JNCI J Natl Cancer Inst (2015) 107(5): djv015

doi:10.1093/jnci/djv015
First published online February 24, 2015
Commentary

COMMENTARY

Breast Cancer Genomics From Microarrays to Massively Parallel Sequencing: Paradigms and New Insights

Charlotte K. Y. Ng*, Anne M. Schultheis*, Francois-Clement Bidard, Britta Weigelt, Jorge S. Reis-Filho

Affiliations of authors: Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY (CKYN, AMS, BW, JSRF); Department of Medical Oncology, SIRIC, Institut Curie, Paris, France (FCB); Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY (JSRF).

*Authors contributed equally to this work.

Correspondence to: Britta Weigelt, PhD, Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065 (e-mail: weigeltb@mskcc.org) or Jorge S. Reis-Filho, MD, PhD, FRCPath, Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065 (e-mail: reisfilj@mskcc.org).

Abstract

Rapid advancements in massively parallel sequencing methods have enabled the analysis of breast cancer genomes at an unprecedented resolution, which have revealed the remarkable heterogeneity of the disease. As a result, we now accept that despite originating in the breast, estrogen receptor (ER)–positive and ER-negative breast cancers are completely different diseases at the molecular level. It has become apparent that there are very few highly recurrently mutated genes such as *TP53*, *PIK3CA*, and *GATA3*, that no two breast cancers display an identical repertoire of somatic genetic alterations at base-pair resolution and that there might not be a single highly recurrently mutated gene that defines each of the "intrinsic" subtypes of breast cancer (ie, basal-like, HER2-enriched, luminal A, and luminal B). Breast cancer heterogeneity, however, extends beyond the diversity between tumors. There is burgeoning evidence to demonstrate that at least some primary breast cancers are composed of multiple, genetically diverse clones at diagnosis and that metastatic lesions may differ in their repertoire of somatic genetic alterations when compared with their respective primary tumors. Several biological phenomena may shape the reported intratumor genetic heterogeneity observed in breast cancers, including the different mutational processes and multiple types of genomic instability. Harnessing the emerging concepts of the diversity of breast cancer genomes and the phenomenon of intratumor genetic heterogeneity will be essential for the development of optimal methods for diagnosis, disease monitoring, and the matching of patients to the drugs that would benefit them the most.

Breast cancer is a heterogeneous disease, comprising numerous entities that differ in their histologic, biological, and clinical behavior (1,2). Long before the advent of molecular biology technologies, histopathologists had classified the disease into groups based on morphology and clinical presentation (2–4). Although different histologic types of breast cancer have been shown to be associated with distinct biological and clinical features (2,5–8), the use of histologic subtyping for clinical decision-making has been limited. Additional criteria such as tumor grade, tumor size, lymph node status and vascular involvement have been shown to be required for breast cancer prognostication and treatment decision making (1,2,5,9). The predictive and prognostic power of these parameters, however, remains limited, and the clinical course of breast

Received: August 12, 2014; **Revised:** November 24, 2014; **Accepted:** January 12, 2015

cancer patients varies tremendously even when tumors of the same histologic grade are considered. High-throughput technologies, including microarray-based gene expression profiling (1,10–14) and massively parallel sequencing (MPS) (15–19), have not only helped define biologically relevant molecular subtypes of breast cancer but have also brought the remarkable heterogeneity of this disease to the fore. We now stand at a turning point, where the understanding of genomic instability and the complex mutational processes that drive tumorigenesis and result in intertumor and intratumor genetic heterogeneity is reshaping how we perceive breast cancer and will likely have a profound impact on how breast cancer patients are treated.

Breast Cancer Classification: The Present

The World Health Organization classification of invasive breast cancer currently used in pathology laboratories worldwide is based on both histopathology and immunohistochemical analyses (5). This classification recognizes the existence of "histologic special types," which account for up to 25% of all invasive breast carcinomas (2,5), while the remaining are classified as invasive carcinomas of no special type (IC-NST, formerly known as invasive ductal carcinoma not otherwise specified or invasive ductal carcinoma of no special type) (5). It should be noted that IC-NST, as the name implies, is a diagnosis of exclusion (ie, tumors that cannot be classified as one of the special histologic types). Histologic typing has not been included in clinical management algorithms, primarily because of the modest interobserver agreement rates and the controversy about the true existence of some entities (eg, medullary carcinoma and apocrine carcinoma) (5). Histologic grading, a measure of the differentiation of human breast cancers based on the analysis of tubule formation, nuclear pleomorphism, and mitotic index, has been shown to be of greater clinical importance than typing; in fact, histologic grading has proven to be an independent prognostic factor and is associated with the benefit patients derive from chemotherapy (20). In a way akin to typing, histologic grading has also proven to suffer from varying degrees of interobserver agreement. In addition to grading and typing, assessment of estrogen receptor (ER) and progesterone receptor (PR) expression by immunohistochemistry and of HER2 status by immunohistochemistry and in situ hybridization have proven to be useful predictive markers for the management of breast cancer patients. Importantly, however, these markers have been shown to have a high negative predictive value (ie, patients who have ER-negative cancers do not respond to endocrine therapy and patients with HER2-negative breast cancers do not respond to anti-HER2 therapies) but limited positive predictive value (ie, not all patients with ER-positive or HER2-positive breast cancers benefit from endocrine or anti-HER2 therapy, respectively). Hence, the development of more robust and reproducible molecular tools to predict the outcome of breast cancer patients and their response to treatment has long been a subject of great interest in translational research endeavors.

Microarray-based gene expression profiling studies carried out in the late 1990s and 2000s provided a paradigm shift such that we now recognize breast cancer not as a single disease but as different diseases with distinct transcriptomic profiles, clinicopathologic features, responses to therapy, and outcomes. Seminal class discovery studies have demonstrated that at the mRNA level, ER-positive and ER-negative breast cancers are fundamentally different diseases (Figure 1) and ER-positive tumors can be further subdivided into luminal A and luminal B "intrinsic" subtypes that differ by the levels of expression of proliferation-related genes and in clinical behavior (1,10–12,21). Within ER-negative tumors, in addition to the basal-like and HER2-enriched "intrinsic" subtypes initially described (10,11), further subtypes have emerged including claudin-low tumors, of which 60% to 70% are of triple-negative phenotype (ie, lacking ER, PR, and HER2 expression) and are potentially enriched for the so-called cancer stem cells (22) and the molecular apocrine subtype, characterized by the expression of androgen receptor, transcriptomic features consistent with activation of the androgen receptor

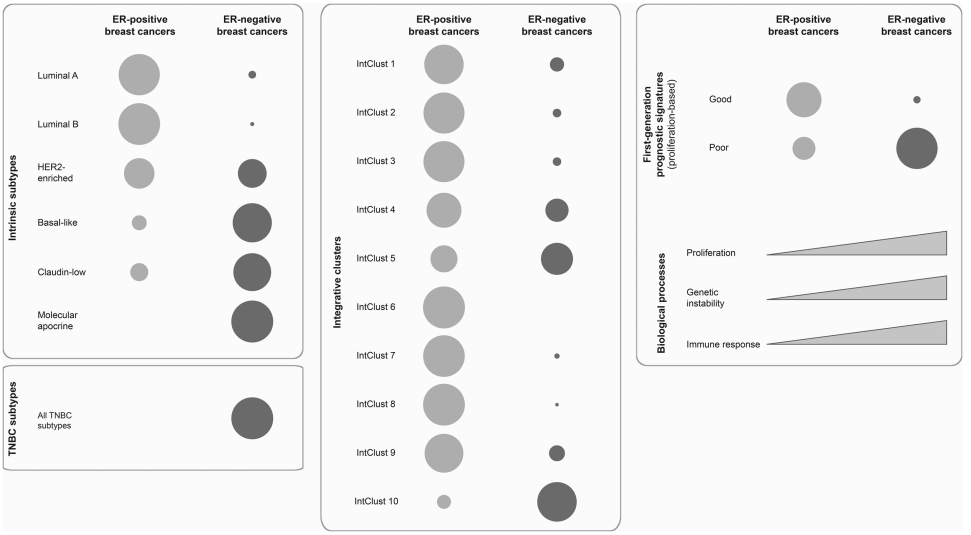

Figure 1. Estrogen receptor–positive and estrogen receptor–negative breast cancers: molecular profiling. Estrogen receptor (ER)–positive and ER-negative breast cancers differ at the molecular level as shown by the distribution of the "intrinsic" subtypes (10,11,22,23), the triple-negative breast cancer subtypes (26–28), the integrative clusters (IntClust) (14,30), first generation prognostic signatures (1), and biological processes as defined by microarray-based profiling (21). The size of each circle is proportional to the percentage of breast cancers harboring the characteristics stratified according to ER status. ER = estrogen receptor; TNBC = triple-negative breast cancer.

pathway, and poor clinical outcome (23–25). More recent independent studies have demonstrated that triple-negative breast cancers (TNBCs) can be subclassified into six subtypes (basal-like I, basal-like II, mesenchymal, mesenchymal stem-like, immunomodulatory, and luminal androgen receptor) (26) or four subtypes (luminal androgen receptor, mesenchymal, basal-like immune-suppressed, and basal-like immune-activated) (27). Together, these studies have resulted in an independent validation of the existence of the luminal androgen receptor and mesenchymal TNBC subtypes; however, the optimal subclassification of the remaining TNBCs remains to be fully elucidated. Importantly, these TNBC classification systems provide an interesting framework for the matching of subtypes of the disease with specific targeted therapies (27,28) and the six-subtype classification has been shown to be associated with distinct responses to neoadjuvant chemotherapy (29). Finally, a landmark study by the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) analyzed approximately 2000 tumors and proposed a genomics-driven classification of breast cancer based on an integrative analysis of gene expression and genome-wide copy number alterations (CNAs) (14) (Figure 1). Through bioinformatics methods, this study demonstrated that the most parsimonious number of molecular subtypes of breast cancer is likely to be 10 and that these subtypes have distinct clinical behaviors (14,30). In fact, gene copy number analyses of ER-positive and ER-negative breast cancers have shown that these tumors differ also in the pattern and type of gene CNAs: while the majority of ER-positive breast cancers (grade 1, 80%; grade 3, 50%) harbor concurrent deletions of 16q and gains of 1q, these concurrent alterations appear to be remarkably rare in ER-negative tumors (31). On the other hand, TNBCs are characterized by complex patterns of copy number gains and losses throughout the genome (32). Although the identification of the METABRIC integrative clusters initially required gene expression and CNA information, the proponents of this classification system have developed a gene expression–based approach to classify breast cancers into the 10 integrative clusters (33). The analysis of 7544 breast cancers with the new classifier has revealed that the METABRIC classification may be more informative in the contextualization of the genomic drivers identified by MPS studies of breast cancer (33) than the "intrinsic" subtypes (15). Microarray-based studies have undoubtedly demonstrated the diversity of breast cancers at the transcriptomic level; however, it is unclear as to how many of these subtypes would have clinical utility, and questions about the robustness of the methods for their identification have been asked (34–36).

Arguably, the most important contribution of microarray-based technologies to molecular tests from a practical standpoint was the development of prognostic signatures. Many commercially available gene expression–based platforms such as Oncotype DX (37), MammaPrint (12), Breast Cancer Index (BCI) (38), PAM50 ROR (11), and EndoPredict (39) have been implemented in the clinical setting to help physicians decide which patients have such a good outcome that they could forgo chemotherapy. These first generation gene signatures are composed of different gene lists; however, all identify very similar sets of breast cancer patients as of good or poor prognosis on the basis of the expression levels of proliferation-associated genes (1,21,40,41). Furthermore, these signatures have the highest discriminatory power in ER-positive disease; their prognostic value in ER-negative tumors is limited, given that more than 95% of ER-negative tumors have high expression levels of proliferation-related genes (1,21). It has been suggested that the prognosis of ER-negative breast cancers may be associated with the expression of immune-related genes (42). As a consequence, prognostic signatures linked to genes involved in immune, inflammatory, and/or chemokine pathways have been developed for hormone receptor–negative/TNBCs, including the STAT1 cluster (43), the IFN cluster (44), the IR-7 (42,45), the Buck-14 (46), TN-45 (47), and a B-cell/IL-8 metagene ratio (48) (Figure 1). At present, however, the prognostic value offered by proliferation-based prognostic signatures has been shown to be complementary to the prognostic information provided by classical clinico pathologic parameters (49), and some have received great acceptance by the medical and research communities (1,50).

The Mutational Landscape of Breast Cancers

MPS methods have allowed for the characterization of breast cancer genomes at the base-pair level and shown that at this resolution each breast cancer is likely unique (15–19). Overall, breast cancers were found to have on average 1.02 to 1.66 somatic mutations per Mb in coding regions (15-19,51), which translate into a mean of 56.9 (range 5–374) somatic mutations per cancer (52). The mutation frequencies found in breast cancers are similar to those of ovarian or renal clear cell carcinomas, but lower than those of bladder urothelial (8.03 somatic mutations/Mb) or lung squamous cell carcinoma (9.92 somatic mutations/Mb) (52). Depending on case selection, sample size, and analysis tools employed, different studies have revealed distinct sets of significantly mutated genes (SMGs) in breast cancer (Figure 2A). Importantly, however, *PIK3CA*, *TP53*, *GATA3*, *MAP3K1*, *AKT1*, and *CBFB* have been shown in multiple independent studies (15,16,52,53) to be SMGs and likely constitute drivers of the disease (Figure 2A). Overall, the genes significantly affected by mutations included genes and pathways known to be aberrant in breast cancers, including *TP53* or the PI3K pathway (eg, *PIK3CA*, *PTEN*, *AKT1*). However, MPS studies have also revealed several SMGs of functional or cellular processes previously not considered to be major players in the biology of breast cancer, including the MAPK/JNK signaling (eg, *MAP3K1*, *MAP2K4*, *NF1*), transcription factors and regulators (eg, *GATA3*, *RUNX1*, *CBFB*), splicing factors (eg, *SF3B1*), and chromatin remodelers (eg, *MLL3*, *ARID1A*) (15,52,53).

Of particular interest is the fact that only *TP53*, *PIK3CA*, and *GATA3* were found to be consistently mutated in more than 10% of unselected breast cancers, while the remaining genes were found to be mutated in less than 7.7% of cases, with a very long list of genes mutated in less than 1% of cases (Figure 2B) (15). Several studies have demonstrated the contribution of numerous low-prevalence mutations (15-18,54); however, algorithms to identify SMGs and potential drivers of the disease rely on mutation frequency and spectrum (55–57). In fact, in Stephens et al., out of the 40 driver genes identified in breast cancer, somatic genetic alterations (ie, mutations or gene amplifications) affecting seven of these genes (ie, *TP53*, *PIK3CA*, *GATA3*, *ERBB2*, *MYC*, *FGFR1*, and *CCND1*) accounted for 58% of all driver genetic alterations; the remaining 42% of somatic alterations affected the remaining 33 driver genes (17). Some low-prevalence mutations, however, affect bona fide genes that do play a role in breast cancer development and/or progression and have been shown to be activating or confer

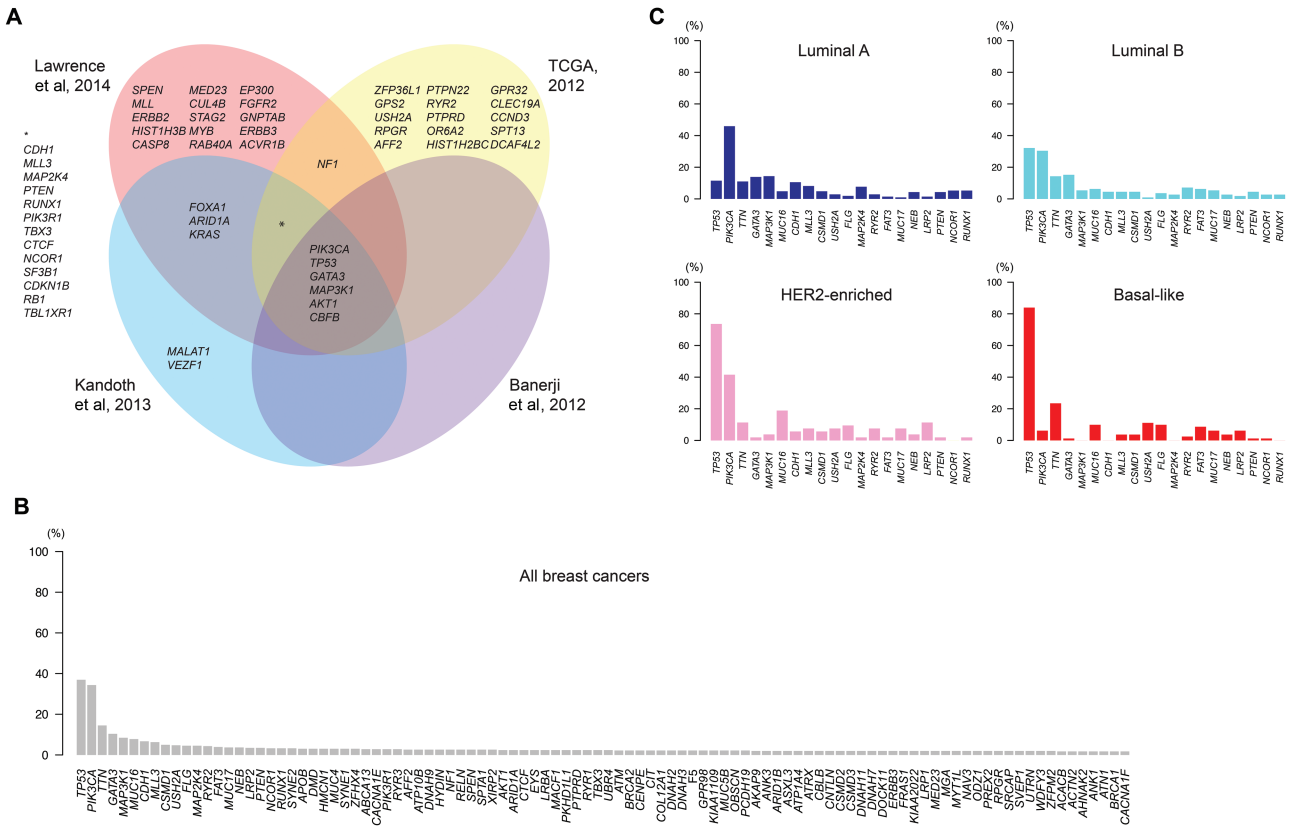

Figure 2. Intertumor genetic heterogeneity in breast cancer. At the genomic level, breast cancers are remarkably heterogeneous and no two tumors display an identical constellation of somatic mutations. **A)** Venn diagram illustrates the significantly mutated genes in breast cancer identified in different sequencing studies (15,16,52,53). **B)** Mutational frequencies of the 100 most frequently mutated genes in all breast cancers (15,52), illustrating the small number of genes highly recurrently mutated and a long "tail" of genes with low mutational frequency. **C)** The mutational frequencies of the 20 most frequently mutated genes in breast cancers of luminal A, luminal B, HER2-enriched and basal-like "intrinsic" subtypes (15,52). TCGA = The Cancer Genome Atlas.

therapy resistance. For instance, activating mutations in the tyrosine kinase domain of *ERBB2* (encoding HER2) have been found in approximately 1.5% of breast cancers and may affect response to anti-HER2 agents (58,59). Similarly, *ESR1* (encoding ER) mutations in the ligand-binding domain potentially confer resistance to endocrine therapy but are only found in 0.6% of luminal breast cancers (60–63). In fact, a recent saturation analysis demonstrated that at least 500 breast cancers need to be subjected to MPS for 90% of the genes mutated in 2% or more of breast cancers to be detected at 90% statistical power (53). This may explain our inability to identify infrequently mutated driver genes such as *ERBB2* and *ESR1* in studies involving unselected breast cancers, particularly because the repertoire of somatic genetic alterations in breast cancer varies according to ER status and "intrinsic" subtypes (see below). One may therefore hypothesize that the analysis of specific subsets of breast cancers, for example stratified according to histologic type, may lead to the identification of additional SMGs and driver genes.

Although it is difficult to ascribe biological or clinical significance to genes infrequently mutated in breast cancer, through the large-scale sequencing efforts of The Cancer Genome Atlas (TCGA), pooling together the major cancer types increases statistical power. Additional cancer genes have since been described across cancer types, and some of those previously implicated in other types of cancer have been found to be significantly though infrequently mutated in breast cancer (52,53). Two pan-cancer studies have identified 127 and 219 SMGs across multiple cancer types and have implicated cellular processes such as transcriptional regulation and genome integrity maintenance as somatically altered processes in breast cancers (52,53). It should be noted that in addition to the 32 SMGs identified when breast cancer was studied alone (53), an additional five genes (*EP300*, *FGFR2*, *GNPTAB*, *ERBB3*, *ACVR1B*) from the 219 SMGs were found to be statistically significantly mutated in breast cancer when the major cancer types were jointly studied (53).

The MPS studies further supported the main observations made using microarray technologies, in that ER-positive and ER-negative breast cancers differ at the molecular level. ER-positive tumors harbor fewer mutations (1.35 nonsilent coding mutations/Mb), which preferentially affect *PIK3CA* (40.1%), *MAP3K1* (11.0%), *MAP2K4* (5.6%), *GATA3* (13.8%), *MLL3* (7.6%), *CDH1* (8.5%), and *AKT1* (3.1%); on the other hand, ER-negative disease has a higher rate of coding mutations (1.94 nonsilent coding mutations/Mb), with *TP53* (84.5%) being the single most recurrently mutated gene (15,52). The microarray-defined "intrinsic" subtypes have also been shown to differ in terms of their constellation of somatic genetic alterations; the mean mutation rates for luminal A, luminal B, HER2-enriched, and basal-like subtypes have been reported to be 0.99, 1.58, 2.35, and 2.01 mutations/Mb, respectively (15,52). Interestingly, even within the HER2-enriched subtype, ER-positive tumors were found to have a lower mutation rate than ER-negative tumors

(1.85 vs 2.81 mutations per Mb) (15,52). Whilst the most frequently mutated gene in breast cancers of basal-like and HER2-enriched subtypes is *TP53* at 80% and 72%, respectively, this gene is reported to be mutated in 12% and 29% of luminal A and B tumors, respectively (15,52) (Figure 2C). In contrast, the most frequently mutated gene in luminal A and B breast cancers is *PIK3CA*, at 45% and 29%, respectively (15). Heterogeneity, again, exists within the HER2-enriched subtype, where only 63% of ER-positive HER2-enriched tumors harbor *TP53* mutations, compared with 92% of ER-negative HER2-enriched cancers (15). It should be noted that mutually exclusive mutations have been found within the "intrinsic" subtypes, namely mutually exclusive *MAP3K1* and *MAP2K4* mutations in the luminal tumors, suggesting that the same pathway can be affected in different cancers by genetic alterations in different components of the same pathways (15). Furthermore, "intrinsic" subtypes have also been shown to differ according to the patterns of mutations affecting specific genes. For instance, *TP53* mutations affecting basal-like tumors are enriched for nonsense and frameshift, whereas in luminal A and luminal B cancers missense mutations of *TP53* are more frequent (15). Another example of this phenomenon is observed in the spectrum of *GATA3* somatic mutations; in the TCGA study, hotspot deletions in intron 4 of *GATA3* were found only in luminal A tumors, whereas seven of nine frame-shift mutations in exon 5 were found in luminal B breast cancers (15). Although no single hotspot mutation or highly recurrently mutated gene defines the individual "intrinsic" subtypes, unsupervised clustering of the mutational repertoire of breast cancers identified five major clusters, defined by *TP53*, *PIK3CA*, *GATA3*, *MAP3K1*, and *CDH1* mutations (52), highlighting the extent of the diversity of somatic genetic alterations in breast cancer.

Taken together, multiple lines of evidence have shown that breast cancers are remarkably diverse in terms of their repertoire of mutations, with few highly recurrently mutated genes, and that there is great variation between tumors, even within "intrinsic" subtypes that were initially perceived to be homogeneous at the molecular level.

Intratumor Genetic Heterogeneity

Heterogeneity in breast cancer is not restricted to the phenotypic and genetic variation between tumors described above. There are several lines of evidence to suggest that solid tumors not only display striking morphologic heterogeneity, but also genetic diversity. Intratumor histologic heterogeneity within breast cancers is a frequent phenomenon, and one subtype of TNBCs, metaplastic breast cancer, is even defined by the presence of histologic heterogeneity (64). In a proof-of-concept study, our group demonstrated that in some cases histologically distinct areas of metaplastic breast cancers, while clonal as shown by the presence of identical *TP53* mutations, harbored distinct repertoires of CNAs (65). Cancers have been shown to be composed of mosaics of tumor cells that, in addition to the founder genetic events, harbor private alterations and follow a branched, Darwinian evolutionary trajectory (19,66–70). Intratumor genetic heterogeneity in the absence of overt phenotypic heterogeneity has also been demonstrated in breast cancer. Various degrees of intratumor genetic heterogeneity are evident in the majority of TNBCs, with approximately two-thirds of TNBCs being composed of genetically distinct clones at diagnosis (19) (Figure 3). Basal-like TNBCs tended to have greater intratumor heterogeneity than non–basal-like TNBCs, and, importantly, mutations in *TP53*, *PIK3CA*, and *PTEN* were usually present in high clonal frequencies, providing evidence that these are usually early driver mutations (19). It should be noted, however, that even mutations affecting these driver genes were also found in minor subclones in a minority of TNBCs, suggesting that some bona fide somatic driver mutations may be subclonal in TNBCs (19). Furthermore, the reconstruction of the evolutionary trajectory of 21 breast cancers suggested that while all tumors have a dominant subclonal lineage according to the definition employed in that study (>50% of tumor cells) (70), several somatic mutations were found in only a fraction of cancer cells. Several studies have documented that the constellations of somatic mutations found between distinct areas within a primary breast cancer (spatial heterogeneity) and between the primary breast tumor and its metastasis (temporal heterogeneity) are similar but not identical, with substantial variations in the number and type of mutations (19,66–68). These findings support the notion that breast cancers, similar to other tumors (69), evolve over the course of the disease.

In a landmark study, Navin et al. performed single-cell sequencing of 100 cells from a genetically heterogeneous breast cancer and showed that the tumor was composed of two largely homogeneous groups of aneuploid cells and more genetically diverse hypodiploid cells with distinct patterns of CNAs and ploidy (67). When the analysis was extended to single nucleotide variants (SNVs), however, it appears that in both an ER-positive breast cancer and a TNBC, although distinct clonal populations could be defined on the basis of their CNAs, no two single tumor cells harbored an identical repertoire of SNVs (71). Interestingly, in contrast to the evolutionary pattern of CNAs, rather than forming distinct subclonal populations, tumor cells displayed an evolutionary continuum with large numbers of subclonal and private SNVs (ie, found in only one tumor cell) (71). Whilst the ER-positive cancer was composed of a single dominant clone, based on CNAs, with numerous private or near-private SNVs, the TNBC had two CNA-defined clones with roughly equal frequency, one of which could be further subdivided into two subclones based on SNVs (71). The relatively higher degree of intratumor genetic heterogeneity observed in the TNBC may be associated with its higher mutation rate per cell division (71), in line with the previous observation that increased mutation rate is associated with increased genetic diversity in TNBCs (19).

Intratumor genetic heterogeneity needs to be contextualized, given that only a minority of the mutations are essential for cancer development and progression, whereas the majority have no significant biological impact or are deleterious (72,73). In some cases, intratumor genetic heterogeneity has been found to affect known driver genes such as *TP53*, *PIK3CA*, and *PTEN* (19). One potential explanation for the functional relevance of the heterogeneity, in particular that of driver genetic alterations, is that rather than competition, these genetically distinct subclones interact cooperatively, as described using a mixed-lineage mammary tumor mouse model, where the *Hras*-mutant basal cells are reported to depend on *Wnt*-expressing luminal cells (74). It has also been suggested that subclones of inferior selective advantage can play a role in tumor growth by inducing microenvironmental changes that promote the growth of all tumor cells in a non–cell-autonomous manner and the elimination of these subclones may result in tumor collapse (75). These observations demonstrate the cooperation of genetically distinct subclones within a tumor and may

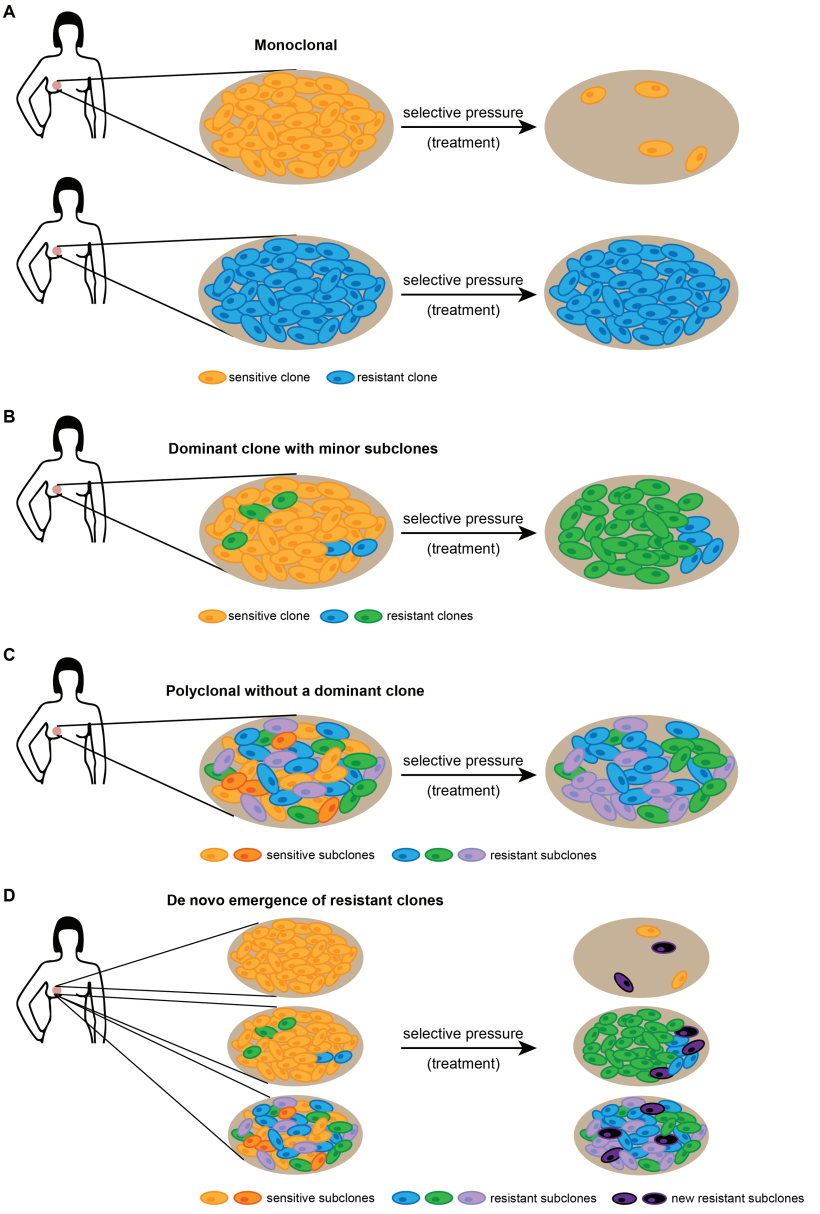

Figure 3. Intratumor genetic heterogeneity in breast cancer. There is evidence to suggest that breast cancers are composed of mosaics of tumor cells at diagnosis (19,66–68,81). Some tumors may only be composed of one or few tumor clones **(A)**, while others may harbor one major clone with several minor subclones **(B)** or may be composed of numerous clones without a dominant clone **(C)**. Under selective pressures, such as treatment, the fittest clones (ie, the resistant clones in the case of treatment-induced selective pressures) may be selected and thus become the so-called "lethal clone" driving tumor progression and finally being the cause of the patient's death. **D)** It is plausible that the de novo emergence of clones resistant to therapy may be causally linked to the therapy offered to the patient, should these therapies directly or indirectly result in mutagenic events that may confer a survival advantage to a given cancer cell.

provide an explanation for the heterogeneity of driver genetic alterations.

Importantly, primary tumors and/or their metastases may also vary in their targetable driver genetic alterations and there are examples of invasive breast cancers containing neoplastic cells with and without *HER2* amplification (76,77), which may result in the selection of clones with specific resistance mechanisms to anti-HER2 therapy that were already present before the onset of therapy (78–80). The recent identification of somatic mutations in the ligand-binding domain of *ESR1* at a much higher frequency in the metastases of breast cancer patients previously treated with aromatase inhibitors than in primary tumors provides further evidence for clonal evolution under selective pressure such as targeted anticancer treatment (60–63). In fact, evidence of *ESR1* gene mutations in patients not treated with aromatase inhibitors or estrogen deprivation is available (81). Importantly though, a given mutation can change from a passenger to a driver mutation under changes of selective pressure (80). For instance, whilst the *ERBB2* L755S mutation did not promote tumor formation in a xenograft model and was not considered to constitute an activating mutation, it is close to the binding site for small-molecule kinase inhibitors and has been shown to confer resistance to the dual HER2/EGFR kinase inhibitor lapatinib (58).

Also of clinical importance is the development of treatment-resistant lethal metastatic breast cancers. It has been suggested that the metastatic process likely constitutes a biological bottleneck (78), where the process of intravasation, survival in circulation, extravasation, and colonization of a distant site, coupled with pressures posed by the different microenvironments, would result in the selection of the fittest clone(s). In a substantial proportion of patients, however, metastases develop after surgical excision of the primary breast cancer and subsequent adjuvant systemic therapy; in this context, the clonal selection would stem from the selective pressures caused by surgical debulking, systemic therapies, and the metastatic process itself, which could result in clonal homogenization (82) (ie, a reduction in clonal diversity). In this model, the metastatic clone after multiple lines of therapy would be primarily constituted by the "lethal clone(s)", that is, the clone that is able to disseminate and colonize distant sites, and is responsible for therapy failure, resulting in the death of the patient. Consistent with this notion, the brain metastasis of one TNBC was shown to be less genetically heterogeneous than its primary tumor, suggesting that the selective pressures of chemotherapy and/or tumor microenvironment may have resulted in clonal homogenization (66). By contrast, the opposite has also been suggested, where distant metastases appeared to have higher clonal diversity than their primary tumors (83). In fact, despite the differences found between primary breast cancers and their matched metastases in individual patients (84), there is emerging evidence that the constellation of highly recurrent drivers of metastatic disease may be similar to that found in primary tumors (85), with a few exceptions, including *ESR1* mutations in ER-positive breast cancer patients treated with estrogen deprivation and *HER2* mutations in patients treated with anti-HER2 therapies. These observations may suggest that there is no common denominator in the form of a highly recurrently mutated gene or a highly recurrent mutation that drives metastatic behavior in breast cancer, that the metastatic process may be a convergent phenotype, and that in different patients distinct genetic and epigenetic mechanisms may result in the acquisition of metastatic properties by cancer cells. Thus, further studies are required to define the effect of the metastatic process and other selective pressures such as cytotoxic and targeted therapies on intratumor heterogeneity.

Mutational Processes Driving Inter- and Intratumor Genetic Heterogeneity

To understand the genomic instability that underlies inter- and intratumor genetic heterogeneity, several studies have sought to define patterns in the genomic alterations across common cancer types. Based on somatic mutations and CNAs, cancers can be classified into two main classes, namely M-class, primarily characterized and driven by recurrent mutations (SNVs and insertions and deletions [indels]) and C-class, by recurrent CNAs (86). Consistent with the notion of chromosomal and microsatellite instability in colorectal cancers, where tumors with microsatellite instability have limited chromosomal instability and vice versa (87), tumors with high numbers of mutations have been recently shown not to have many CNAs and those with many CNAs not to have many mutations (86). The vast majority of breast cancers (76%) belong to the C-class; M-class breast cancers are almost exclusively of luminal "intrinsic" subtype (92%), but only 66.5% of luminal breast cancers are in the M-class, with the remaining 33.5% in the C-class, suggesting that luminal breast cancers can be driven by both mutations and CNAs. By contrast, 99% of TNBCs are of C-class (86). Paradoxically, TNBCs, of which 83% have *TP53* mutations (15), have mutation rates similar to those of melanoma (15,19,86) but also have a large number of CNAs, similar to high-grade serous ovarian cancer (15,86). In addition, luminal breast cancers with *TP53* mutations overwhelmingly belong to the C-class (90%) and have relatively high mutation rates (1.73 mutations per Mb) compared with those without *TP53* mutations (59% C-class; 1.08 mutations per Mb) (15,52,86). These observations suggest that early *TP53* mutations may confer a mutator phenotype (88) and lead to copy number genetic instability. Although the paucity of recurrent SNVs and indels in TNBCs and *TP53*-mutant luminal tumors could be interpreted as suggestive of CNAs being the main drivers in these cancers, it is plausible that these tumors may constitute convergent phenotypes and be driven by mutations that either affect different components of the same pathway or different components of distinct pathways/networks whose alterations would result in a similar biological outputs (80).

These observations suggest that in some subtypes of breast cancer CNAs likely contain important drivers of the disease. Although this is immediately apparent in HER2-positive breast cancers, which are defined by the presence of *HER2* gene amplification, it is plausible that small subsets of ER-positive and TNBCs may be driven by amplification of specific genes (eg, *FGFR1*, *ZNF703*, and *CCND1* in a subset of ER-positive breast cancers, *FGFR2* in a subset of TNBCs, *PPM1D* in a subset of ER-positive and HER2-positive breast cancers) (Supplementary Table 1, available online) (14,15,30,79,89–91). One of the major challenges in translating the information stemming from gene copy number analyses of breast cancers lies in the fact that focal regions of high-level amplification often contain multiple genes, and defining whether an amplicon is driven by a single driver, the exact identity of the driver gene, whether the driver varies according to breast cancer subtype, or if there is cooperation between multiple genes amplified in the same amplicon or in other regions of the genome has proven challenging. It should be noted, however, that in addition to *HER2* gene amplification, amplification of other genes is currently being tested in breast cancer patients as potential biomarkers for specific targeted therapies (eg, NCT01795768, NCT02053636, and NCT00979134).

Genetic alterations in breast cancers have been found to be generated via a number of mutational processes, endogenous mutagens, and biological phenomena (86,92–96) (Figure 4). Overexpression of APOBEC3B and the APOBEC family of cytidine deaminases has been reported to catalyze deamination inducing C>T and C>G mutations at TpCpN in breast and other cancers (92–95). APOBEC activity has been found to contribute to, in particular, the large number of mutations in breast cancers with a localized hypermutator phenotype known as "kataegis" (Figure 4) (94,96). Of note, HER2-enriched breast cancers are particularly enriched for displaying the APOBEC mutation pattern (95). The association of APOBEC-mediated mutagenesis with *TP53* mutations suggests this may be a major endogenous mutagen in the presence of a checkpoint defect (92). In contrast, breast cancers with the fewest mutations were shown to be associated with the aging mutational signature (C>T at NpCpG) (96) (Figure 4), whereas tumors arising in *BRCA1* or *BRCA2* germline mutation carriers lack specific SNV signatures but are associated with deletions around 50 bp with overlapping

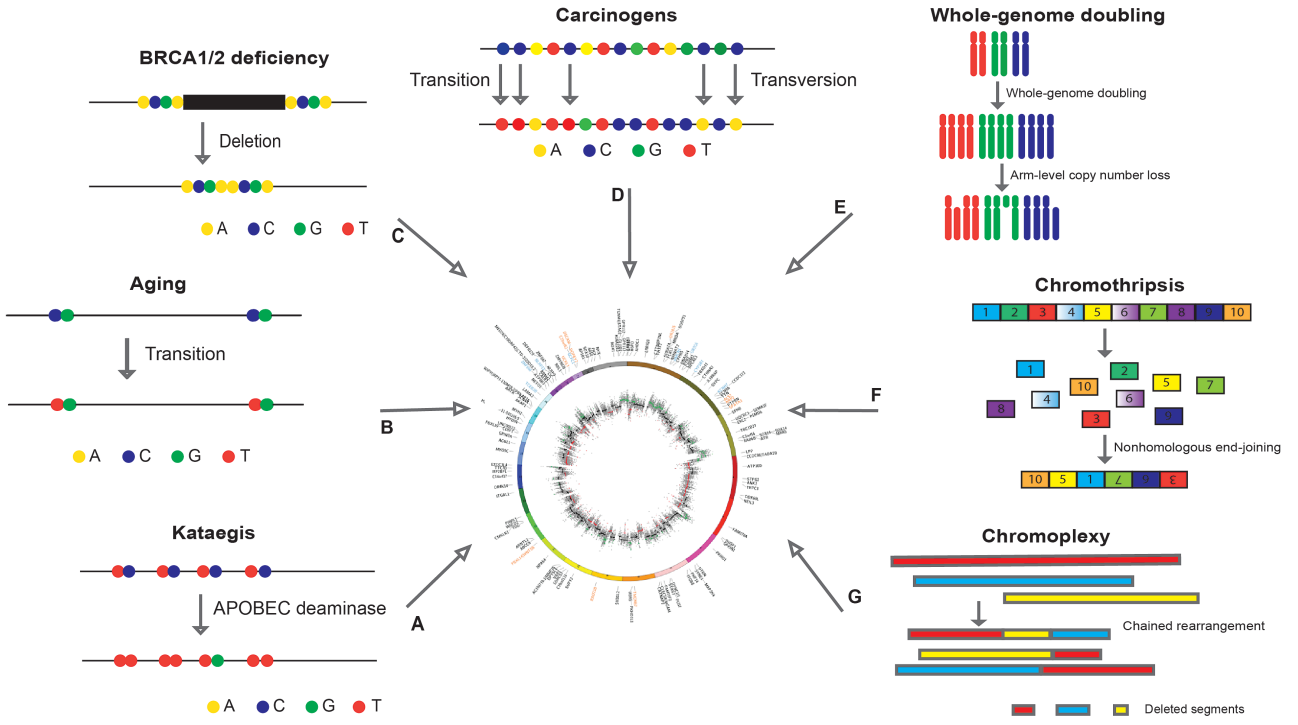

Figure 4. Mutational processes occurring in cancer. Breast cancer genomes are shaped through different mutational processes. **A)** Kataegis is a localized hypermutator phenotype associated with the overexpression of the APOBEC family of cytidine deaminases, which catalyze and induce C>T and C>G mutations at TpCpN. **B)** The mutational signature C>T at NpCpG has been associated with aging and is often observed in breast cancers with the lowest number of somatic mutations. **C)** Deletions of around 50 bp with overlapping microhomology are more frequently found in tumors arising in *BRCA1* and *BRCA2* mutation carriers. **D)** Different exogenous carcinogens cause specific mutational signatures. Tobacco use, for example, is highly associated with C>A base pair transversions, whereas ultraviolet light has been shown shown to predominantly cause C>T transitions. These mechanisms, however, do not appear to constitute major sources of mutagenesis in breast cancers. **E)** Whole-genome doubling is associated with increased rates of somatic copy number alterations, and arm-length copy number losses preferentially occur after whole-genome doubling. **F)** Chromothripsis, a localized catastrophic event of shattering and restitching of chromosomal segments, is found in a small subset of breast cancers. **G)** Chromoplexy, the rearrangement of large chromosomal parts as complex chains, is often associated with large DNA deletions at their junction, also referred to as "deletion bridges"; this process has not been documented in breast cancer as yet.

microhomology, as previously described (Figure 4) (51,94,96). Unlike other types of tumors such as lung cancers and melanoma, exogenous carcinogens do not appear to constitute major sources of mutagenesis in breast cancer (Figure 4), as robust evidence for the presence of carcinogen-induced mutation signatures has not been observed in this disease (96).

In terms of CNAs, whole-genome duplication and chromothripsis appear to play a role in shaping the genomic landscapes of breast cancers (Figure 4). Although it has long been known that breast cancers are frequently aneuploid, a recent analysis demonstrated that 45% of breast cancers have undergone at least one iteration of whole-genome duplication and that whole-genome duplication is associated with an increased rate of other types of somatic CNAs (97). Less common in breast cancers is chromothripsis (ie, localized catastrophic shattering and restitching of chromosomal segments), which has been found in 2% to 5% of all cancers and approximately 7% of breast cancers (97–99). Through whole-genome sequencing, chromothripsis has been reported to co-occur with kataegis (94), suggesting that multiple types of catastrophic events may contribute to the evolution of a tumor. Chromoplexy, however, defined as the formation of a chained pattern of rearrangements formed by interdependent rearrangements, has been observed in prostate cancer (100) but has yet to be documented in breast cancer (Figure 4).

Tackling Inter- and Intratumor Genetic Heterogeneity

While MPS studies have provided great insights into the genetics of breast cancers, inter- and intratumor genetic heterogeneity pose important challenges. There is evidence to demonstrate that the repertoire of somatic alterations found in a single sample of a primary tumor may not be representative of the entire disease, as illustrated by MPS analysis of primary breast cancers and their metastases (66,68,81) or even representative of a single tumor, demonstrated by divergent CNAs from anatomically distinct areas of primary tumors (65,71,101). Additional in-depth multiregional profiling of primary tumors and their respective metastases and serial tumor sampling at crucial time points of the disease are required to ascertain the level of spatial and temporal heterogeneity within cancers, although serial sampling may prove challenging in the clinical setting. Another clinical question that needs to be addressed is whether genomic profiling of single biopsies of the primary tumor would be sufficiently representative for critical therapeutic decisions to be rendered (79,80).

Actionable mutations or aberrations (ie, driver aberrations that can be targeted with specific therapies), even at subclonal levels, are of particular clinical interest. The thresholds at which subclonal mutations should be considered actionable (that is, the cellular frequency at which an actionable mutation should

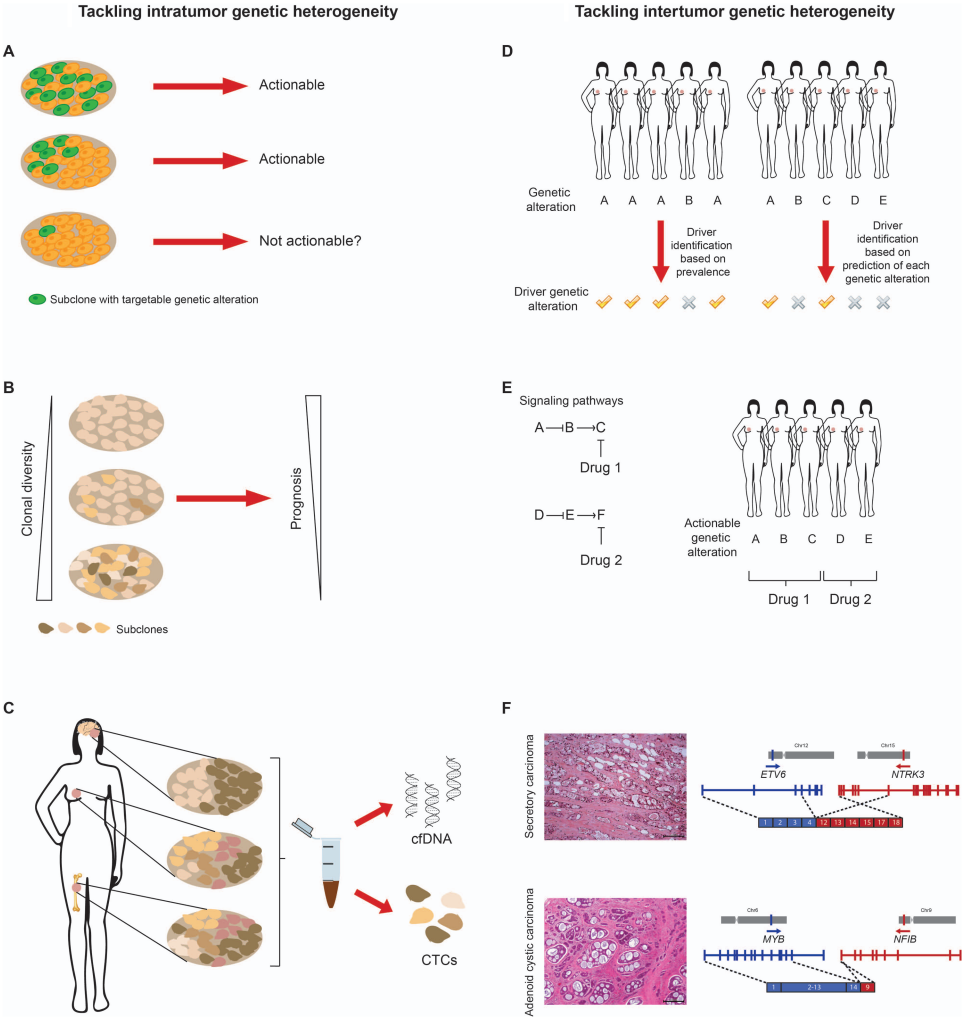

Figure 5. Tackling inter- and intratumor genetic heterogeneity. **A)** Actionable genetic alterations present even at the subclonal level in a given cancer are of particular clinical interest. A consensus about the thresholds at which subclonal genetic alterations should be considered actionable has yet to be reached. **B)** It has been hypothesized that the subclonal structure and heterogeneity within tumors may be associated with response and relapse rates and thus may be associated with outcome. **C)** Genetic analysis of circulating tumor cells (CTCs) and cell-free DNA (cfDNA) may overcome sampling biases and may serve as surrogate sources for the assessment of spatial and temporal intratumor genetic heterogeneity as well as minimal residual disease. The accurate detection of genetic alterations in CTCs and cfDNA, in particular in patients with early-stage breast cancer, remains challenging. **D)** Algorithms used for the identification of driver and passenger mutations often rely on the prevalence of a given genetic alteration across breast cancers; however, low-frequency genetic alterations may be bona fide drivers of the disease and their identification needs to be performed on an individual basis. **E)** Convergent phenotypes have been observed in breast cancer, where a given signaling pathway is dysregulated by genetic alterations targeting different components of the pathway, and may be exploited therapeutically. Using this approach, many tumors previously thought to be unsuitable for targeted therapies may harbor mutations that are in fact clinically actionable. **F)** Some special histologic types of breast cancer are underpinned by highly recurrent genetic alterations such as adenoid cystic carcinomas and secretory carcinomas, which harbor the oncogenic *MYB-NFIB* and *ETV6-NTRK3* fusion genes, respectively. Scale bars = 500μm.

be considered a potential drug target) have yet to be fully defined (Figure 5A). Intuitively, the driver mutations would have to be present in a substantial proportion of tumor cells to be considered candidate drug targets; however, this does not mean that mutations found at low mutant allelic fraction are of no biological and/or clinical importance. In fact, while subclonal resistance-associated mutations may not be of immediate clinical interest at the outset, the clinically relevant thresholds for these mutations may have to be much lower than current thresholds for "actionable mutations", given that enrichment of subclonal mutations after treatment may contribute to disease relapse (102,103). Furthermore, submodal mutations may still be driver events if their biological impact results in paracrine signals essential for the dominant clone (75). It is, therefore, not only important to employ sufficiently sensitive and specific techniques, but also to define which subclonal mutations and mutant allelic fractions should be considered clinically relevant at diagnosis.

While MPS has provided an unprecedented view of the subclonal structure of tumors, its translation into benefit for cancer patients has yet to take place. In theory, the subclonal structure and heterogeneity within the primary tumor may be associated with response and relapse rates (78,79) (Figure 5B). Mutation rate, one of the factors that influences the extent of intratumor genetic heterogeneity, has been shown to be higher in aromatase inhibitor–resistant than in aromatase inhibitor–sensitive breast cancers (18). The quantification of intratumor genetic heterogeneity may provide clinically useful information; several indices

have been employed, including the Shannon index (83,104) and the measure mutant-allele tumor heterogeneity (MATH) (105). There is no consensus, however, on which approach would be ideal for breast cancer prognostication and prediction of therapy response.

More recently, circulating blood biomarkers such as circulating tumor cells (CTCs) and cell-free plasma DNA (cfDNA) have gained attention as potential sources of tumor material for genetic analyses (ie, liquid biopsies) (106–109). Although CTCs have been shown to be prognostic for patients with breast cancer (108,110), their use as a source of biological material for MPS analyses of tumors has proven challenging but possible (111). For reviews on the use of CTCs for the molecular characterization of cancers, the readers are referred to Alix-Panabieres and Pantel (106) and Bidard et al. (108). cfDNA, however, has been more enthusiastically embraced as a surrogate of tumor genetic material, given that it can be obtained with simple DNA extraction from plasma. Circulating tumor DNA in plasma (ctDNA) is present at varying proportions in cfDNA, likely originates from all tumor masses (ie, primary tumors and metastases), and has been shown to be associated with disease burden (108,112,113). In addition, there is evidence that ctDNA may overcome sampling bias and may serve as a less invasive surrogate biomarker for spatial and temporal intratumor genetic heterogeneity and for the monitoring of minimal residual disease (Figure 5C) (81,112–114). Most of the studies on the genetic characterization of ctDNA have been performed in patients with metastatic breast cancer; its usefulness in early-stage disease has yet to be fully established; however, highly sensitive detection methods have demonstrated that approximately one in two patients with early-stage breast cancer have detectable levels of ctDNA (113). Although cfDNA detection of specific mutations previously identified through genetic analysis of the primary tumor or metastatic lesions can be robustly performed, de novo mutation detection on the basis of whole-exome or targeted capture MPS has proven challenging, owing to the limited proportion of ctDNA in cfDNA in the majority of cases (81,114).

One question that is germane to the translation of MPS findings into advancements in biology and therapy decision-making is the differentiation of driver from passenger mutations (Figure 5D). The current tools used to identify candidate driver genes, such as MuSiC and MutSig, rely heavily on mutational frequency, while variably taking into account background mutation rate, transcriptional regulation, DNA replication timing, and gene size (55,56). As a complementary approach, computational algorithms, such as cancer-specific high-throughput annotation of somatic mutations (CHASM), functional analysis through hidden Markov models (FATHMM), combined annotation dependent depletion (CADD), and Mutation Assessor, have been proposed to distinguish potentially pathogenic from nonpathogenic mutations on an individual mutation basis (115–118). These algorithms are based on various evolutionary, structural, and sequence annotations and in some cases cancer-specific frequency information, but their performance varies. When benchmarked against a set of mutations found in the COSMIC database, the various algorithms showed accuracy ranging from 61% to 89% (119). At present, there is no infallible algorithm to differentiate pathogenic from nonpathogenic mutations, and combinations of algorithms have been shown to result in modest increases in the overall prediction performance (119,120). Careful benchmarking of these algorithms to ensure their optimal use is warranted.

In terms of precision medicine, we may be able to exploit the fact that cancers exhibit convergent phenotypes, as mutations in different parts of the same activated pathway lead to the net effect of dysregulating the pathway (Figure 5E). In fact, it was found that 12% to 19% of luminal breast cancers likely have mutually exclusive mutations of *MAP2K4* and *MAP3K1* in the MAPK signaling pathway, and 4%-8% of breast cancers have mutually exclusive mutations in *RUNX1* and its binding partner *CBFB* (15,16). Similarly, 33% of cases showed mutations in either *AKT1* or *PIK3CA*, both in the PI3K pathway (16). The identification of convergent phenotypes suggests that many tumors previously thought unsuitable for targeted therapies may harbor mutations that are in fact clinically actionable. In a different study, by whole-genome and whole-transcriptome sequencing, Craig et al. found that 10 of 14 metastatic, chemotherapy-resistant TNBCs had at least one alteration in the RAS/RAF/MEK/ERK or PI3K/AKT/mTOR pathways that may be actionable (121). These results suggest that even genes that are only mutated in a small proportion of breast cancer (1% to 3%) may affect components of a potentially druggable pathway; however, whether these genetic alterations would constitute actual targetable drivers of TNBCs remains to be investigated.

As discussed above, the highly recurrent driver genetic alterations found in breast cancers may have already been identified (53); however, it is plausible that bona fide driver genes may be found in the large list of genes mutated in less than 2% of breast cancers. Importantly, most of the studies have primarily analyzed IC-NSTs (ie, the common type of breast cancer) or a combination of IC-NSTs and a limited number of cases of special histologic types. At the molecular level, each special subtype appears to be more homogeneous than IC-NSTs as a whole (2,122) and the study of these rare but phenotypically homogeneous forms of breast cancer has led to the identification of pathognomonic genetic aberrations that are distinctively characteristic of and underpin these special types (Figure 5F). Adenoid cystic carcinomas, for example, account for less than 0.1% of all invasive breast cancers, have been shown to harbor the t(6;9)(q22-23;p23-24) translocation involving the genes *MYB* and *NFIB* (123,124) and secretory carcinomas, which account for 0.15% of all invasive breast carcinomas, harbor the recurrent t(12;15)(p13;q25) translocation resulting in the *ETV6-NTRK3* fusion gene in more than 90% of cases (125). These two rare types of breast cancer driven by recurrent fusion genes share additional characteristics, particularly as they display a triple-negative phenotype but have a remarkably indolent clinical course. Based on these observations, further studies investigating the constellation of somatic genetic alterations found in special histologic types of breast cancer may result in the identification of novel drivers of the disease.

Conclusions

Genomic analyses of breast cancer have reshaped our understanding of the disease and resulted in novel classification systems, which herald a new era for therapeutic options for breast cancer patients. The initial class discovery analyses in combination with TCGA and other large MPS studies have further brought the remarkable diversity of breast cancer to the forefront of cancer research. Inter- and intratumor genetic heterogeneity pose formidable challenges for the implementation of precision medicine for patients with breast cancer; however, we would contend that it is only by harnessing the complexities posed by this heterogeneity and its underlying biological causes that critical decisions about the targeted agents and combinatorial therapies will be rendered on the basis of disease biology rather than empiricism and anatomy. Germane to these endeavors is the realization

that not all information essential for the optimal matching of cancer patients with specific therapeutic agents may stem from the analysis of the genome. Additional integrative approaches employing multiple types of data, including more comprehensive analysis of the transcriptome (eg, noncoding RNAs and splice variants), epigenetic regulators of the genome (ie, capitalizing on the results of the ENCODE project [126]) and modern quantitative proteomics methods, coupled with a conceptual framework and bioinformatics and statistical methods that incorporate the intratumor genetic and phenotypic heterogeneity found in cancers, may result in fundamental discoveries and potentially their translation into benefit for cancer patients.

Funding

AMS is supported in part by a stipend from the German Cancer Aid (Dr. Mildred Scheel Stiftung).

Notes

The authors have no conflicts of interest to declare.

References

1. Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet*. 2011;378(9805):1812–1823.
2. Weigelt B, Reis-Filho JS. Histological and molecular types of breast cancer: is there a unifying taxonomy? *Nat Rev Clin Oncol*. 2009;6(12):718–730.
3. Azzopardi JG, Ahmed A, Millis RR. Problems in breast pathology. *Major Probl Pathol*. 1979;11:i-xvi, 1–466.
4. Scarff RW, Torloni H. *Histological typing of breast tumours. World Health Organization*. Geneva; 1968.
5. Lakhani SR, Ellis IO, Schnitt SJ, et al. *WHO classification of Tumours of the Breast, 4th edition*. Lyon: IARC press; 2012.
6. Colleoni M, Rotmensz N, Maisonneuve P, et al. Outcome of special types of luminal breast cancer. *Ann Oncol*. 2012;23(6):1428–1436.
7. Rakha EA, Lee AH, Evans AJ, et al. Tubular carcinoma of the breast: further evidence to support its excellent prognosis. *J Clin Oncol*. 2010;28(1):99–104.
8. Ellis IO, Galea M, Broughton N, et al. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. *Histopathology*. 1992;20(6):479–489.
9. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. 1991;19(5):403–410.
10. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–752.
11. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160–1167.
12. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530–536.
13. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999–2009.
14. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012;486(7403):346–352.
15. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70.
16. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. 2012;486(7403):405–409.
17. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012;486(7403):400–404.
18. Ellis MJ, Ding L, Shen D, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature*. 2012;486(7403):353–360.
19. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*. 2012;486(7403):395–399.
20. Rakha EA, Reis-Filho JS, Baehner F, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res*. 2010;12(4):207.
21. Wirapati P, Sotiriou C, Kunkel S, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res*. 2008;10(4):R65.
22. Prat A, Parker JS, Karginova O, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res*. 2010;12(5):R68.
23. Guedj M, Marisa L, de Reynies A, et al. A refined molecular taxonomy of breast cancer. *Oncogene*. 2012;31(9):1196–1206.
24. Doane AS, Danso M, Lal P, et al. An estrogen receptor-negative breast cancer subset characterized by a hormonally regulated transcriptional program and response to androgen. *Oncogene*. 2006;25(28):3994–4008.
25. Farmer P, Bonnefoi H, Becette V, et al. Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene*. 2005;24(29):4660–4671.
26. Lehmann BD, Bauer JA, Chen X, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest*. 2011;121(7):2750–2767.
27. Burstein MD, Tsimelzon A, Poage GM, et al. Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-negative Breast Cancer. *Clin Cancer Res*. 2014; doi: clincanres.0432.2014.
28. Lehmann BD, Pietenpol JA. Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes. *J Pathol*. 2014;232(2):142–150.
29. Masuda H, Baggerly KA, Wang Y, et al. Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. *Clin Cancer Res*. 2013;19(19):5533–5540.
30. Dawson SJ, Rueda OM, Aparicio S, et al. A new genome-driven integrated classification of breast cancer and its implications. *EMBO J*. 2013;32(5):617–628.
31. Natrajan R, Lambros MB, Geyer FC, et al. Loss of 16q in high grade breast cancer is associated with estrogen receptor status: Evidence for progression in tumors with a luminal phenotype? *Genes Chromosomes Cancer*. 2009;48(4):351–365.
32. Hicks J, Krasnitz A, Lakshmi B, et al. Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res*. 2006;16(12):1465–1479.
33. Ali HR, Rueda OM, Chin SF, et al. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. *Genome Biol*. 2014;15(8):431.
34. Weigelt B, Mackay A, A'Hern R, et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol*. 2010;11(4):339–349.
35. Mackay A, Weigelt B, Grigoriadis A, et al. Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement. *J Natl Cancer Inst*. 2011;103(8):662–673.
36. Haibe-Kains B, Desmedt C, Loi S, et al. A three-gene model to robustly identify breast cancer molecular subtypes. *J Natl Cancer Inst*. 2012;104(4):311–325.
37. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817–2826.
38. Zhang Y, Schnabel CA, Schroeder BE, et al. Breast cancer index identifies early-stage estrogen receptor-positive breast cancer patients at risk for early- and late-distant recurrence. *Clin Cancer Res*. 2013;19(15):4196–4205.
39. Filipits M, Rudas M, Jakesz R, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. *Clin Cancer Res*. 2011;17(18):6012–6020.
40. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med*. 2006;355(6):560–569.
41. Kelly CM, Bernard PS, Krishnamurthy S, et al. Agreement in risk prediction between the 21-gene recurrence score assay (Oncotype DX(R)) and the PAM50 breast cancer intrinsic Classifier in early-stage estrogen receptor-positive breast cancer. *Oncologist*. 2012;17(4):492–498.
42. Teschendorff AE, Miremadi A, Pinder SE, et al. An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol*. 2007;8(8):R157.
43. Desmedt C, Haibe-Kains B, Wirapati P, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res*. 2008;14(16):5158–5165.
44. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*. 2006;7:96.
45. Teschendorff AE, Caldas C. A robust classifier of high predictive value to identify good prognosis patients in ER-negative breast cancer. *Breast Cancer Res*. 2008;10(4):R73.
46. Yau C, Esserman L, Moore DH, et al. A multigene predictor of metastatic outcome in early stage hormone receptor-negative and triple-negative breast cancer. *Breast Cancer Res*. 2010;12(5):R85.
47. Kuo WH, Chang YY, Lai LC, et al. Molecular characteristics and metastasis predictor genes of triple-negative breast cancer: a clinical study of triple-negative breast carcinomas. *PLoS One*. 2012;7(9):e45831.
48. Rody A, Karn T, Liedtke C, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast Cancer Res*. 2011;13(5):R97.
49. Azim HA Jr, Michiels S, Zagouri F, et al. Utility of prognostic genomic tests in breast cancer practice: The IMPAKT 2012 Working Group Consensus Statement. *Ann Oncol*. 2013;24(3):647–654.
50. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol*. 2013;24(9):2206–2223.

51. Natrajan R, Mackay A, Lambros MB, et al. A whole-genome massively parallel sequencing analysis of BRCA1 mutant oestrogen receptor-negative and -positive breast cancers. *J Pathol*. 2012;227(1):29–41.
52. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013;502(7471):333–339.
53. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505(7484):495–501.
54. Kan Z, Jaiswal BS, Stinson J, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature*. 2010;466(7308):869–873.
55. Dees ND, Zhang Q, Kandoth C, et al. MuSiC: identifying mutational significance in cancer genomes. *Genome Res*. 2012;22(8):1589–1598.
56. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214–218.
57. Youn A, Simon R. Identifying cancer driver genes in tumor genome sequencing studies. *Bioinformatics*. 2011;27(2):175–181.
58. Bose R, Kavuri SM, Searleman AC, et al. Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer Discov*. 2013;3(2):224–237.
59. Greulich H, Kaplan B, Mertins P, et al. Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of ERBB2. *Proc Natl Acad Sci U S A*. 2012;109(36):14476–14481.
60. Toy W, Shen Y, Won H, et al. ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet*. 2013;45(12):1439–1445.
61. Robinson DR, Wu YM, Vats P, et al. Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat Genet*. 2013;45(12):1446–1451.
62. Jeselsohn R, Yelensky R, Buchwalter G, et al. Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin Cancer Res*. 2014;20(7):1757–1767.
63. Li S, Shen D, Shao J, et al. Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep*. 2013;4(6):1116–1130.
64. Weigelt B, Eberle C, Cowell CF, et al. Metaplastic breast carcinoma: more than a special type. *Nature Rev Cancer*. 2014;14:147–148.
65. Geyer FC, Weigelt B, Natrajan R, et al. Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas. *J Pathol*. 2010;220(5):562–573.
66. Ding L, Ellis MJ, Li S, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*. 2010;464(7291):999–1005.
67. Navin N, Kendall J, Troge J, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011;472(7341):90–94.
68. Shah SP, Morin RD, Khattra J, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*. 2009;461(7265):809–813.
69. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012;366(10):883–892.
70. Nik-Zainal S, Van Loo P, Wedge DC, et al. The life history of 21 breast cancers. *Cell*. 2012;149(5):994–1007.
71. Wang Y, Waters J, Leung ML, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*. 2014;512(7513):155–160.
72. Stratton MR. Exploring the genomes of cancer cells: progress and promise. *Science*. 2011;331(6024):1553–1558.
73. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976;194(4260):23–28.
74. Cleary AS, Leonard TL, Gestl SA, et al. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature*. 2014;508(7494):113–117.
75. Marusyk A, Tabassum DP, Altrock PM, et al. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*. 2014;514(7520):54–58.
76. Seol H, Lee HJ, Choi Y, et al. Intratumoral heterogeneity of HER2 gene amplification in breast cancer: its clinicopathological significance. *Mod Pathol*. 2012;25(7):938–948.
77. Cottu PH, Asselah J, Lae M, et al. Intratumoral heterogeneity of HER2/neu expression and its consequences for the management of advanced breast cancer. *Ann Oncol*. 2008;19(3):595–597.
78. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta*. 2010;1805(1):105–117.
79. Turner NC, Reis-Filho JS. Genetic heterogeneity and cancer drug resistance. *Lancet Oncol*. 2012;13(4):e178–e185.
80. Weigelt B, Reis-Filho JS. Epistatic interactions and drug response. *J Pathol*. 2014;232(2):255–263.
81. De Mattos-Arruda L, Weigelt B, Cortes J, et al. Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann Oncol*. 2014;25(9):1729–1735.
82. Klein CA. Selection and adaptation during metastatic cancer progression. *Nature*. 2013;501(7467):365–372.
83. Almendro V, Kim HJ, Cheng YK, et al. Genetic and phenotypic diversity in breast tumor metastases. *Cancer Res*. 2014;74(5):1338–1348.
84. Bidard FC, Ng CK, Piscuoglio S, et al. Abstract S6-06: High-depth massively parallel sequencing reveals heterogeneity between primary tumor and metastatic deposits in de novo metastatic breast cancer patients prior to exposure to systemic therapy. *Cancer Res*. 2013;73(24 Suppl):S6–06.
85. Bachelot T, Lefebvre C, Campone M, et al. Abstract S6-07: Genomic characterisation of metastatic samples from breast cancer patients using next generation sequencing. *Cancer Res*. 2013;73(24 Suppl):S6–07.
86. Ciriello G, Miller ML, Aksoy BA, et al. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet*. 2013;45(10):1127–1133.
87. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature*. 1998;396(6712):643–649.
88. Youn A, Simon R. Using passenger mutations to estimate the timing of driver mutations and identify mutator alterations. *BMC Bioinformatics*. 2013;14:363.
89. Natrajan R, Lambros MB, Rodriguez-Pinilla SM, et al. Tiling path genomic profiling of grade 3 invasive ductal breast cancers. *Clin Cancer Res*. 2009;15(8):2711–2722.
90. Chin SF, Teschendorff AE, Marioni JC, et al. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol*. 2007;8(10):R215.
91. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell*. 2006;10(6):529–541.
92. Burns MB, Lackey L, Carpenter MA, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature*. 2013;494(7437):366–370.
93. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet*. 2013;45(9):977–983.
94. Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012;149(5):979–993.
95. Roberts SA, Lawrence MS, Klimczak LJ, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet*. 2013;45(9):970–976.
96. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415–421.
97. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. 2013;45(10):1134–1140.
98. Stephens PJ, Greenman CD, Fu B, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*. 2011;144(1):27–40.
99. Korbel JO, Campbell PJ. Criteria for inference of chromothripsis in cancer genomes. *Cell*. 2013;152(6):1226–1236.
100. Baca SC, Prandi D, Lawrence MS, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013;153(3):666–677.
101. Navin N, Krasnitz A, Rodgers L, et al. Inferring tumor progression from genomic heterogeneity. *Genome Res*. 2010;20(1):68–80.
102. Heinrich MC, Maki RG, Corless CL, et al. Primary and secondary kinase genotypes correlate with the biological and clinical activity of sunitinib in imatinib-resistant gastrointestinal stromal tumor. *J Clin Oncol*. 2008;26(33):5352–5359.
103. Sequist LV, Waltman BA, Dias-Santagata D, et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci Transl Med*. 2011;3(75):75ra26.
104. Maley CC, Galipeau PC, Finley JC, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet*. 2006;38(4):468–473.
105. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol*. 2013;49(3):211–215.
106. Alix-Panabieres C, Pantel K. Circulating tumor cells: liquid biopsy of cancer. *Clin Chem*. 2013;59(1):110–118.
107. Yu M, Stott S, Toner M, et al. Circulating tumor cells: approaches to isolation and characterization. *J Cell Biol*. 2011;192(3):373–382.
108. Bidard FC, Weigelt B, Reis-Filho JS. Going with the flow: from circulating tumor cells to DNA. *Sci Transl Med*. 2013;5(207):207ps14.
109. De Mattos-Arruda L, Cortes J, Santarpia L, et al. Circulating tumour cells and cell-free DNA as tools for managing breast cancer. *Nat Rev Clin Oncol*. 2013;10(7):377–389.
110. Bidard FC, Peeters DJ, Fehm T, et al. Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. *Lancet Oncol*. 2014;15(4):406–414.
111. Lohr JG, Adalsteinsson VA, Cibulskis K, et al. Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat Biotechnol*. 2014;32(5):479–484.
112. Dawson SJ, Tsui DW, Murtaza M, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med*. 2013;368(13):1199–1209.
113. Bettegowda C, Sausen M, Leary RJ, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 2014;6(224):224ra24.
114. Murtaza M, Dawson SJ, Tsui DW, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*. 2013;497(7447):108–112.
115. Carter H, Chen S, Isik L, et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res*. 2009;69(16):6660–6667.
116. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 2011;39(17):e118.

117. Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310–315.
118. Shihab HA, Gough J, Cooper DN, et al. Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics*. 2013;29(12):1504–1510.
119. Gnad F, Baucom A, Mukhyala K, et al. Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics*. 2013;14(Suppl 3):S7.
120. Martelotto LG, Ng C, De Filippo MR. et al. Benchmarking mutation effect prediction algorithms using functionally validated cancer-related missense mutations. *Genome Biol*. 2014;15(10):484.
121. Craig DW, O'Shaughnessy JA, Kiefer JA, et al. Genome and transcriptome sequencing in prospective metastatic triple-negative breast cancer uncovers therapeutic vulnerabilities. *Mol Cancer Ther*. 2013;12(1):104–116.
122. Weigelt B, Horlings HM, Kreike B, et al. Refinement of breast cancer classification by molecular characterization of histological special types. *J Pathol*. 2008;216(2):141–150.
123. Persson M, Andren Y, Mark J, et al. Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. *Proc Natl Acad Sci U S A*. 2009;106(44):18740–18744.
124. Wetterskog D, Lopez-Garcia MA, Lambros MB, et al. Adenoid cystic carcinomas constitute a genomically distinct subgroup of triple-negative and basal-like breast cancers. *J Pathol*. 2012;226(1):84–96.
125. Tognon C, Knezevich SR, Huntsman D, et al. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell*. 2002;2(5):367–376.
126. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.