given the extensive differences observed at the genomic level we were interested to check different MCF-7 sublines at the transcriptome level. Our RNA expression profiling results confirmed the divergent position of MCF-7-ATCC cells, which clustered with at some distance of other MCF-7 sublines. It, thus, appears from the expression profiling analysis, that MCF-7 sublines can show substantial

differences at both the genomic and RNA expression levels and this strongly suggests that the genomic differences could translate into phenotypic differences of possibly equivalent importance. MCF-7 cells are the most commonly used model for hormone responsive breast cancer and there is generally little knowledge concerning the variant used. Our data indicate that this may bear some importance, given the level of genetic variability these cells show and the rapidity with which they evolve.

Conclusions

In conclusion we want to propose that MCF-7 cells could represent an interesting model for genetic evolution of a subset of breast tumors. Breast tumors are prone to chromosomal instability and frequently show cytogenetic oligoclonality [1]. While some cancers were shown to fit the linear progression model, in which each step corresponded to the occurrence of an additional anomaly [36], other data brought evidence of more complex molecular evolution schemes [37]. This latter study compared CGH patterns of matched sets of primary breast tumors and asynchronous metastases. A number of metastases fitted the linear progression model, but it was noticeable that some presented very divergent sets of anomalies compared to their matched primary tumor. Only a limited set of (in some cases none detectable) aberrations were shared. The authors proposed the existence of a common early stem clone which diverged, evolved independently and ultimately lead to the formation of tumors with different locations. This scheme is very similar to what we observed in MCF-7 cells when the ATCC subline was compared to more distant offshoots of MCF-7-R. This leads us to propose that the capacity to generate clonal heterogeneity could represent an important selective advantage in some cancers and lead to aggressive and metastatic forms of the disease.

Abbreviations

CNA, copy number alterations; CGH, comparative genomic hybridization; ERE, estrogen responsive element.

Competing interests

None declared.

Authors' contribution

MN carried out most of the cell culture work, CGH and expression array analyses. PC(1) set up and carried out the analysis of CGH patterns by maximum parcimony as well as the clustering analysis of RNA expression profiles. JV helped setting up the expression profiling studies. BO coordinated the CGH analysis of breast cancer cell lines. LU carried out part of the CGH. CN helped setting up the expression profiling studies. DB helped setting up the expression profiling studies. EJPD helped carrying out the maximum parcimony study. PC(2) helped setting up the

expression profiling technology in the lab. CT conceived the study, its design and coordination and drafted the manuscript.

Acknowledgments

This work was supported by funds from the Association pour la Recherche sur le Cancer grant#5122, la Ligue Contre le Cancer program CIT, le Ministère de la Recherche et de la Technologie, Groupement des Entreprises en Lutte contre le Cancer and Gépole Montpellier-Languedoc-Roussillon. cDNA clones used to set up the arrays were provided by the HGMP (UK-MRC, Cambridge, UK). The contribution of EJPD was the one 2003-026 of the Institut des Sciences de l'Évolution de Montpellier (UMR 5554 CNRS). The authors wish to thank Drs Stephen Ethier, Alain Puisieux, Françoise Vignon and Jean-Claude Nicolas for providing cell lines.

References

1. Heim S, Teixeira MR, Dietrich CU and Pandis N **Cytogenetic polyclonality in tumors of the breast** *Cancer Genet.Cytogenet.* 1997, **95**:16-19
2. Flagiello D, Gerbault-Seureau M, Sastre-Garau X, Padoy E, Vielh P and Dutrillaux B **Highly recurrent der(1;16)(q10;p10) and other 16q arm alterations in lobular breast cancer [In Process Citation]** *Genes Chromosomes Cancer* 1998, **23**:300-306
3. Teixeira MR, Pandis N, Bardi G, Andersen JA, Mitelman F and Heim S **Clonal heterogeneity in breast cancer: karyotypic comparisons of multiple intra- and extra-tumorous samples from 3 patients** *Int.J.Cancer* 1995, **63**:63-68
4. Visscher D, Dolan P, Ottosen S and Crissman J **Geographic distribution of co-dominant DNA stemlines in breast carcinoma** *Cytometry* 1995, **21**:14-17
5. Bonsing BA, Corver WE, Fleuren GJ, Cleton-Jansen AM, Devilee P and Cornelisse CJ **Allelotyping analysis of flow-sorted breast cancer cells demonstrates genetically related diploid and aneuploid subpopulations in primary tumors and lymph node metastases** *Genes Chromosomes Cancer* 2000, **28**:173-183
6. Soule HD, Vazquez J, Long A, Albert S and Brennan M **A human cell line from a pleural effusion derived from a breast carcinoma** *J Natl Cancer Inst* 1973, **51**:1409-1416
7. Seibert K, Shafie SM, Triche TJ, Whang-Peng JJ, O'Brien SJ, Toney JH, Huff KK and Lippman ME **Clonal variation of MCF-7 breast cancer cells in vitro and in athymic nude mice** *Cancer Res* 1983, **43**:2223-2239
8. Whang-Peng J, Lee EC, Kao-Shan CS, Seibert K and Lippman M **Cytogenetic studies of human breast cancer lines: MCF-7 and derived variant sublines** *J Natl Cancer Inst* 1983, **71**:687-695
9. Butler WB, Berlinski PJ, Hillman RM, Kelsey WH and Toenniges MM **Relation of in vitro properties to tumorigenicity for a series of sublines of the human breast cancer cell line MCF-7** *Cancer Res* 1986, **46**:6339-6348
10. Osborne CK, Hobbs K and Trent JM **Biological differences among MCF-7 human breast cancer cell lines from different laboratories** *Breast Cancer Res Treat* 1987, **9**:111-121
11. Resnicoff M, Medrano EE, Podhajcer OL, Bravo AI, Bover L and Mordoh J **Subpopulations of MCF7 cells separated by Percoll gradient centrifugation: a model to analyze the heterogeneity of human breast cancer** *Proc Natl Acad Sci U S A* 1987, **84**:7295-7299
12. Davidson JM, Gorringe KL, Chin SF, Orsetti B, Besret C, Courtay-Cahen C, Roberts I, Theillet C, Caldas C and Edwards PA **Molecular cytogenetic analysis of breast cancer cell lines** *Br J Cancer* 2000, **83**:1309-1317
13. Jones C, Payne J, Wells D, Delhanty JD, Lakhani SR and Kortenkamp A **Comparative genomic hybridization reveals extensive variation among different MCF-7 cell stocks** *Cancer Genet Cytogenet* 2000, **117**:153-158
14. Courjal F, Cuny M, Rodriguez C, Louason G, Speiser P, Katsaros D, Tanner MM, Zeillinger R and Theillet C **DNA amplifications at 20q13 and MDM2 define distinct subsets of evolved breast and ovarian tumours** *Br.J.Cancer* 1996, **74**:1984-1989
15. Orsetti B, Courjal F, Cuny M, Rodriguez C and Theillet C **17q21-q25 aberrations in breast cancer: combined allelotyping and CGH analysis reveals 5 regions of allelic imbalance among**

- which two correspond to DNA amplification [In Process Citation] *Oncogene* 1999, **18**:6262-6270
16. Courjal F and Theillet C **Comparative genomic hybridization analysis of breast tumors with predetermined profiles of DNA amplification** *Cancer Res.* 1997, **57**:4368-4377
 17. Fitch WM **Toward defining the course of evolution: minimum change for a specific tree topology** *Syst. Zool.* 1971, **20**:406-416
 18. Swofford DL **PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4** Sunderland, Massachusetts, Sinauer Associates 1998,
 19. Maddison WP and Maddison DR **MacClade: analysis of phylogeny and character evolution, version 3.04** Sunderland, MA, Sinauer Associates 1992,
 20. Felsenstein J **Confidence limits on phylogenies: an approach using the bootstrap** *Evolution* 1985, **39**:783-791
 21. Bremer K **Branch support and tree stability** *Cladistics* 1994, **10**:295-304
 22. Bertucci F, Bernard K, Loriod B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K and Jordan BR **Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples** *Hum Mol Genet* 1999, **8**:1715-1722
 23. Eisen MB, Spellman PT, Brown PO and Botstein D **Cluster analysis and display of genome-wide expression patterns** *Proc Natl Acad Sci U S A* 1998, **95**:14863-14868
 24. Pons M, Gagne D, Nicolas JC and Mehtali M **A new cellular model of response to estrogens: a bioluminescent test to characterize (anti) estrogen molecules** *Biotechniques* 1990, **9**:450-459
 25. Bertucci F, Viens P, Hingamp P, Nasser V, Houlgatte R and Birnbaum D **Breast cancer revisited using DNA array-based gene expression profiling** *Int J Cancer* 2003, **103**:565-571
 26. Kytola S, Rummukainen J, Nordgren A, Karhu R, Farnebo F, Isola J and Larsson C **Chromosomal alterations in 15 breast cancer cell lines by comparative genomic hybridization and spectral karyotyping** *Genes Chromosomes Cancer* 2000, **28**:308-317
 27. Michel LS, Liberal V, Chatterjee A, Kirchwegger R, Pasche B, Gerald W, Dobles M, Sorger PK, Murty VV and Benezra R **MAD2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells** *Nature* 2001, **409**:355-359
 28. Jallepalli PV, Waizenegger IC, Bunz F, Langer S, Speicher MR, Peters JM, Kinzler KW, Vogelstein B and Lengauer C **Securin is required for chromosomal stability in human cells** *Cell* 2001, **105**:445-457
 29. Wang X, Jin DY, Ng RW, Feng H, Wong YC, Cheung AL and Tsao SW **Significance of MAD2 expression to mitotic checkpoint control in ovarian cancer cells** *Cancer Res* 2002, **62**:1662-1668
 30. Saunders WS, Shuster M, Huang X, Gharaibeh B, Enyenihi AH, Petersen I and Gollin SM **Chromosomal instability and cytoskeletal defects in oral cancer cells** *Proc Natl Acad Sci U S A* 2000, **97**:303-308
 31. Yoon DS, Wersto RP, Zhou W, Chrest FJ, Garrett ES, Kwon TK and Gabrielson E **Variable levels of chromosomal instability and mitotic spindle checkpoint defects in breast cancer** *Am J Pathol* 2002, **161**:391-397
 32. van Tilborg AA, de Vries A, de Bont M, Groenfeld LE, van der Kwast TH and Zwarthoff EC **Molecular evolution of multiple recurrent cancers of the bladder** *Hum Mol Genet* 2000, **9**:2973-2980
 33. Jiang F, Desper R, Papadimitriou CH, Schaffer AA, Kallioniemi OP, Richter J, Schraml P, Sauter G, Mihatsch MJ and Moch H **Construction of evolutionary tree models for renal cell carcinoma from comparative genomic hybridization data** *Cancer Res* 2000, **60**:6503-6509
 34. Kainu T, Juo SH, Desper R, Schaffer AA, Gillanders E, Rozenblum E, Freas-Lutz D, Weaver D, Stephan D, Bailey-Wilson J, Kallioniemi OP, Tirkkonen M, Syrjakoski K, Kuukasjarvi T, Koivisto P, Karhu R, Holli K, Arason A, Johannesdottir G, Bergthorsson JT, Johannsdottir H, Egilsson V, Barkardottir RB, Johannsson O, Haraldsson K, Sandberg T, Holmberg E, Gronberg H, Olsson H, Borg A, Vehmanen P, Eerola H, Heikkila P, Pyrhonen S and Nevanlinna H **Somatic deletions in hereditary breast cancers implicate 13q21 as a putative novel breast cancer susceptibility locus** *Proc Natl Acad Sci U S A* 2000, **97**:9603-9608
 35. Planet PJ, DeSalle R, Siddall M, Bael T, Sarkar IN and Stanley SE **Systematic analysis of DNA microarray data: ordering and interpreting patterns of gene expression** *Genome Res* 2001, **11**:1149-1155
 36. Larramendy ML, Lushnikova T, Bjorkqvist AM, Wistuba JI, Virmani AK, Shivapurkar N, Gazdar AF and Knuutila S **Comparative genomic hybridization reveals complex genetic changes in primary breast cancer tumors and their cell lines** *Cancer Genet Cytogenet* 2000, **119**:132-138
 37. Kuukasjarvi T, Karhu R, Tanner M, Kahkonen M, Schaffer A, Nupponen N, Pennanen S, Kallioniemi A, Kallioniemi OP and Isola J **Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer** *Cancer Res.* 1997, **57**:1597-1604

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2407/3/13/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

