Pathology and Biological Markers of Invasive Breast Cancer
CHAPTER 24

Genomic Events in Breast Cancer Progression

Rodrigo Goncalves, Burcu Cakar, Ron Bose, Cynthia X. Ma, and Matthew J. Ellis

CHAPTER CONTENTS

Next-Generation DNA Sequencing
Genomic Models of Breast Cancer Progression
Genomic Alterations Supporting Preinvasive Lesions as Precursors of Invasive Breast Cancer
LOH and CGH Studies
Expression Profiling
Genetic Alterations Indicating Distinct Genomic Pathways Associated with Low- and High-Grade Breast Cancers
Evolution from Lower to Higher Grade Lesions

Molecular Markers of DCIS That Predict Recurrence and Invasive Progression
Prognostic Markers
Candidate Drivers of Invasive Progression
The Role of the Tumor Microenvironment
Clonal Evolution During Invasive Progression
Sequencing of a Breast Cancer Primary and Metastasis
Comparison of the Clonal Evolution Model and the Cancer Stem Cell Model
Effects of Treatment on Clonal Evolution

INTRODUCTION

Breast cancers are thought usually to develop from regions of cellular atypia into clinically evident preinvasive or invasive lesions with subsequent evolution of lethal features such as metastatic spread and therapeutic resistance. Biological and statistical models that guide our thinking about breast cancer genomic progression events that underlie this process include linear, stepwise models, such as those proposed by Vogelstein and coworkers for colon cancer (Fig. 24-1A) and branched, Darwinian evolutionary models, such as that proposed by Nowell, where there is no fixed order of events or steps (Fig. 24-1B) (1,2). The recent application of next-generation sequencing technology has uncovered immense genomic complexity but has mostly focused on a single time point—the breast cancer genome at diagnosis. Longitudinal studies of individual cancer genomes over time will ultimately be required to completely characterize the genomics of breast cancer progression and are now possible as massively parallel sequencing platforms become more accessible. In this chapter, we will describe the theoretical basis for breast cancer evolution, from the earliest stages to advanced disease. We will review what is known about the dynamic structural changes in the genome that underlie the disease.

NEXT-GENERATION DNA SEQUENCING

The revolution in DNA sequencing technology, with the development of next-generation sequencing instruments, has provided a wealth of new data on the genomic evolution of cancer, and it is worthwhile beginning our discussion with an overview of this technology as many of the discussions in this chapter rely on data derived from these “massively parallel sequencing” (MPS) approaches. The development and commercialization of next-generation DNA sequencing instruments began around 2006 and has made DNA sequencing thousands of times faster and considerably cheaper. With this technological advance, projects to sequence large numbers of cancer genomes became feasible. Here, we will mostly focus on an MPS process called “sequencing-by-synthesis,” as most data sets have used Illumina-based technology. The steps in this process involve generating a DNA library from the sample of interest (Fig. 24-2A) and attaching individual DNA molecules to a glass slide called a flow cell. The individual sequences are then amplified into “clusters” to increase signal intensity (Fig. 24-2B). A sequencing reaction is then performed whereby a different colored fluorophor is activated when each of the four different nucleotides are added to the DNA polymer (Fig. 24-2C) (3,4). These sequential light reactions are captured by a sensitive light detector and the sequences entered into programs that align the sequences to the reference human genome. The process is remarkably efficient, and sequencing throughput is advancing rapidly. Currently more than 100 billion base pairs (Gigabases, Gb) can be generated per instrument run. Other methodologies and companies for next-generation DNA sequencing include the SOLID system developed by Applied Biosystems, and single-molecule DNA sequencers, which are still under development (3).

A key advantage of the MPS approach is that the analysis begins with an individual DNA molecule. This is fundamentally different than the original Sanger DNA sequencing method, developed over 35 years ago, which averages sequences across millions of DNA molecules and, as a
result, has limited ability to distinguish variations in DNA sequences. Because MPS provides data on the frequency of a DNA mutation within a population, tumor clonality or heterogeneity can be inferred from the variant allele frequency, and this provides new means for studying the clonal progression of cancer. Rare alleles can be detected down to a frequency of between 0.1% and 1% depending on the depth of the sequencing instrument and the fidelity of the enzymes used to generate the original clusters on the flow cell.

Much of the sequencing funded by the National Cancer Institute has focused on the initial diagnostic sample from invasive cancers through The Cancer Genome Atlas Project, with the goal to catalogue all the somatic mutations and structural abnormalities in breast cancer (5). The application of this approach in the study of breast cancer is still in its infancy. Our current understanding of the genomic alterations that promote preinvasive to invasive breast cancer is still largely based on comparative genomic hybridization (CGH) techniques, loss of heterozygosity (LOH) analysis, gene expression profiling, and selective gene sequencing studies that have been applied to synchronous preinvasive and invasive breast cancers.

**GENOMIC MODELS OF BREAST CANCER PROGRESSION**

To understand breast cancer progression, a description of the anatomy of the breast and the histology of preinvasive versus invasive breast cancer is required; the reader is guided to Chapters 1, 9, 21, and 25 which deal with these issues in depth. In brief, several models have been proposed to describe the development of breast cancer that focus on the relationship between preinvasive and invasive breast cancers. The most
**FIGURE 24-2** Next-generation DNA sequencing on the Illumina sequencing instruments. (A) DNA Sample preparation to create the library for sequencing. (B) Attaching the DNA to the surface of the flow cell and generating clusters of identical DNA molecules for sequencing. (C) Sequencing by synthesis. Each round attaches a DNA nucleotide to the DNA molecule and releases light, which is detected by a very sensitive camera. (From Ansourge WJ. Next-generation DNA sequencing techniques. *New Biotechnol* 2009;25(4):195–203.)
findings introduced the concept of usual ductal hyperplasia (UDH), in which cells pile up to fill the terminal duct (TD) and acini compared to the single or minimally pseudostratified layer of cells in FEA that distends the TDLU, as the direct precursor to ADH (12,13). However, UDH as the precursor for ADH has not been supported by recent immunohistochemical and molecular evidence. The LOH pattern observed in UDH is notably different from that associated with ADH and DCIS (8,14–18). This linear model of breast cancer progression provided the rationale for detection methods such as mammography in the hope of diagnosing and treating breast cancer at earlier and preinvasive stages before lethal features of the disease have developed (19). However, the occurrence of a preinvasive lesion is probably not an obligatory event in the development of invasive breast cancer. Although many premalignant lesions progress through the lifespan of some widely accepted “linear” multistep model suggests a transition from normal epithelium to invasive breast cancer via non-atypical and atypical hyperplasia and in situ carcinoma through accumulation of genetic mutations (6). In this classic Wellings model, premalignant breast lesions arise from terminal ductal lobular units (TDLUs) and give rise to flat epithelial atypia (FEA), atypical ductal hyperplasia (ADH), and ductal carcinoma in situ (DCIS), which subsequently progresses, over a long period of time, to invasive ductal carcinoma (IDC), whereas atypical lobular hyperplasia (ALH) and lobular carcinoma in situ (LCIS) progress to invasive lobular cancer (ILC) (Fig. 24-3) (7). This theory was originally based on the histologic observation of the gradual histologic continuity, but it also has been supported by analyses of genetic alterations that compared preinvasive and invasive breast cancers, especially when they occur in the same breast (8–11). Subsequent

![Multipathway Model](image-url)
patients, others might stay stable throughout their lives; it remains unclear which lesions have the capacity to progress to invasive cancer. The "nonlinear" or "branched" model builds upon the "linear" model in that it agrees that DCIS is the precursor for IDC, but hypothesizes that different grades of DCIS progress to corresponding grades of IDC. In contrast, the "parallel" model hypothesizes that DCIS and IDC are parallel and independent developments from a common progenitor cell through different grades (20). This is supported by the investigation of gene copy number changes in synchronous DCIS and IDC lesions, which demonstrated changes that are specific to DCIS or IDC (21).

**GENOMIC ALTERATIONS SUPPORTING PREINVASIVE LESIONS AS PRECURSORS OF INVASIVE BREAST CANCER**

The molecular differences among the preinvasive and invasive breast cancers are largely unknown but have been an area of great research interest with the hope to identify the key events that drive the development and progression of invasive breast cancer. The pathological heterogeneity and the microscopic size of the preinvasive lesions have posed a practical challenge in isolating sufficient material that is devoid of contaminating tissues. The availability of laser capture microdissection (LCM) and genome-wide analysis tools provide a new opportunity to discover genetic events specifically activated or inactivated in the course of breast cancer development.

**LOH and CGH Studies**

Multiple studies indicate that genetic alterations that confer the potential for invasive growth already exist in the earliest phenotypically recognized preinvasive stages. Initial studies of the genetic evolution of breast cancer progression used relatively insensitive loss of heterozygosity (LOH)/comparative genomic hybridization (CGH) techniques (8,10,22,23). For example, O’Connell et al. studied 399 microdissected preinvasive lesions (211 UD, 51 AD, 81 noncomedo DCIS, and 56 comedo DCIS) for LOH at 15 polymorphic genetic loci known to exhibit high rates of loss in invasive breast cancer (IBC) and assessed the sharing of LOH between synchronous DCIS and invasive cancers. For breast samples without DCIS and IBC, 37% of UDH and 42% of ADH lesions showed LOH in at least one locus, although loss at any given locus was uncommon (range, 0%–15%), suggesting that the development of hyperplasias can involve many different genes. In breast samples without IBC, LOH was common in DCIS, with 70% of noncomedo lesions and 79% of comedo lesions showing at least one loss with up to 37% of samples harboring LOH on chromosomes 16q, 17p, and 17q. When DCIS lesions from breasts with or without IBC were compared, substantially more LOH was observed in the breast with IBC at a few loci (on chromosomes 2p, 11p, and 17q), suggesting that genetic alterations in these regions may be important in the progression of DCIS to invasive disease. Among specimens harvested from breasts with IBC, 37% of concurrent UDH, 45% of ADH, 77% of noncomedo DCIS, and 80% of comedo DCIS lesions shared LOH with synchronous cancers at one locus or more. Similarly, in another CGH study performed on a panel of breast tumors that included 10 DCIS, 18 IBC, and two lymph node metastasis, there was an overall trend toward an increase in the number of genetic gains and losses in the IBC (24). In a study of 41 cases of sporadic breast cancer that focused on LOH of chromosome 11q13, LOH on chromosome 11q13 was present in 24 of 36 (67%) of the informative invasive breast cancer cases. The identical allelic loss was shared in the microdissected DCIS and the corresponding invasive breast cancer in 71% (15 of 21) of the available cases (23). Modrzejewski et al. analyzed 21 genes including transcription factors and thyrosine kinases in DCIS and adjacent IDC (25) and found that there were no copy number differences between them. These studies provided molecular genetic support for the notion that invasive breast cancer arises from preinvasive lesions and that DCIS is genetically as advanced as IDC and the driving genetic events already occurred at the preinvasive stage. This conclusion is thus somewhat paradoxical because DCIS is a benign disorder and invasive disease is not. We therefore still do not have a clear idea of the genomic determinants of the DCIS to invasive transition, which is a clear impetus for more detailed studies using MPS.

**Expression Profiling**

The molecular similarity between preinvasive lesions and invasive breast cancer has also been observed at the level of gene expression. Using LCM, T-7 based RNA amplification and DNA microarrays analysis, Ma et al. compared the gene expression profile of normal TDLU epithelium and synchronous ADH, DCIS, or IDC in a study of 36 breast cancer specimens (26). Compared to patient-matched normal epithelial cells, significant alterations in global gene expression occurred at ADH, which persisted in the later stages of DCIS and IDC. There were extensive similarities at the transcriptome level among the paired ADH, DCIS, and IDC without any consistent gene expression unique to each of the three identities. These observations were consistent with an earlier study of global gene expression profiles using serial analysis of gene expression (SAGE), although performed on a limited cohort of normal mammary epithelial cells, DCIS, IDC, and metastatic disease (27).

Similar to IBC, DCIS exhibits significant histologic and biological diversity between different cases. Under microarray gene expression analysis, intrinsic subtypes originally described for IBC have also been observed in DCIS (28–30). In a recent immunohistochemistry (IHC) analysis of a tissue microarray composed with 188 cases of pure DCIS (31), a frequency of 38.3% for Luminal A (ER+/PR+/HER2−), 9.9% Luminal B (ER+ and PR− and/or HER2+), 14.9% HER2 (ER−/PR−/HER2+), 7.5% TN (ER−/PR−/HER2−) and 4.2% basal-like (ER−/PR−/HER2−/CK5/6 and/or EGFR+) was observed according to IHC criteria of intrinsic subtypes (32). These studies indicated that the molecular heterogeneity of IBC is reflected at the stage of DCIS, and DCIS may be classified in a manner similar to invasive breast cancer.

**GENETIC ALTERATIONS INDICATING DISTINCT GENOMIC PATHWAYS ASSOCIATED WITH LOW- AND HIGH-GRADE BREAST CANCERS**

In gene expression profiling studies of synchronous DCIS and IDC lesions, the greatest alterations were among the different histological grades of DCIS and IDC (26). Notably, the grade I and grade III tumors demonstrated reciprocal gene expression patterns, whereas grade II tumors exhibited a hybrid pattern of grade I and grade III signatures. ADH samples demonstrated a grade I gene expression signature and clustered with the low-grade DCIS and IDC. Similarly, several comparative genomic hybridization studies revealed that the low-grade and high-grade DCIS have distinct gains and losses of genetic material. In the CGH study by Buerger et al. on 38 DCIS and 6 associated invasive breast cancers,


**TABLE 24-1**

<table>
<thead>
<tr>
<th>Chromosomal Aberrations of DCIS and IDC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low-grade DCIS</strong></td>
</tr>
<tr>
<td>16 q loss</td>
</tr>
<tr>
<td>11 q loss</td>
</tr>
<tr>
<td>17q12 and 11q13 amplification</td>
</tr>
</tbody>
</table>

Intermediate-grade DCIS lesions have common features with low- and high-grade IDC.

**EVOLUTION FROM LOWER TO HIGHER GRADE LESIONS**

The significant intratumoral histological and biological diversity within cases of DCIS argues for an evolution of DCIS from lower grade to higher grade (28). In an analysis of 120 consecutive cases of pure DCIS, 45.8% of cases showed areas of diversity in nuclear grades, including 30% with grades I and II, 6.6% with grades II and III, and 9.2% with grades I, II, and III. In addition, about one-third of the cases showing histologic diversity also showed biologic diversity for one or more biomarkers that included ER, GATA3, HER2, CK5/6, CK18, and p53 by immunohistochemistry staining. Similarly, in studies assessing the LOH, DCIS contained many of the same specific genetic defects regardless of histologic differentiation, although the absolute number of defects was found to be higher in higher grade lesions (8). Furthermore, a significant subset of genes expressed at higher levels in grade III DCIS compared to grade I DCIS were further elevated in IDC. In addition, the link between tumor grade and transition from DCIS to IDC is consistent with the clinical observation that grade III DCIS is more likely to be associated with occult invasive disease than grade I DCIS (36).

**MOLECULAR MARKERS OF DCIS THAT PREDICT RECURRENT AND INVASIVE PROGRESSION**

The evolution of genetic events that drives the process of breast tumorigenesis is poorly understood. As demonstrated in previous studies, many genetic alterations present in the invasive breast cancer already exist in the earliest phenotype-identifiable lesions such as ADH. However, gene expression profiling, CGH, and candidate gene approach, which are the main technologies used in these studies, are limited in their capacity for detailed genomic interrogation. A comprehensive genomic sequencing study of synchronous preinvasive and invasive cancer in comparison to the normal breast epithelial cells is needed but has not been reported. In this section, we will focus on available data investigating molecular markers that may predict the progression of DCIS and potential candidate drivers in cancer progression.

**Prognostic Markers**

Low ER or PR expression or HER2 amplification is associated with higher grades of DCIS and recurrence (37–41). A molecular signature of lack of ER and PR, HER2 overexpression, accumulation of p53, and high Ki67 expression was proposed to predict recurrence (42). COX-2 and p16 have also been associated with progression or recurrence. In a retrospective study, DCIS lesions that were positive for p16, COX-2, and Ki67 expression were significantly associated with risk of subsequent invasive cancer, whereas DCIS lesions that either lacked ER but were positive for HER2 and Ki67 or that lacked COX2 but were positive for p16 and Ki67 were associated with recurrence of DCIS (43). However, with the exception of ER/PR, none of these molecular markers are routinely assessed in the clinic due to the lack of sufficient evidence or established interventions.

Several other molecular markers, including cell cycle regulation and apoptotic markers (cyclin D1, cyclin A, cyclin E, p21, p27, p53, Bcl-2, Bax, Survivin, c-myc, and Rb), angiogenesis-related proteins (VEGF and heparanase-1), and extracellular matrix-related proteins (CD10, secreted protein acidic and rich in cysteine), have been investigated in molecular epidemiology studies; however, the data have not been conclusive (44).

**Candidate Drivers of Invasive Progression**

ER is commonly expressed in preinvasive lesions, 95% and 75% in ADH/LCIS and DCIS lesions, respectively (10). Binding
of estrogen to ER stimulates the growth and differentiation of breast epithelium; therefore, prolonged estrogen exposure in preinvasive disease might have a role in the development of breast cancer. Consistent with the important role of estrogen in breast cancer progression, tamoxifen has been shown to be an effective drug for prevention of breast cancer in high risk patients as well as an effective adjuvant hormonal therapy for patients with resected ER-DCIS.

HER2 overexpression occurs commonly in high-grade lesions (60%) compared to low-grade lesions (10%). In addition, HER2 is not overexpressed in TDLUs, very rarely in ADH, and about 2% in LCIS. The absence of HER2 overexpression in the earliest phase of preinvasive disease and its association with higher grade DCIS and more aggressive clinic behavior suggest HER2 overexpression is a driving event in cancer progression (45). A randomized phase III trial of adjuvant radiotherapy with or without trastuzumab in patients with HER2-DCIS resected by lumpectomy is ongoing to evaluate the effect of HER2-targeting in ipsilateral breast cancer recurrence (NCT00769379).

As mentioned above, the precise genetic event(s) triggered during the transition from DCIS to IDC is a critical unknown in the study of breast cancer. A few limited studies that compared the IDC with the adjacent DCIS suggested that c-Myc or FGFR1 amplification may be involved in this process because these genetic events occurred more frequently in the IDC compared to the adjacent DCIS lesions (46–48), while PIK3CA, AKT1, and TP53 mutations are early events that appeared to already exist at the DCIS stage (Table 24-2). Knudson et al. confirmed in their study that DCIS present in concert with IBC harbors gene expression profiles similar to IBC; however, when IBC and pure DCIS were compared, the expression differences became clearer (49). Genes associated with epithelial-to-mesenchymal transition and myoepithelial specific genes were enriched in IBC relative to DCIS, particularly in the stromal component. There have been few in vivo studies addressing the function of genes in the progression of DCIS to IDC. Using a “mammary intraductal DCIS xenograft model, Lee et al. studied the progression of DCIS to invasive breast cancer in vivo by introducing specific genes in the human DCIS cell line. Four genes, including a protease inhibitor (CSTa) and three genes involved in cell adhesion and signaling (FAT1, DST, and TMEM45A), which were usually elevated in clinical samples of DCIS, were found to suppress the progression of DCIS to invasive cancer (50).

### THE ROLE OF THE TUMOR MICROENVIRONMENT

The morphogenesis and functional differentiation of mammary epithelium are known to depend on signals from systemic hormones and on cues from the local tissue microenvironment, and epithelial-mesenchymal interactions are important for breast cancer tumorigenesis (51). Multiple lines of evidence point to the potential importance of tumor microenvironment, which is composed of fibroblasts, myoepithelial cells, endothelial cells, and various immune cells or leukocytes during the transition from invasive to metastatic breast cancer (52–54) and from DCIS to IDC (55,56). To analyze the contribution of tumor microenvironment, several groups have performed unbiased high-throughput genomic and transcriptomic analyses on different tissue/ cellular compartments of preinvasive and invasive breast cancer. Using cell-type specific antibodies, Allinen et al. isolated different cell types including epithelial cells, myoepithelial cells, myofibroblasts, leukocytes, and endothelial cells from normal breast, DCIS, or invasive breast cancer specimens and performed comprehensive gene expression profile and aCGH analysis of each cell type (55). While genetic changes by CGH were restricted to tumor epithelial cells of DCIS and IDC, gene expression changes are present

---

**TABLE 24-2**

<table>
<thead>
<tr>
<th>Genes</th>
<th>Method</th>
<th>Patient Features</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>c-Myc(45)</td>
<td>CGH and FISH</td>
<td>n = 12 IBC with large in situ component</td>
<td>c-Myc is amplified in IBC but not paired DCIS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n = 13 synchronous DCIS and IDC</td>
<td>c-Myc gain and CDH1 loss were the most frequent changes between DCIS and IDC</td>
</tr>
<tr>
<td>FGFR1(47)</td>
<td>FISH performed on selected gene on tissue microarray</td>
<td>n = 179 pure DCIS, n = 438 invasive carcinoma, n = 216 with DCIS component</td>
<td>FGFR1 amplification is more frequent in invasive carcinoma, associated with decreased overall survival (OS)</td>
</tr>
<tr>
<td>PIK3CA(48)</td>
<td>LCM, PCR sequencing for exon 9 and exon 20</td>
<td>n = 125 DCIS, n = 108 IBC</td>
<td>Similar in DCIS and IDC</td>
</tr>
<tr>
<td>AKT/PIK3CA(49)</td>
<td>LCM and PCR mutation analysis</td>
<td>n = 81 invasive <em>in situ</em> carcinoma</td>
<td>12/81 PIK3CA mutation, 3/78 AKT mutation, no difference in DCIS and IBC</td>
</tr>
<tr>
<td>TP53(50)</td>
<td>LCM, PCR mutation analyses</td>
<td>n = 32 DCIS, n = 38 IBC, n = 48 mixture</td>
<td>No significant difference in TP53 mutations in DCIS and IBC</td>
</tr>
</tbody>
</table>

CGH, comparative genomic hybridization; DCIS, ductal carcinoma in situ; FISH, fluorescence in situ hybridization; FGFR1, fibroblast growth factor receptor 1; IDC, invasive ductal carcinoma; LCM, laser capture microdissection; PCR, polymerase chain reaction; PIK3CA, phosphatidylinositol 3,4,5-bisphosphate 3-kinase; TP53, tumor protein 53.
in all cell types. The most consistent and dramatic gene expression changes occurred in myoepithelial cells from normal breast and those from DCIS samples. Interestingly, a significant fraction of these genes were secreted or cell-surface proteins, including CXCL12 and CXCL14 chemokines, suggesting paracrine interactions between myoepithelial and other cell types. Similarly, in an oligonucleotide microarray study performed on 14 patient-matched normal epithelium, normal stroma, tumor epithelium, and tumor-associated stroma from DCIS and invasive cancer, the transition from DCIS to invasive carcinoma was accompanied by significant increases in the expression of genes encoding extracellular matrix proteins and matrix metalloproteases (MMP2, MMP11, and MMP14) and cell cycle-related genes, in the stroma compartment (36). In contrast, the epithelial compartment demonstrates no or rare gene expression changes during the DCIS to IDC transition. These findings support the notion that stroma-produced MMPs may be key players driving the DCIS-to-IDC transition. Studies indicate the presence of distinct epigenetic changes in tumor-associated stroma cells (57,58). Future studies investigating the mechanisms of epigenetic changes may shed new light on the control of gene expression during breast tumorigenesis and tumor progression.

**CLONAL EVOLUTION DURING INVASIVE PROGRESSION**

Single-cell genetic analysis and next-generation sequencing studies support the hypothesis that IBC is the result of clonal evolution driven by a combination of an increased mutation rate and selection pressure on cells within the evolving malignant focus. Using a 4-FISH probe panel that targets 8 candidate genes, including oncogenes COX2, c-Myc, HER2, CCND1, and ZNF217 and tumor suppressor genes DBC2, CDH1, and TP53 and 2 centromere probes, single-cell analysis of copy number changes of the 8 genes was performed on 13 cases of synchronous DCIS and IDC (47). Signal patterns were counted in 76 to 220 nuclei per sample. A high degree of chromosomal instability, defined as variability in the signal patterns from one cell to another in a tumor population (Fig 24-4), was observed in both DCIS and IDC samples. Despite enormous intercellular heterogeneity in DCIS and IDC, nonrandom distribution of genomic imbalances was observed. The progression from DCIS to IDC was commonly accompanied by loss of CDH1 and gain of MYC (c-Myc). Four of 13 DCIS showed identical clonal imbalances in the IBC (see Fig. 24-4, Category I). Six cases revealed a switch, four of which acquired a gain of MYC in IDC (see Fig. 24-4, Category II). In one case, the major clone in the IDC was one of several clones in the DCIS (see Fig. 24-4, Category IV), and in another case, the major clone in the DCIS became one of the two major clones in the IDC (see Fig. 24-4, Category III). This data suggest that transition from DCIS to IDC is driven by a selection of clone(s) with a specific repertoire of genetic alterations. This hypothesis was further supported by another study of 13 matched DCIS and IDC pairs by a CGH and Sequenom MassARRAY (59). Although the genomic profiles of matched DCIS and IDCs were similar, amplification of distinct loci (i.e., 1q41, 2q24.2, 6q22.31, 1q11.21, 9q21.2, and 9p13.3) was either restricted to, or more prevalent in, one of the components in 3 pairs. **PIK3CA** mutations were restricted to the DCIS component in two cases, and reduced from 49% in the DCIS to 25% in the IDC component in the third case. Similarly, it is well known that some DCIS harboring **HER2** gene amplification are associated with **HER2-negative invasive carcinomas** (60,61).

Using newly developed bioinformatic algorithms (62), Nik-Zainal et al. reconstructed the genomic evolution and a model of breast cancer development over molecular time (Fig. 24-5A) based on analysis of NGS data obtained for 21 breast cancers that included ER+ (n = 5), HER2+ (n = 4), triple negative breast cancer (TNBC) (n = 3), **BRCA1** mutant (n = 5), and **BRCA2** mutant (n = 4) cases. An example of the phylogenetic tree constructed for PD4120, which was sequenced to 188-fold depth is shown in Figure 24-5B. The chronological orders of copy number gains in 16 informative breast cancer genomes are shown in Figure 24-5C. A key milestone in this evolutionary process is the appearance of the "most-recent common ancestor"—the cell with the full range of somatic mutations found in all tumor cells, which demarcates the point when divergent subclones branch out from the initial clone (see Fig. 24-5A,B) (63,64). Strikingly, many oncogenic events, including several driver mutations such as mutations in **PIK3CA** and **TP53**, amplifications of **ERBB2**, MYC, and **CCND1**, and somatic loss of the **BRCA1** and **BRCA2** alleles, accumulated before the emergence of the most-recent common ancestor and were identified in all tumor cells among the 21 breast cancers studied. Another important finding from the study was that all of the tumors contained a dominant subclonal lineage, accounting for more than 50% of cancer cells in the sample and carrying many hundreds or thousands of point mutations. Using PD4120a as an example, 26,762 of the 70,690 somatic substitutions genome-wide were present in all tumor cells, and 4 major subclones were present by statistical modeling of the distribution of clonal and subclonal mutations, with the dominant clone composed of an estimated 70% of the cells in the tumor sample (35% of sequencing reads reported this variant) (Fig. 24-5D). Chromosomal instability was found to be common throughout the history of the cancer although not usually the earliest genomic event. This results in the clonal acquisition of many recurrent abnormalities, such as gains of 1q and 8q and losses of 17p, and considerable divergence among subclones. Similarly, other mutations accumulate during the tumor’s development. Once again, it is not clear what triggers the development of the dominant clone. One theory is that this involves an event referred to as chromotrisomy (Greek; **chromos** for chromosome, **tris** for tripled into pieces). Chromotrisomy describes a genetic event in which tens to hundreds of genomic rearrangements interspersed with widespread losses of sequence fragments occur in a one-off cellular crisis (65). This is accompanied occasionally by the formation of small circular DNA molecules (double-minute chromosomes), which could become amplified with oncogenes. Strikingly, this genomic rearrangement has been found to be limited to one or a few chromosomes, with affected regions criss-crossing back and forth and showing the characteristic pattern of copy number oscillations between two copy number states. This phenomenon was recently discovered amidst a flood of information from the NGS, in which both ends of 50 to 100 million genomic DNA fragments per sample are sequenced and aligned to a reference genome. In the initial analysis of 10 chronic lymphocytic leukemia (CLL) cases, chromotrisomy was detected in one sample (Fig. 24-6) (65). Chromotrisomy was identified in 18 of 746 (2.4%; 95% CI, 1.5%-3.9%) cancer cell lines subsequently analyzed using the high-resolution single nucleotide polymorphism (SNP) array data (65). The affected cell lines were across many different tumor types including melanoma; small-cell lung cancer; glioma; non-small-cell lung cancer; synovial sarcoma; and esophageal, colorectal, renal, and thyroid cancers. Additionally, a similar proportion of cases demonstrated evidence of chromotrisomy in the analysis of SNP array data from 2,792 cancer specimens that composed 80% of primary tumors. This phenomenon
FIGURE 24-4  Schematic presentation of examples of clonal evolution in categories I (A and B), II (C), III (D), and IV (E and F) based on the presence of imbalance clones. In case 8 (A) the major clone in DCIS [an 18.8% gain of \( \text{MYC} \) (+8q), losses of \( \text{DBC2} \) (-8p) and \( \text{CDH1} \) (-16q)] was also the major clone in IDC (13.3%). A rare clone in the DCIS (a gain of \( \text{MYC} \), losses of \( \text{DBC2} \), \( \text{CDH1} \), and \( \text{ZNF217} \)) expanded to become the second largest clone, whereas the second largest clone in DCIS (a gain of \( \text{MYC} \), a loss of \( \text{CDH1} \)) became rare in the IDC. The text on the left of each panel denotes whether specific chromosome arms are gained (+) or lost (-). The sizes of the circles reflect the frequency with which a clone occurred, which is specified by the percentages in the circles as well. In (C), the clone that occurred in 7.4% of the IDC could have emerged by losses of 17p from either the major clone in the DCIS (+1q, -16q) or the IDC clone present in 13.7% of the cells. Note that one of the major clones in the DCIS (+1q, +20q) vanished in the IDC. (From Heselmeyer-Haddad K, et al. Single-cell genetic analysis of ductal carcinoma in situ and invasive breast cancer reveals enormous tumor heterogeneity yet conserved genomic imbalances and gain of \( \text{MYC} \) during progression. Am J Pathol 2012;181:1807–1822.)
FIGURE 24-5 (A) A model for breast cancer development over molecular time. The cancer evolves through acquisitions of driver mutations (black stars), which produce clonal expansions. These driver mutations occur only infrequently in long-lived lineages of cells, which passively accumulate many mutations without expansion.

(Continued)
has been found to be particularly common in bone cancers (9 or 20 tumors identified). The argument that the chromotriptic changes are a result of a single catastrophic event was based on the observation of the coordinated gene arrangement, with the restriction of two copy number and preservation of LOH of involved regions and was supported by a statistical analysis using Monte Carlo simulations of the progressive model of gradual accumulation of random alterations (65). The mechanisms underlying chromotripsy are unknown. One hypothesis is that chromosomes can be “pulverized” or undergo premature chromosome compaction (66), a phenomenon observed during cell-fusion experiments, in which incompletely replicated chromosomes from the S phase nucleus shatter when induced to undergo chromosomal condensation by signals from the host cell in mitosis (67,68). But how this process involves only one or two chromosomes or a single chromosome arm remains to be explained. A lack of sequence homology between joined segments of the regions affected argues that the nonhomologous end-joining DNA repair system is involved after the massive DNA fragmentation. The end results of chromotripsy are the survival advantage that could be offered when tumor suppressors are lost and the generation of new fusion genes in the disrupted chromosome, as well as amplified oncogenes occurring on the derivative chromosomes. Examples include the identification of a normal copy of chromosome 8 as well as a large number of double-minute chromosomes that are composed of 15 distinct segments of chromosome 8, leading to amplification of the MYC oncogene in a small-cell lung cancer cell line and the identification of simultaneous loss of several tumor suppressor genes including CDKN2A, WRN, and FBXW7 in a chordoma sample (65).

FIGURE 24-5 (Continued) (B) Reconstruction of the phylogenetic tree for PD4120a. The thickness of the branches reflects the proportion of tumor cells comprising that lineage. The length of the branches reflects the number of mutations specific to that lineage.

In 2009, Shah et al. described the mutational evolution of a lobular breast carcinoma (69). The DNA sequence of a metastatic, lobular breast carcinoma was obtained using next-generation DNA sequencing and comparison was made to the patient’s original primary breast cancer, which was resected 9 years previously. The metastasis contained 32 somatic, protein-coding mutations. Of these, five mutations were prevalent in the primary cancer, six were present at lower frequency in the primary cancer (between 1% and 13%) and were more prevalent in the metastasis, and nineteen mutations could not be detected at all in the primary. Another study, conducted by Ding et al., investigated the genomic differences between DNA derived from a primary
**FIGURE 24-5** (Continued) (C) Timing of copy number gains in 16 informative breast cancer genomes from the ploidy of mutations. The point estimates of timing for specific copy number gains are shown as arrows colored by the type of chromosomal aberration, with 95% confidence intervals generated by bootstrapping shown as horizontal lines. Molecular time is shown as an arrow, with the timing estimated as a fraction of point mutation time.
Genomic architecture of PD4120a, a breast cancer genome sequenced to 188-fold coverage: (1) Copy number profile of the sample, with the upper panel showing the logR of intensity, and the middle panel showing the B allele fraction (BAF) of germline heterozygous SNPs. Genomic segments of constant logR and BAF value were identified by the ASCAT algorithm (green lines). These were interpreted to give estimated overall copy number (purple lines) and copy number of the minor allele (blue lines) across the genome (lower panel). (2) Distribution of 70,690 somatically acquired base substitutions according to the total number of reads across that base (x-axis) and the fraction of those reads reporting the variant (y-axis). Points are colored according to the chromosome the mutation derives from. (3) Statistical modeling of the distribution of clonal and subclonal mutations by a Bayesian Dirichlet process. The empiric histogram of mutations is shown in pale blue, with the fitted distribution as a dark green line. Also shown are the 95% posterior confidence intervals for the fitted distribution (pale green area). Four separate clusters of mutations, named A–D, are identified. (4) Estimated number of mutations found in clusters A–D, with the error bars representing the 95% posterior confidence intervals. (From Nik-Zainal S, Alexandrov LB, Wedge DC, et al. The life history of 21 breast cancers. Cell 2012;149:994–1007.)
tumor, a metastasis, and a xenograft sample from the same patient (70). The samples were obtained from a 44-year-old African American woman with triple-negative breast cancer resistant to initial chemotherapy. The primary breast cancer contained 48 somatic, protein-coding mutations, which had a wide range of variant allele frequencies. The metastasis contained all 48 of these mutations, but about half of these mutations showed higher variant allele frequency in the metastasis, indicating enrichment or clonal selection in the metastasis. This enrichment was also seen in the xenograft that was derived from the primary cancer, and, because the sample to establish the xenograft was obtained prior to any cancer treatment, this argues that this enrichment or clonal selection is an intrinsic property of the cancer and not due to the effects of treatment. Further, two new mutations and one large DNA deletion were present in the metastasis but not in the primary cancer, indicating that there was some degree of genomic evolution in the cells making up the metastasis.

Studies of other cancer types also provide guidance about the type of genomic progression that can occur in breast cancer. Sequencing of a renal cell cancer that had metastasized to the lung and chest wall showed substantial intratumor genomic heterogeneity (71). The researchers sampled 9 different areas within the primary tumor and 3 metastases (1 from the perinephric fat metastasis; 2 from the chest wall metastasis) and found that only 31% to 37% of the mutations were common to all samples. They constructed a phylogenetic tree based on these results (Fig. 24-7) and proposed that the shared mutations are found in the trunk of this tree (ubiquitous mutations, indicated in blue) whereas the remainder of mutations (63% to 69%) are located in the branches of the tree (indicated in yellow, green, or red). Based on these cases, a schema of clonal evolution in both the primary tumor and metastasis can be proposed (Fig. 24-8) (72). Because of genomic instability in the cancer cells, heterogeneity and different subclones develop within the primary tumor. Metastases can develop either early or late in the cancer and are an opportunity for one or several subclones to grow at a distant site. The metastasis can derive from a dominant clone or a minor clone of the primary cancer, which will influence how similar the metastasis and primary cancer are in mutation pattern or even in response to treatment. The ability to sequence individual cancer cells (73) is providing further information about this clonal evolution process and will likely lead to future advances in this area.

**FIGURE 24-6** Clustered rearrangements on chromosome 4q in a patient with chronic lymphocytic leukemia. (A) Copy number between 70 Mb and 170 Mb of the chromosome oscillates between a copy number of 1 and 2, demarcated by back-and-forth interchromosomal rearrangements of all four possible orientations, as well as several interchromosomal rearrangements. (B) PCR gel of 12 putative genomic rearrangements identified by sequencing. PCR across the breakpoint is performed for each rearrangement on tumor DNA for samples taken at initial presentation (T1) and relapse (T2) as well as germline DNA (N). (C) Genome-wide profile of rearrangements in a sample taken before chemotherapy. Chromosomes range around the outside of the circle, copy number changes are shown by the blue line in the inner ring, and somatically acquired genomic rearrangements are shown as arcs linking the two relevant genomic points. (D) Genome-wide profile of rearrangements from the same patient 31 months later, at relapse after therapy, which showed all rearrangements present at initial presentation were present at relapse, and the striking copy number profile persisted. There were no new genomic rearrangements, suggesting that the process generating this complex regional remodeling had resolved before the patient was first diagnosed. (From Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 2011;144:27-40.)
Phylogenetic relationships of tumor regions

- Ubiquitous
- Shared primary
- Transient
- Shared metastasis
- Private

**Figure 24-7** Phylogenetic relationships of the different renal cell cancer regions and metastases. A single renal cell cancer was subdivided into 9 regions (labeled R1–R9). Additionally, a perinephric metastasis (M1) and a chest wall metastasis (subdivided into two halves, M2a and M2b) were sequenced. R4a and R4b are the subclones detected in R4. A question mark indicates that the detected SETD2 splice-site mutation probably resides in R4a, whereas R4b most likely shares the SETD2 frameshift mutation also found in other primary-tumor regions. Branch lengths are proportional to the number of non-silent mutations separating the branching points. Potential driver mutations were acquired by the indicated genes in the branch (arrows).


**Comparison of the Clonal Evolution Model and the Cancer Stem Cell Model**

In addition to the clonal evolution model discussed extensively up to this point, the cancer stem cell model also offers an explanation for tumor heterogeneity and cancer progression. The cancer stem cell hypothesis proposes that there is a subpopulation of cells within the tumor that are capable of self-renewal and multi-lineage differentiation (74,75). These cells can be isolated based on low expression of the heat stable antigen (CD24), high level expression of the hyaluronic acid receptor (CD44), and expression of aldehyde dehydrogenase (ALDH1) (76–78). In the cancer stem cell model, only mutations in these cells are propagated, and the clonal evolution in them gives rise to the genomic heterogeneity in the cancer (Fig. 24-9). Data to support this concept come from Nik-Zainal et al., where 21 primary breast cancer samples were sequenced and their clonal evolution was analyzed with bioinformatic algorithms as described earlier (63). They observed that all tumors contained a dominant subclone that accounted for more than 50% of cancer cells in the sample. They postulated that this expansion of a dominant clone is the final step in the development of a tumor that is responsible for triggering diagnosis, due to the emergence of a palpable mass. As there is minimal evidence of clonal expansion before the accumulation of all mutations in the dominant subclone, they suggest that the dominant clone becomes a cancer-initiating population, which is conceptually similar to a cancer stem cell (79). Similarly, Shah et al. performed deep sequencing, to a depth of 20,000X, on 104 triple-negative breast cancers (80). They observed that groups of mutations within individual cases have different clonal frequencies, indicative of distinct clonal genotypes. These triple-negative breast cancers had a wide range of clonal frequencies in the mutations sequenced, with some cases showing only one or two clonal populations (indicating a smaller number of clonal genotypes), whereas other tumours exhibited more extensive clonal evolution. The findings that many breast cancers have a dominant clone could be the result of this clone having a competitive advantage and taking over the tumor (the clonal evolution model) or could result from one or a few clones in the cancer stem cells, which then propagate and fill the tumor with their progeny.

The cancer stem cell model has also been proposed to explain the existence of the intrinsic molecular subtypes of breast cancer defined by gene expression (Fig. 24-10) (81,82). With the comprehensive genomic sequencing analysis of hundreds and thousands of invasive breast cancers in recent years, it is well established that the profound difference in the gene expression patterns among the various molecular subtypes is accompanied by subtype-specific genetic alterations (Table 24-3) (80,83–90). This suggests that the molecular subtypes are mechanistically different and perhaps derived from progenitor cells (or stem cells) at different stages of differentiation (81,91,92). Lim et al. isolated and functionally characterized the 4 populations of breast cells by fluorescence-activated cell sorting (FACS) analysis of EpCAM and CD49f (81). Using transplantation assays in immunocompromised mice, the CD49fEpCAM+ subpopulation was enriched with bipotent (ability to derive luminal and myoepithelial progenitors) mammary stem cell progenitors (MaSC), the CD49fEpCAM+ subpopulation was enriched for luminal progenitor, the CD49fEpCAM– subpopulation were committed luminal cells, and the CD49fEpCAM+ subpopulation were stromal cells. The data comparing the gene expression profiles of these different subpopulations of cells and that of invasive breast cancers suggest that Luminal A, Luminal B, and HER2-enriched breast cancers are likely derived from the more mature luminal cells, while basal-like tumors are derived from the less differentiated “luminal restricted progenitor” cells, and the claudin-low tumors are derived from the pluripotent or bipotent stem cell population (see Fig. 24-10) (81,82,91–93).

**Effects of Treatment on Clonal Evolution**

The development of endocrine therapy resistance or chemotherapy resistance in advanced breast cancers contributes significantly to patient mortality. Cancer treatment imposes a selective pressure on a tumor that can create an evolutionary bottleneck (Fig. 24-11). Drug resistant clones, which already existed as a minor population within the cancer, can be selected for and expand after cancer treatment (72,94). Genomic studies investigating the effects of treatment on breast cancer cells are still in progress, but evidence for this phenomenon comes from other cancer types. BRCA2 mutant ovarian cancers are sensitive to cisplatin and PARP inhibitors, but secondary mutations in the BRCA2 gene have been identified which give rise to resistance to these drugs (95,96). Similarly, resistance to imatinib in chronic myelogenous leukemia (CML) arises from secondary mutations in the BCR-ABL fusion protein (97). Chemotherapy or radiation therapy can itself induce mutations through DNA damage. Temozolomide treatment of glioblastoma multiforme...
Tumor heterogeneity in the primary tumor and metastases. Two models of clonal evolution are diagrammed here, with either early or late dissemination of cancer cells.

**FIGURE 24.8**

A. Clonal evolution model

B. Cancer stem cell model

**FIGURE 24.9** Comparison of the clonal evolution and cancer stem cell models. In the clonal evolution model, mutations can occur in any cell within the tumor.
FIGURE 24-10  Breast cancer intrinsic subtype in correlation to stages of normal mammary development.

![Diagram of breast cancer intrinsic subtype](image)

<table>
<thead>
<tr>
<th>Table 24-3</th>
<th>Highlights of Genomic, Clinical, and Proteomic Features of Subtypes&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtype</td>
<td>Luminal A</td>
</tr>
<tr>
<td>ER+/HER2− (%)</td>
<td>87</td>
</tr>
<tr>
<td>HER2+ (%)</td>
<td>7</td>
</tr>
<tr>
<td>TNBCs (%)</td>
<td>2</td>
</tr>
<tr>
<td>TP53 pathway</td>
<td>TP53 mut (12%); gain of MDM2 (14%)</td>
</tr>
<tr>
<td>PIK3CA/PTEN</td>
<td>PIK3CA mut (49%); PTEK mut/loss (13%); INPP4B loss (9%)</td>
</tr>
<tr>
<td>RB1 pathway</td>
<td>Cyclin D1 amp (29%); CDK4 gain (14%); low expression of CDKN2C; high expression of RB1</td>
</tr>
<tr>
<td>mRNA expression</td>
<td>High ER cluster; low proliferation</td>
</tr>
<tr>
<td>Copy number</td>
<td>Most diploid; many with quiet genomes; 1q, 8q, 8p11 gain; 8p, 16q loss; 11q13.3 amp (24%)</td>
</tr>
<tr>
<td>DNA mutations</td>
<td>PIK3CA (49%); TP53 (12%); GATA3 (14%); MAP3K1 (14%)</td>
</tr>
<tr>
<td>DNA methylation</td>
<td>—</td>
</tr>
<tr>
<td>Protein expression</td>
<td>High estrogen signaling; high MYB; RPPA reactive subtypes</td>
</tr>
</tbody>
</table>

<sup>a</sup>Percentages are based on a 466 tumor overlap list.

produces a statistically significant increase in patient survival, but subsequent recurrence and drug resistance in the cancer is frequently seen. Hunter et al. sequenced two recurrent glioma that had been treated with temozolomide and identified large numbers of somatic mutations in a pattern that was consistent with alkylating agent–induced mutations (98). They identified inactivating somatic mutations of the mismatch repair gene MSH6 in both of these cancers, and they proposed that these MSH6 mutations both conferred resistance to alkylating agents and triggered accelerated mutagenesis in the resistant clones. Accelerating mutagenesis will result in more clonal heterogeneity in the cancer and the potential to give rise to cancers that are more difficult to treat. In 1976, Nowell proposed that more research should be directed toward understanding and controlling the evolutionary process in tumors (2). Next-generation DNA sequencing provides a higher volume of data to understand this process, but the ability to control or reduce clonal evolution in cancer is still beyond our current capabilities.

CONCLUSIONS

Despite recent advances, it remains clear that insufficient effort has been placed on acquiring samples from patients with a range of premalignant lesions in order to determine the molecular landscape of precursor lesions. Similarly, at the other end of the scale, we still do not have a comprehensive catalog of the genomic landscape of advanced breast cancer. A comprehensive effort should be made to longitudinally sample the disease so that we can understand the genome dynamics of disease progression. It does appear that the process cannot be described as linear with a series of checkpoints, as Vogelstein imagined; this process is much more chaotic and complex than that, with any one tumor containing a spectrum of dominant and subdominant clones that constantly evolve in response to environmental and therapeutic stresses. Despite this, breast cancer is a curable illness in the majority of cases, which means that heterogeneity can be successfully overcome with the right treatment. However, for many patients who are not cured, genomic heterogeneity and Darwinian evolution at the cellular level is the root cause of their incurability.

REFERENCES


