Insight into the heterogeneity of breast cancer through next-generation sequencing

Hege G. Russnes,[1,2,3,4] Nicholas Navin,[5,6] James Hicks,[7] and Anne-Lise Borresen-Dale[2,3]

[1]Laboratory of Molecular Pathology, Division of Pathology, and [2]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. [3]Institute for Clinical Medicine, University of Oslo, Oslo, Norway. [4]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA. [5]Department of Genetics and [6]Department of Bioinformatics and Computational Biology, MD Anderson Cancer Center, Houston, Texas, USA. [7]Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, USA.

Rapid and sophisticated improvements in molecular analysis have allowed us to sequence whole human genomes as well as cancer genomes, and the findings suggest that we may be approaching the ability to individualize the diagnosis and treatment of cancer. This paradigmatic shift in approach will require clinicians and researchers to overcome several challenges including the huge spectrum of tumor types within a given cancer, as well as the cell-to-cell variations observed within tumors. This review discusses how next-generation sequencing of breast cancer genomes already reveals insight into tumor heterogeneity and how it can contribute to future breast cancer classification and management.

Introduction

Next-generation sequencing. DNA sequencing using dideoxynucleotide termination chemistry was first described by Fred Sanger in the 1970s and subsequently automated by capillary sequencing by Applied Biosystems in the 1990s. However, this "first-generation" sequencing method was limited to sequencing targeted regions of DNA spanning approximately 700 nucleotides at a time. This brute-force method was the workhorse of the Human Genome Project, which sequenced all 3.2 billion bp at high coverage over a period of 10 years. Today, advanced systems such as the Illumina HiSeq 2000 are capable of sequencing a human genome at 30× coverage in about one week. These next-generation sequencing (NGS) systems use massively parallel sequencing to generate hundreds of millions of short (36- to 150-bp) DNA reads that can be aligned to the human genome. In addition to using single-end reads, it is now possible to sequence both ends of DNA library molecules (paired ends) to identify discordant pairs that represent deletions, amplifications, inversions, or translocations (see recent reviews for more details about the technology; refs. 1, 2). Thus, in addition to identifying point mutations, this strategy provides a wealth of information about a range of genetic aberrations that can occur in a cancer genome, including copy number variation. Although a number of different highly parallel NGS strategies have been developed, the paired-end strategy from Illumina Inc. has become the tool of choice for most cancer genome studies published to date. While most cancer genome studies so far have focused on single patients, this pattern is likely to change as a result of ongoing international collaborations (3) and steep decreases in the cost of sequencing. The hope is that NGS data will shorten the road to personalized medicine, in which treatments and therapies are tailored to target the unique spectrum of mutations that define individual tumors and tumor subpopulations (4–8). However, this challenge, which has been referred to as "the $1,000 genomes, the $100,000 analysis" problem (9), will only continue to grow.

Tumor heterogeneity. Variation between patients is often referred to as intertumor heterogeneity and is classically recognized through different morphology types, expression subtypes, or classes of genomic copy number patterns, among other differences. Variation within a single tumor, intratumor heterogeneity, has long been observed by histopathologists as sectors of different morphology or staining behavior and has more recently been defined at the molecular level by the genetic differences observed in tumor subpopulations and even among individual malignant cells. As we begin to enter the era of whole-genome DNA sequencing, a wealth of data is starting to emerge shedding light on the evolution of cancer (2, 10). However, high-resolution DNA sequence data are currently available for only a handful of cases, and at the present level of technology, incorporation of whole-genome sequencing into clinical trials is problematic. The challenge is to make the best use of the great body of knowledge that has been gained using lower-resolution methods on thousands of cases to direct the NGS studies to have the greatest impact on clinical management of the disease.

Intertumor heterogeneity

Subgrouping tumors by phenotype. The reference book from the WHO groups breast cancer into 17 different types based on their microscopic appearance, but the clinical impact of this classification is debated and has little impact on clinical decision making (11, 12). The so-called histological "special" subtypes have distinct molecular alterations and clinical behavior, but more than 70% of breast carcinomas are categorized as invasive ductal carcinoma, not otherwise specified (IDC NOS), tumors that do not display sufficient characteristics of any of the special types (11–13). Still, information about the microscopic features of the tumor and its cells are very informative; the histological grading system developed more than 50 years ago, which is used to estimate the level of differentiation, number of mitoses, and nuclear pleomorphism, is of major impact in clinical decision making today (6, 14).

Since the introduction of microarray analyses, the rapid advancement in technology and data handling has yielded an enormous increase in our knowledge about the molecular disruptions in cancer cells. As the variations from case to case became evident, two main approaches were applied to explore the clinical utility of such

---
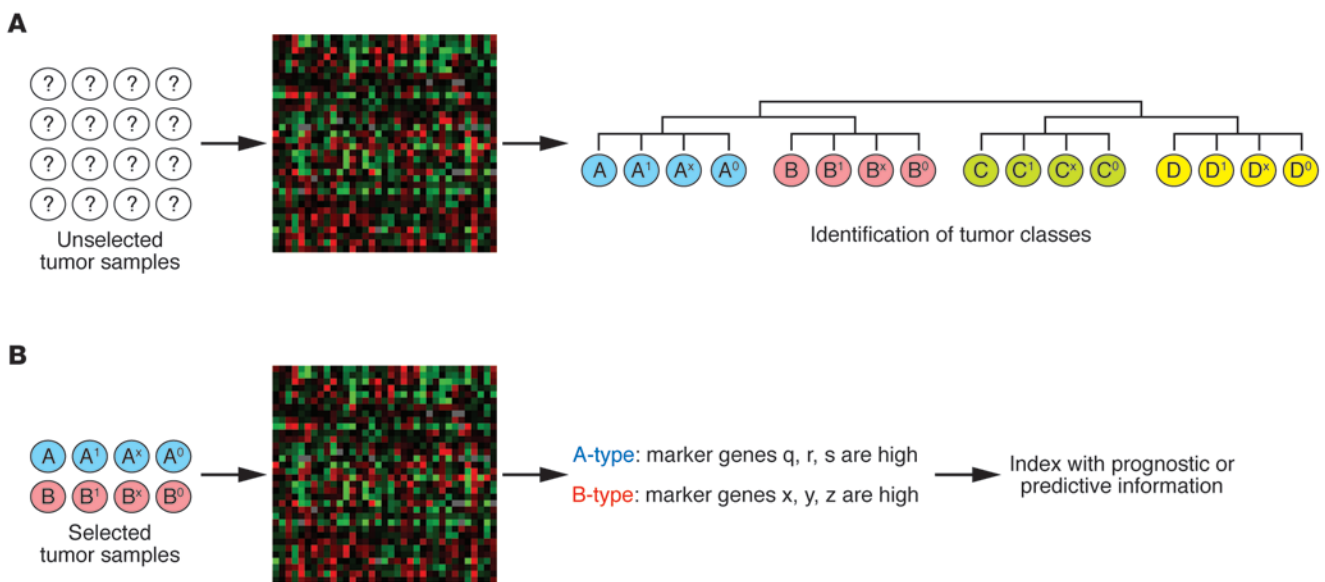
Figure 1
Different study designs for array-based gene expression studies. (**A**) Studies aimed at identifying different subgroups investigate a mixed population of patients to group tumors with similar alterations together, and markers that recognize each type can then be identified. (**B**) This in contrast to studies that search for markers for prediction of therapy response or outcome; here, selected groups of patients are analyzed to identify the most discriminating alterations.

information (Figure 1). Some studies were designed to search for a taxonomy that could define distinct subtypes of breast carcinomas, while others were designed to elucidate specific alterations of predictive or prognostic value. It is important to recognize the difference in design of these two approaches. The first approach investigates an unsorted population of cases and aims to group tumors by common alterations (Figure 1A). The latter interrogates predefined groups of tumors such as clinical trial cohorts and looks for biomarkers that can predict a given clinical parameter, such as outcome or response to therapy (Figure 1B). Some of the most highly cited gene expression microarray studies were aimed at identifying alterations that predict prognosis (15, 16) or propensity to metastasize (17). While such studies do not reveal fundamental differences among tumors, others, such as the study by Sorlie et al. (18), were able to group tumors based on alterations in expression of a set of predefined "intrinsic genes" (19). In these studies five main intrinsic subtypes were identified: luminal A, luminal B, basal-like, human epidermal growth factor receptor 2–related (HER2-related), and normal-like, each with a different prognosis and distributions of known or promising targets for therapy. The introduction of this classification system into clinical use has not been easy, as robust single-sample predictors are needed (20–23). Identifying molecular classes in samples from breast cancer patients is dependent on the composition of the given cohort; the initial study of Perou et al. was performed on fewer than 40 carcinomas, the majority of which were of the IDC NOS class (18). Distinct tumor types that were not included in this study might not be appreciated at all and thus not recognized in validation studies. With increased knowledge from thousands of investigated breast carcinomas, combined with technical advancements including more sophisticated bioinformatics tools, a variety of more or less related classifiers are being recognized, each developed from different data sets on different gene expression platforms (24–27). As an example, one of the newer

variants of the "intrinsic classification," the PAM50 assay, derived its classification algorithm and parameter values from an independent training set and was designed for independent single-sample classification (27, 28). Validation of PAM50 as a classifier is ongoing, and a gold standard for molecular classification by gene expression is not yet available. Other groups have shown that by grouping breast cancer by the expression of the established markers estrogen receptor (ER), progesterone receptor (PR), and HER2 (also known as ERBB2), distinct molecular alterations and outcomes can be identified (29). The same markers are frequently used as surrogates for the intrinsic subtypes — luminal A subtype is defined as ER+ and/or PR+, HER2-; luminal B subtype is defined as ER+ and/or PR+, HER2+; HER2-related subtype is defined as ER-, PR-, HER2+, and Basal-like tumors are defined as ER-, PR-, and HER2- (30).

Genomic alterations are linked to phenotype. A few decades ago, cytogenetic studies showed that breast cancer had different types of alterations; analysis of near-diploid tumors identified the most common alteration as a translocation resulting in a der(1;16)(10q;10p) and formation of isochromosome 1q (31–34). Other tumors had multiple rearrangements affecting a multitude of chromosomes indicating that the heterogeneity of breast cancer is also present at the genetic level even in an early phase of tumor progression (33). The introduction of microarray analyses to assess genomic copy number variation (array comparative genomic hybridization [aCGH] and SNP arrays) gave increased resolution and more precise quantification, but physical rearrangements in the genome could not be assessed by these methods (35–38). Despite this limitation, studies of aCGH data analyzed without prior knowledge of molecular subtype showed that breast cancer could be divided into groups based on the architecture of the genomic alterations, probably reflecting different types of genomic instability (35, 37). Three main patterns were recognized: tumors with few rearrangements (dominated by gain of chromosome 1q and/or loss of 16q), tumors with complex
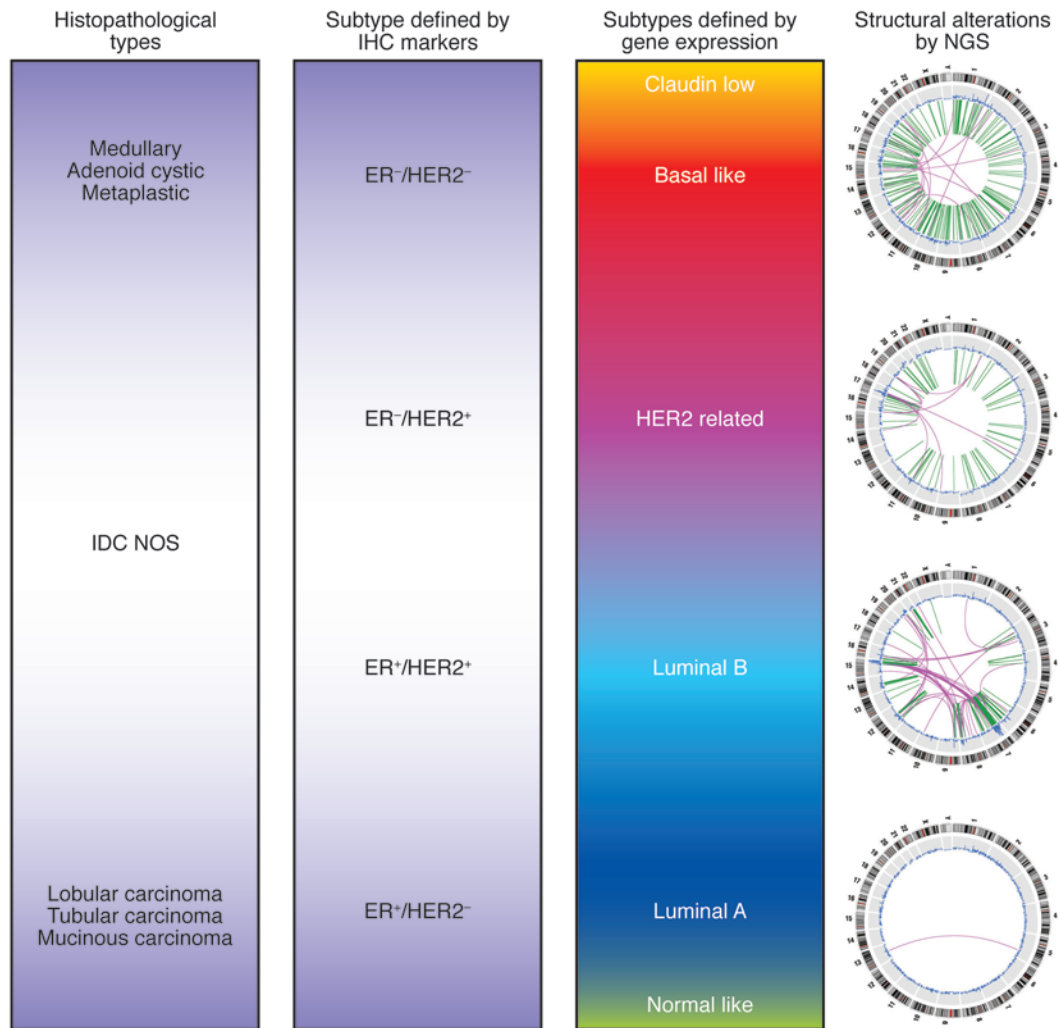
Figure 2
Subtypes of breast cancer. Hypothetically, subtypes of breast cancer can be viewed as a spectrum of more or less related entities. The majority are classified through histopathology as IDC NOS, but some types have defined histopathological traits. Such groups have tumors that are frequently either ER⁻/HER2⁻ or ER⁺/HER2⁻, which also corresponds to the outer part of a spectrum of intrinsic subtypes, namely the basal-like and luminal A types of breast cancer. NGS of a basal-like (top), a HER2-related (second from top), a luminal B (third from top), and a luminal A tumor (bottom) show distinct structural characteristics. The circos plots show intrachromosomal rearrangements in green and interchromosomal rearrangements in purple (circos plots used with permission from *Nature*; ref. 43).

alterations, and tumors with tightly packed, high-level amplicons. Although such patterns of alterations can be objectively quantified and have prognostic information (39), in analyzing these patterns, knowledge about outcome of rearrangements such as fusion genes or disrupted genomic elements is lost. Advances in NGS have enabled researchers to characterize the full spectrum of mutations in a limited number of breast cancer genomes including the architectural pattern (40–45). These studies often reveal that tumor genomes are littered with diverse types of mutations — segmental duplications, amplifications and deletions, translocations, inversions, small insertion-deletions, and point mutations. One of the first large-scale sequencing studies of primary breast tumors and breast cancer cell lines by Stephens and coworkers revealed different types of structural alterations consistent with those deduced from aCGH (43). Three major patterns were seen: (a) few, interchromosomal translocations with copy number alterations involving large DNA fragments or whole chromosome arms; (b) complex, interchromosomal translocations affecting shorter regions with high-level amplifications; and (c) small, intrachromosomal segmental alterations such as duplications, deletions and/or inversions, termed the "mutator phenotype". Moreover, the subtypes exhibited distinctly different microhomologies at translocation breakpoints, making it a reasonable hypothesis that the structural rearrangements are caused by subtype-specific mechanisms. Consistent with this notion, analyses of other cancers have shown that the type and distribution of rearrangement patterns characteristically vary among diseases (40, 42, 46).

A relationship between these classifications and the gene expression subtypes has been described. A luminal A tumor was found to have few chromosomal rearrangements by aCGH and had only one translocation by paired-end sequencing, in contrast to a basal-like tumor, which had a complex aCGH profile and a typical mutator
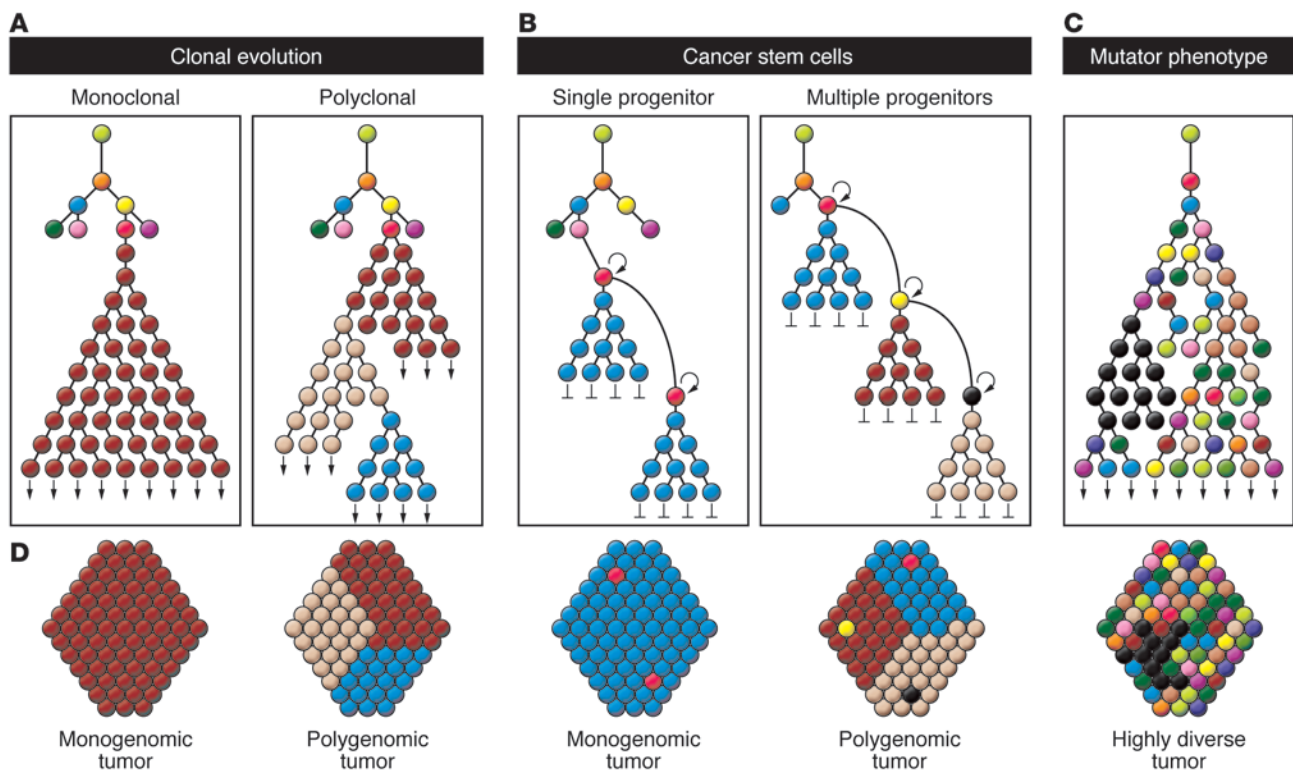
Figure 3
Hypothetical models explaining intratumor heterogeneity. (**A–C**) Different models of tumor progression can give rise to distinct types of intratumor heterogeneity, exemplified here by the clonal evolution (**A**), the cancer stem cell (**B**), and the mutator phenotype (**C**) models. (**D**) The different models can result in distinct spatial distributions of subpopulations.

phenotype pattern detected by paired-end sequencing (39). Interestingly, one luminal B and one HER2-enriched tumor were found to have high-level amplifications, but the latter also had complex alterations by aCGH analysis (39). Lobular carcinomas are known to frequently be of the normal-like or luminal A type (47) and may have no or few structural alterations (43, 45). This is in contrast to a basal-like tumor analyzed by Ding et al. that, with a corresponding brain metastasis xenograft, revealed multiple translocations and segmental rearrangements in all three tumors from this patient (44).

Breast cancer subtypes defined by different methods share some overlapping molecular features. As illustrated in Figure 2, histopathological types such as lobular and medullary carcinomas correlate with ER⁺/HER2⁻ and ER⁻/HER2⁻ receptor status respectively as well as with luminal A and basal-like expression subtypes. Recent NGS studies have revealed that these major expression subtypes display different classes of mutations, and it will be of considerable interest to determine if NGS data can define additional subtypes of breast carcinomas.

Nonrecurrent mutations in cancer genomes. Perhaps the biggest surprise of detailed sequencing studies of cancer has been the failure to identify recurrent mutations in cancer genes when mutational profiles are compared from patient to patient (1, 43, 48). The picture emerging is that individual tumors are unique, each harboring large numbers of "private" mutations that uniquely characterize its genome. Even when mutations occur in the same cancer genes, they often occur in different codons or protein domains, revealing an element of randomness in their genesis. A recent large-scale study of 50 luminal A breast cancer genomes sequenced at high

coverage (30×) identified over 1,700 genic mutations, but only 3 of these genes were mutated at frequencies that approached or exceeded 10%: *PIK3CA* (43%), *TP53* (15.2%), and *MAP3K1* (9.3%) (49). When stratified by expression subtypes, it was reported that mutation of *TP53* is more frequent in basal-like and HER2-enriched disease, while mutation of *PIK3CA* is found to be over-represented in luminal A tumors (25, 50–52). The spectrum of mutations found by NGS seems also to differ; even though the predominant type of point mutations were CG-to-TA transitions in both the basal-like tumor and the lobular tumor sequenced, only the former had CG-to-AT transversions (44). This shows the importance of taking intertumor heterogeneity into account when designing experiments to detect novel mutations. And, as the vast majority of somatic mutations occur at very low frequencies in cancer genomes, it raises the question of whether common signaling pathways, rather than individual genes, are mutated. This question will become more addressable as we sequence more cancer genomes, and large-scale international sequencing projects will certainly shed light on this question (3, 53).

Intratumor heterogeneity

Subpopulations within tumors. In addition to the vast heterogeneity among breast tumors, many studies have reported extensive genomic diversity within tumors. As early as the 1800s, scientists such as the revered Rudolf Virchow, recorded the morphological heterogeneity of malignant cells within individual tumors (reviewed in ref. 54). The development and progress in cell-staining methods subsequently enabled pathologists to characterize tumor
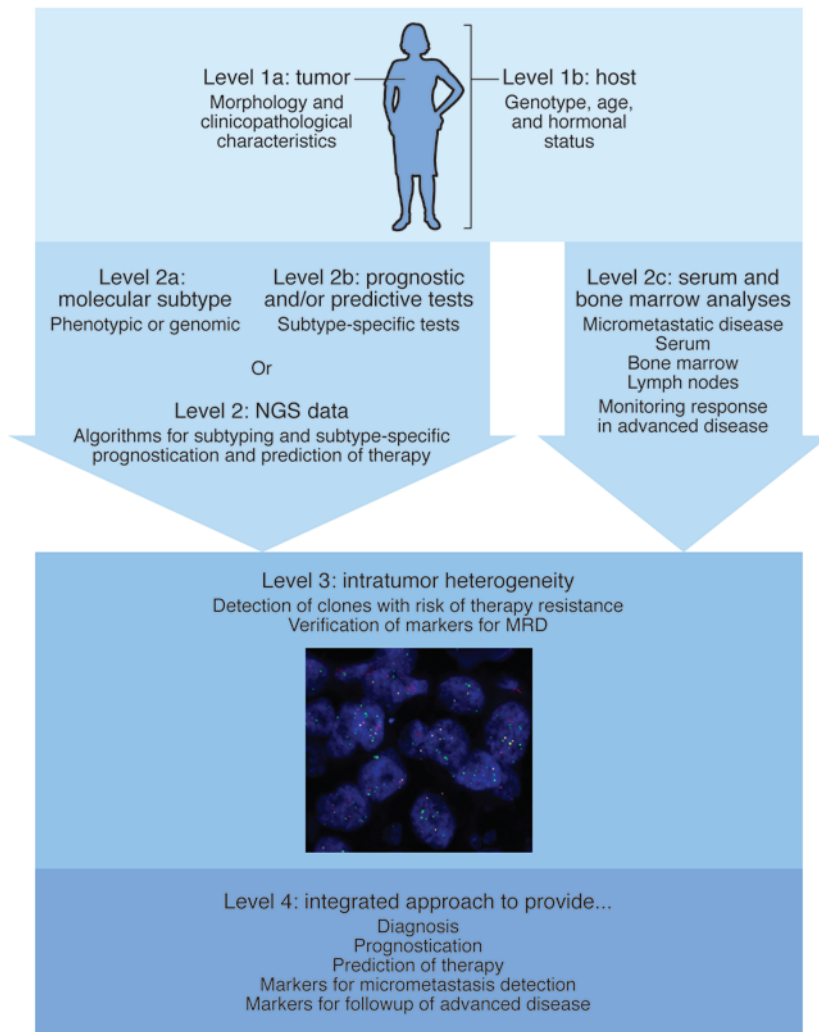
Figure 4
A multilevel approach for a dynamic classification system. The first level is defined by tumor and patient characteristics. The second level includes detailed genomic and translational analyses of tumor to define molecular type and selection of appropriate tests. Parallel to that, tumor-specific serum markers can be assessed. The third level determines intratumor heterogeneity and is crucial for selection of appropriate markers for micrometastatic disease detection in serum, bone marrow, or lymph nodes. MRD, minimal residual disease. The fourth level integrates all available information to produce a diagnosis, prognostication, prediction of therapy, and program for disease monitoring.

cells using different morphological parameters, including nuclear pleomorphism, number of mitoses, and differentiated structures, the basis for the histological grading system. However, this system is challenging due to the morphological heterogeneity of malignant cells within some tumors (55–58). In fact, pathologists are well aware of this phenomenon and will examine tissue sections from different regions of the same tumor, reporting the highest grade observed (59, 60). Giemsa staining, spectral karyotyping, and FISH enabled biologists to directly visualize chromosomal aberrations in individual tumor cells. Results from such studies clearly show that breast tumors commonly exhibit genetic heterogeneity at preferred loci including duplications, deletions, and distinctive chromosomal rearrangements (61–69).

Microarray technologies have made it possible to conduct genome-wide measurements of gene expression and chromosome copy number in tumors, providing quantitative data that can be subjected to statistical analysis (70, 71). However, the aCGH methods used until recently have required larger quantities of input DNA, and thus their signals were limited to averaging copy number signal over populations of tumor cells, leukocytes, and stromal tissue. Efforts have been made to isolate and compare genetically distinct subpopulations prior to array; we have used

regional macro-dissection of tumors to show that genetically defined subpopulations could be found in geographically distinct sectors of the tumors, and further analyzed subfractions by using FACS to sort cells by DNA content (72). Others have employed flow sorting based on surface markers to separate phenotypically distinct subpopulations for genomic analysis (73, 74).

Intratumor heterogeneity inferred from NGS data. The pioneering NGS studies of breast cancer patient samples and cell lines have provided an excellent overview of spectra of mutations, but they cannot resolve the combinations of mutations present in any given subpopulation from a heterogeneous tumor. Despite this, deep sequencing of bulk tumors provides a major advantage over microarray methods for studying tumor heterogeneity, since sequencing can measure the distribution of allele frequencies in a population of cells. This feature was particularly useful in the study by Ding et al. in analyzing the metastatic progression of a basal-like breast cancer to the brain (44). In this study, roughly the same set of 50 coding mutations was observed in the primary tumor and the metastasis. Few de novo mutations were seen in the metastasis; however, gross changes in allelic frequencies of these mutations were observed, suggesting that minor subpopulations of cells with metastatic potential were pre-existing in the primary tumor. In addition, the study by Shah et al. revealed allelic variation, indicat-

ing intratumor variation at the genomic level (45). Such studies would benefit greatly by first isolating tumor subpopulations by macro-dissection, DNA ploidy, laser capture microdissection, or cell surface receptors prior to deep sequencing, or — even better — by sequencing the genomes of single tumor cells.

Genomic heterogeneity at single-cell resolution. Intratumor heterogeneity studies profiling or sequencing DNA from individual tumor cells require whole-genome amplification (WGA). By using commercially available methods for WGA, it is possible to amplify DNA from a single cell to a level where it can be profiled by microarrays, but these studies have been challenged by technical difficulty and limited reproducibility (75–78). Analyzing WGA fragments from single cells using targeted approaches such as DNA microarrays is problematic because fragments are randomly amplified from a small fraction (<10%) of the genome, and thus many fail to hybridize to their target probes. An alternative approach is to measure the randomly amplified WGA fragments from single cells using NGS, which has the advantage of providing a non-targeted approach. In a recent study combining flow sorting, WGA, and NGS, in an approach called single-nucleus sequencing (SNS), genomic copy number profiles of single cells were quantified at high resolution (50 kb) (79). The SNS strategy involves sparsely sequencing (0.1× coverage) the genome of a single cell and measuring copy number from sequence read depth. By binning intervals across the genome, counting the number of sequence reads, segmenting the data, and sampling copy number states, the authors showed that genomic copy number profile of a single cell could be quantified at high resolution (50 kb). By comparing multiple single-cell copy number profiles, they could provide highly accurate measures of genomic heterogeneity within solid tumors. Furthermore, by comparing multiple single-cell profiles, they showed that it was possible to reconstruct the evolutionary lineage of a tumor and understand its pattern of progression. In this study, 100 single cells were profiled from a triple-negative heterogeneous breast tumor, in addition to 100 single cells from a homogeneous primary breast tumor and its paired liver metastasis. This analysis revealed a punctuated model of clonal evolution, in which tumors evolve by one or more sequential clonal expansions with few gradual intermediates, challenging the paradigm of evolution through the gradual accumulation of mutations over a long period of time (79). In future studies it will be of significant interest to correlate tumor heterogeneity, as measured by SNS, with overall survival of the patient and response to chemotherapy. These single-cell genomic methods are likely to have additional clinical value in the early detection of tumor cells or tumor DNA in scarce clinical samples (urine, blood, fine-needle aspirates) and monitoring of circulating tumor cells after remission.

Causations of intertumor and intratumor heterogeneity

The causations of both inter- and intratumor heterogeneity in breast cancer is debated (80), partly because knowledge about the hierarchical relationship between different epithelial cells in the normal breast is still at the hypothetical stage, but also because cell-of-origin and tumor progression paths of breast cancers are not yet defined.

Two hypothetical models explaining intertumor heterogeneity are frequently proposed (recently reviewed by Visvader; ref. 81). The genetic model points to the same cell of origin but different initiating events that will lead to different molecular subtypes. The other model points to each subtype having different cells of origin. It is also acknowledged that a combinatory model might be plausible as well, in which not only different cells of origin but also dif-

ferent initial events can explain the diversity in molecular subtypes (82). The differences in the genome and the transcriptome between luminal A and basal-like tumors indicate that these diseases have very distinct pathogenesis. Several studies have also shown that genome-wide patterns of DNA methylation differ between luminal and basal-like tumors, with similarities to CD24+ cells (luminal cells) and CD44+ cells (progenitor cells) (83–86). Although it is tempting to speculate that this is related to cell of origin, recent work has pointed to luminal progenitors as the cell of origin for both basal-like and luminal tumors (87, 88).

Tumor progression is an important basis to explain intratumor heterogeneity, and different models are plausible (89, 90). The clonal evolution model originally proposed by Nowell in 1976 suggests that tumors evolve by the expansion of one (monoclonal) or multiple (polyclonal) subpopulations to form the tumor mass (Figure 3A and ref. 91). In this egalitarian model, all clones have the potential for continued proliferation and Darwinian selection. In contrast, the cancer stem cell model suggests a hierarchical organization in which tumor heterogeneity is explained by several rare precursor cells, each giving rise to a different subpopulation within the tumor (Figure 3B). Another model for tumor progression, the mutator hypothesis, suggests that tumors evolve by the gradual and random accumulation of mutations as the tumor grows (Figure 3C), which suggests a vast degree of diversity in the tumor rather than clonal subpopulations (92). As illustrated in Figure 3D, different progression models can result in distinct spatial distribution of subpopulations, but whether such patterns are subtype specific is still unknown.

Clinical decision making in the era of NGS

Prognostication and prediction of drug response. While prognostic markers aim at identifying patients with a probability of having a better or worse outcome, markers with predictive potential can also be used for selection of patients with high probability of response to a given drug or treatment. Prediction is important both to spare non-responders from the side effects that come along with treatment and to minimize the overall cost by only treating those that have a good chance to respond. Today the most established markers in breast cancer are ER and HER2 status. Both markers have prognostic and predictive information and are themselves the targets for therapy. While these markers have a clear utility, it is important to acknowledge that tumors display a wide range of expression of each; for instance, ER+ patients can have anywhere from 1% to 100% of the tumor cells scored positive. Heterogeneity among ER+ tumors is also evident by microarray analyses at the genomic, transcriptomic, and epigenetic level (19, 37, 84), and this was addressed by the so-called "recurrence score"/OncotypeDx, which estimates the expression level of 21 genes (16 tumor genes and 5 reference genes) stratifying ER+ breast cancer into groups with high or low risk of distant recurrences (93). The work by Desmedt and colleagues demonstrated that different clinical variables had prognostic value only for subsets of the patients, again illustrating the importance of selecting subgroups of patients to identify prognostic or predictive markers or profiles (94). Discoveries of mutated genes have led to development of several targeted therapies (95), and future research must further these efforts.

Intratumor heterogeneity adds a second level of complexity; an optimal diagnostic test will need to identify even minor subpopulations of cells with alterations related to increased aggressiveness or therapy resistance. Some subtypes seem to have greater intratumoral hetero-

geneity than others (69), but the clinical importance of each given subpopulation is not yet clear. Intratumor heterogeneity is likely to play an important role in responsiveness to chemotherapy (5, 96, 97), and results from adjuvant-treated ovary, cervical, and tongue cancer suggest that resistant tumor subpopulations pre-exist and expand after treatment (98–100). The study of Jones et al. illustrates how such information can be used to determine the next level of treatment in an adjuvant setting (100). Still more studies using in situ–based techniques are needed to fully appreciate the intratumor heterogeneity and its prognostic and predictive impact.

Monitoring disease progression; detection of minimal residual disease. The detection of circulating tumor cells in blood or of disseminated tumor cells in bone marrow has a prognostic impact in breast cancer in general and in luminal disease in particular (101–103). Two studies have shown how sequencing of breast cancer genomes and subsequent design of patient-specific probes for nested real-time PCR can be used to detect tumor DNA in serum at the time of relapse (104, 105) and to monitor the effect of treatment (104). NGS has also been used to detect naked DNA in serum in order to enable early diagnosis of breast cancer; Beck and colleagues were able to identify the prevalence of tumor DNA in serum from breast cancer patients compared with controls (106). This study detected a differential representation of some repetitive DNA elements and was thus not patient specific. Knowing a patient's tumor genome or transcriptome down to single base level at time of diagnosis can provide tailored markers for follow-up. While still at an early stage, these results nevertheless show the potential power of exploiting genomic rearrangements in fluids to measure subclinical disease at time of diagnosis and after treatment.

Toward an integrated classification? A multi-level classification based on combining both clinical and molecular information (Figure 4) could be very useful to tailor therapy and disease monitoring. The first level will be tumor-specific information revealed by pathology as well as clinical information about the given patient. At the second level, assessment of the molecular subtype by phenotype and/or genotype will be followed by subtype-specific prognostic and predictive tests. This would be an efficient and practical approach; tests that have more clinical value for some subtypes than others should be restricted to validated groups. This second level should indeed be dynamic; we expect new subtype definitions, development of novel diagnostic tests, and changes in technology. As most tests today are only validated for a given platform, a panel of tests might be required (illustrated in Figure 4), but we envision that NGS might be able to overcome this obstacle, since information from NGS might reveal the subtype as well as provide the data for subtype-specific algorithms, allowing prognosis and prediction of therapy responsiveness.

A separate level will be the assessment of intratumor heterogeneity; although in situ techniques are the most usable at the moment, this might be solved by NGS of subpopulations or single cells in the future. Finally, observed alterations in the tumor cells (even in smaller fractions) can be used as individual markers for follow-up analyses in blood and/or bone marrow.

For such a classification to be meaningful, an approach is needed in which each patient has a combination of parameters outlined and uniformed into both a prognostic and a predictive index for a given therapy; these will direct the clinician to select optimal therapy and design follow-up and provide markers for monitoring the disease at the molecular level (Figure 4). It is important to acknowledge that prospective studies are always needed, but a combinatory model could be built over time, including validated parameters.

Although such a detailed classification system will be challenging to manage, it seems evident that the minor tumor subgroups will require such a tailored scheme.

Perspectives

The improvement in sequencing technology that has made possible even whole-genome sequencing of single cells has already given novel insight into breast cancer heterogeneity. The technique is still challenging because deep sequencing requires large numbers of cells or genome-wide amplification of single cells. Mapping and exclusion of artifacts is challenging, and the development of methods for data interpretation is still in an early phase. In spite of this, the first studies of a handful of breast carcinomas by NGS have revealed exciting new insights into genomic variability. First, selected tumors known to belong to different molecular subtypes have different rearrangement patterns and frequencies of mutations (43). Second, alterations in metastases are present in primary tumors, but the latter show a wider distribution of additional changes (44, 45). Third, analysis of single cells from the same tumor reveal different clonal relationships, supporting the notion that tumor progression can follow distinctly different pathways (79). From these few studies it is tempting to conclude that the detailed genomic knowledge allowed by NGS can provide markers for individualized disease monitoring.

From a clinical point of view, the insight from sequencing of more tumor genomes will provide a step toward defining more robust subsets of breast cancer types. Another step will be more detailed knowledge about aberrant translation that alters the tumor proteome or introduces dysfunctions in the epigenome. Taken together this can be the fundament of dynamic molecular classification guiding therapy choices and disease monitoring as well as being adjustable for prediction of novel therapies.

Many questions remain unresolved. The challenges in analyzing such huge amounts of data are not yet overcome; we do not have a nomenclature suitable for classifying all types of genomic alterations revealed and integrative approaches needed for robust classification are still in an early phase. Exploring intratumor heterogeneity is challenging, but microdissection or sampling of tumors as viable tumor cells will enable analysis of subpopulations or cells. This, together with sequencing of circulating or disseminating cells, is needed to elucidate the impact of intratumor heterogeneity and micrometastatic disease on clinical outcome. Still, the fast pace in technical advancements in NGS combined with reduced costs and increased availability should be encouraging for projects focusing on these issues.

1. Mardis ER. A decade's perspective on DNA sequencing technology. *Nature*. 2011;470(7333):198–203.
2. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet*. 2010;11(10):685–696.
3. Hudson TJ, et al. International network of cancer genome projects. *Nature*. 2010;464(7291):993–998.
4. Swanton C, Caldas C. Molecular classification of solid tumours: towards pathway-driven therapeutics. *Br J Cancer*. 2009;100(10):1517–1522.
5. Oakman C, Santarpia L, Di Leo A. Breast cancer assessment tools and optimizing adjuvant therapy. *Nat Rev Clin Oncol*. 2010;7(12):725–732.
6. Rakha EA, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res*. 2010;12(4):207.
7. Rakha EA, Reis-Filho JS, Ellis IO. Combinatorial biomarker expression in breast cancer. *Breast Cancer Res Treat*. 2010;120(2):293–308.
8. Perez EA. Breast cancer management: opportunities and barriers to an individualized approach. *Oncologist*. 2011;16 suppl 1:20–22.
9. Mardis ER. The $1,000 genome, the $100,000 analysis? *Genome Med*. 2010;2(11):84.
10. Pfeifer GP, Hainaut P. Next-generation sequencing: emerging lessons on the origins of human cancer. *Curr Opin Oncol*. 2011;23(1):62–68.
11. Elston CW, Ellis IO, Pinder SE. Prognostic factors in invasive carcinoma of the breast. *Clin Oncol (R Coll Radiol)*. 1998;10(1):14–17.
12. The International Agency for Research on Cancer. World Health Organization: Tumors Of The Breast And Female Genital Organs (Who/iarc Classification Of Tumors). Tavassoéli FA, Devilee P, eds. Washington, DC, USA: IARC Press WHO; 2003.
13. Weigelt B, et al. Refinement of breast cancer classification by molecular characterization of histological special types. *J Pathol*. 2008;216(2):141–150.
14. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer*. 1957;11(3):359–377.
15. van 't Veer LJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530–536.
16. van de Vijver MJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999–2009.
17. Wang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005;365(9460):671–679.
18. Sorlie T, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869–10874.
19. Perou CM, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747–752.
20. Dunning MJ, Curtis C, Barbosa-Morais NL, Caldas C, Tavare S, Lynch AG. The importance of platform annotation in interpreting microarray data. *Lancet Oncol*. 2010;11(8):717.
21. Perou CM, Parker JS, Prat A, Ellis MJ, Bernard PS. Clinical implementation of the intrinsic subtypes of breast cancer. *Lancet Oncol*. 2010;11(8):718–719.
22. Sorlie T, et al. The importance of gene–centring microarray data. *Lancet Oncol*. 2010;11(8):719–720.
23. Weigelt B, et al. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *Lancet Oncol*. 2010;11(4):339–349.
24. Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer*. 2007;7(7):545–553.
25. Hennessy BT, et al. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res*. 2009;69(10):4116–4124.
26. Gatza ML, et al. A pathway-based classification of human breast cancer. *Proc Natl Acad Sci U S A*. 2010;107(15):6994–6999.
27. Parker JS, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160–1167.
28. Nielsen TO, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res*. 2010;16(21):5222–5232.
29. Wirapati P, 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.
30. Nielsen TO, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res*. 2004;10(16):5367–5374.
31. Flagiello D, Gerbault-Seureau M, Sastre-Garau X, Padoy E, Vielh P, Dutrillaux B. Highly recurrent der(1;16)(q10;p10) and other 16q arm alterations in lobular breast cancer. *Genes Chromosomes Cancer*. 1998;23(4):300–306.
32. Tsarouha H, Pandis N, Bardi G, Teixeira MR, Andersen JA, Heim S. Karyotypic evolution in breast carcinomas with i(1)(q10) and der(1;16)(q10;p10) as the primary chromosome abnormality. *Cancer Genet Cytogenet*. 1999;113(2):156–161.
33. Gerbault-Seureau M, Vielh P, Zafrani B, Salmon R, Dutrillaux B. Cytogenetic study of twelve human near-diploid breast cancers with chromosomal changes. *Ann Genet*. 1987;30(3):138–145.
34. Pandis N, et al. Chromosome analysis of 97 primary breast carcinomas: identification of eight karyotypic subgroups. *Genes Chromosomes Cancer*. 1995;12(3):173–185.
35. Hicks J, et al. Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res*. 2006;16(12):1465–1479.
36. Chin SF, et al. Using array-comparative genomic hybridization to define molecular portraits of primary breast cancers. *Oncogene*. 2007;26(13):1959–1970.
37. Chin K, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell*. 2006;10(6):529–541.
38. Bergamaschi A, et al. Extracellular matrix signature identifies breast cancer subgroups with different clinical outcome. *J Pathol*. 2008;214(3):357–367.
39. Russnes HG, et al. Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci Transl Med*. 2010;2(38):38ra47.
40. Campbell PJ, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet*. 2008;40(6):722–729.
41. Hampton OA, et al. A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome. *Genome Res*. 2009;19(2):167–177.
42. Stephens PJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*. 2011;144(1):27–40.
43. Stephens PJ, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature*. 2009;462(7276):1005–1010.
44. Ding L, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*. 2010;464(7291):999–1005.
45. Shah SP, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*. 2009;461(7265):809–813.
46. Berger MF, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011;470(7333):214–220.
47. Zhao H, et al. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol Biol Cell*. 2004;15(6):2523–2536.
48. Kan Z, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature*. 2010;466(7308):869–873.
49. Ellis MJ, et al. Analysis of luminal-type breast cancer by massively parallel sequencing. Presented Saturday, April 2, 2011 at the 102nd Annual Meeting of the American Association for Cancer Research in Orlando, Florida, USA.
50. Sorlie T, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A*. 2003;100(14):8418–8423.
51. Samuels Y, Velculescu VE. Oncogenic mutations of PIK3CA in human cancers. *Cell Cycle*. 2004;3(10):1221–1224.
52. Langerod A, et al. TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Res*. 2007;9(3):R30.
53. Stratton MR. Exploring the genomes of cancer cells: progress and promise. *Science*. 2011;331(6024):1553–1558.
54. Brown TM, Fee E. Rudolf Carl Virchow: medical scientist, social reformer, role model. *Am J Public Health*. 2006;96(12):2104–2105.
55. Fitzgerald PJ. Homogeneity and heterogeneity in pancreas cancer: presence of predominant and minor morphological types and implications. *Int J Pancreatol*. 1986;1(2):91–94.
56. Hirsch FR, Ottesen G, Podenphant J, Olsen J. Tumor heterogeneity in lung cancer based on light microscopic features. A retrospective study of a consecutive series of 200 patients, treated surgically. *Virchows Arch A Pathol Anat Histopathol*. 1983;402(2):147–153.
57. Kruger S, Thorns C, Bohle A, Feller AC. Prognostic significance of a grading system considering tumor heterogeneity in muscle-invasive urothelial carcinoma of the urinary bladder. *Int Urol Nephrol*. 2003;35(2):169–173.
58. van der Poel HG, Oosterhof GO, Schaafsma HE, Debruyne FM, Schalken JA. Intratumoral nuclear morphologic heterogeneity in prostate cancer. *Urology*. 1997;49(4):652–657.
59. Ignatiadis M, Sotiriou C. Understanding the molecular basis of histologic grade. *Pathobiology*. 2008;75(2):104–111.
60. Komaki K, Sano N, Tangoku A. Problems in histological grading of malignancy and its clinical significance in patients with operable breast cancer. *Breast Cancer*. 2006;13(3):249–253.
61. Farabegoli F, Santini D, Ceccarelli C, Taffurelli M, Marrano D, Baldini N. Clone heterogeneity in diploid and aneuploid breast carcinomas as detected by FISH. *Cytometry*. 2001;46(1):50–56.
62. Maley CC, et al. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat Genet*. 2006;38(4):468–473.
63. Mora J, Cheung NK, Gerald WL. Genetic heterogeneity and clonal evolution in neuroblastoma. *Br J Cancer*. 2001;85(2):182–189.
64. Pantou D, et al. Cytogenetic manifestations of multiple myeloma heterogeneity. *Genes Chromosomes Cancer*. 2005;42(1):44–57.
65. Roka S, et al. Aneuploidy of chromosome 8 as detected by interphase fluorescence in situ hybridization is a recurrent finding in primary and metastatic breast cancer. *Breast Cancer Res Treat*. 1998;48(2):125–133.
66. Sauter G, Moch H, Gasser TC, Mihatsch MJ, Waldman FM. Heterogeneity of chromosome 17 and erbB-2 gene copy number in primary and metastatic bladder cancer. *Cytometry*. 1995;21(1):40–46.
67. Teixeira MR, Pandis N, Bardi G, Andersen JA, Heim S. Karyotypic comparisons of multiple tumorous and macroscopically normal surrounding tissue samples from patients with breast cancer. *Cancer Res*. 1996;56(4):855–859.
68. Teixeira MR, Pandis N, Bardi G, Andersen JA, Mitelman F, Heim S. Clonal heterogeneity in breast cancer: karyotypic comparisons of multiple intra-

and extra-tumorous samples from 3 patients. *Int J Cancer*. 1995;63(1):63–68.

69. Park SY, Lee HE, Li H, Shipitsin M, Gelman R, Polyak K. Heterogeneity for stem cell-related markers according to tumor subtype and histologic stage in breast cancer. *Clin Cancer Res*. 2010;16(3):876–887.

70. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*. 1995;270(5235):467–470.

71. Snijders AM, et al. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet*. 2001;29(3):263–264.

72. Navin N, et al. Inferring tumor progression from genomic heterogeneity. *Genome Res*. 2010; 20(1):68–80.

73. Shipitsin M, et al. Molecular definition of breast tumor heterogeneity. *Cancer Cell*. 2007; 11(3):259–273.

74. Keller PJ, et al. Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. *Breast Cancer Res*. 2010;12(5):R87.

75. Geigl JB, et al. Identification of small gains and losses in single cells after whole genome amplification on tiling oligo arrays. *Nucleic Acids Res*. 2009;37(15):e105.

76. Fuhrmann C, et al. High-resolution array comparative genomic hybridization of single micrometastatic tumor cells. *Nucleic Acids Res*. 2008;36(7):e39.

77. Fiegler H, et al. High resolution array-CGH analysis of single cells. *Nucleic Acids Res*. 2007;35(3):e15.

78. Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmuller G. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci U S A*. 1999;96(8):4494–4499.

79. Navin N, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011;472(7341):90–94.

80. Campbell LL, Polyak K. Breast tumor heterogeneity: cancer stem cells or clonal evolution? *Cell Cycle*. 2007;6(19):2332–2338.

81. Visvader JE. Cells of origin in cancer. *Nature*. 2011;469(7330):314–322.

82. Polyak K. Breast cancer: origins and evolution. *J Clin Invest*. 2007;117(11):3155–3163.

83. Bloushtain-Qimron N, et al. Cell type-specific DNA methylation patterns in the human breast. *Proc Natl Acad Sci U S A*. 2008;105(37):14076–14081.

84. Holm K, et al. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res*. 2010;12(3):R36.

85. Kamalakaran S, et al. DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables. *Mol Oncol*. 2011;5(1):77–92.

86. Ronneberg JA, et al. Methylation profiling with a panel of cancer related genes: Association with estrogen receptor, TP53 mutation status and expression subtypes in sporadic breast cancer. *Mol Oncol*. 2011;5(1):61–76.

87. Lim E, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med*. 2009;15(8):907–913.

88. Molyneux G, et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell*. 2010;7(3):403–417.

89. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta*. 2010;1805(1):105–117.

90. Navin NE, Hicks J. Tracing the tumor lineage. *Mol Oncol*. 2010;4(3):267–283.

91. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976;194(4260):23–28.

92. Loeb LA, Springgate CF, Battula N. Errors in DNA replication as a basis of malignant changes. *Cancer Res*. 1974;34(9):2311–2321.

93. Paik S, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817–2826.

94. Desmedt C, et al. Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res*. 2008;

14(16):5158–5165.

95. Stuart D, Sellers WR. Linking somatic genetic alterations in cancer to therapeutics. *Curr Opin Cell Biol*. 2009;21(2):304–310.

96. Gerlinger M, Swanton C. How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine. *Br J Cancer*. 2010;103(8):1139–1143.

97. Merlo LM, et al. A comprehensive survey of clonal diversity measures in Barrett's esophagus as biomarkers of progression to esophageal adenocarcinoma. *Cancer Prev Res (Phila)*. 2010;3(11):1388–1397.

98. Cooke SL, et al. Genomic analysis of genetic heterogeneity and evolution in high-grade serous ovarian carcinoma. *Oncogene*. 2010;29(35):4905–4913.

99. Cooke SL, et al. Intra-tumour genetic heterogeneity and poor chemoradiotherapy response in cervical cancer. *Br J Cancer*. 2011;104(2):361–368.

100. Jones SJ, et al. Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biol*. 2010;11(8):R82.

101. Braun S, et al. A pooled analysis of bone marrow micrometastasis in breast cancer. *N Engl J Med*. 2005;353(8):793–802.

102. Naume B, et al. Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer. *Mol Oncol*. 2007;1(2):160–171.

103. Wiedswang G, et al. Detection of isolated tumor cells in bone marrow is an independent prognostic factor in breast cancer. *J Clin Oncol*. 2003;21(18):3469–3478.

104. Leary RJ, et al. Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med*. 2010;2(20):20ra14.

105. McBride DJ, et al. Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer*. 2010;49(11):1062–1069.

106. Beck J, Urnovitz HB, Mitchell WM, Schutz E. Next generation sequencing of serum circulating nucleic acids from patients with invasive ductal breast cancer reveals differences to healthy and nonmalignant controls. *Mol Cancer Res*. 2010;8(3):335–342.