# Translating Genomic Research into Clinical Practice: Promise and Pitfalls

*Joseph A. Sparano, MD, Harry Ostrer, MD, and Paraic A. Kenny, PhD*

## OVERVIEW

**Breast cancer is a heterogeneous disease associated with variable clinical outcomes despite standard local therapy for the primary tumor and systemic adjuvant therapy to prevent distant recurrence. Management decisions are typically made using classical prognostic and predictive clinicopathologic factors, and more recently gene expression profiling assays are commonly used in practice. Recent advances in genomic sequencing—often referred to collectively as next-generation sequencing (NGS)—have facilitated more in-depth evaluation of the cancer genome than could be afforded by the initial generation of gene expression studies, including DNA single nucleotide variants, small insertions and deletions, structural alterations, and copy number alterations (CNAs). In addition, this information has been integrated with other molecular profiling methods of processes that affect gene transcription (e.g., epigenetic, microRNA) and protein expression—the ultimate readout of the genetic code. Although NGS has provided new insights on the classification of breast cancer and identified potential predictive biomarkers and novel targets, there are formidable logistical and scientific obstacles that must be addressed before the promise of this technology is fully realized.**

Genomics is defined as the study of all of the nucleotide sequences in an organism. The original sequencing methods relied on Sanger sequencing using synthesis with DNA polymerase and termination with dideoxynucleotides described in 1975, then modified in 1977 to be more rapid and accurate.[1,2] These methods were used to sequence the first human genomes that were reported in 2001.[3,4] Over the past decade, newer methods allow sequencing to be done more quickly, accurately, and cheaply, which are often referred to as NGS.[5] Such assays are now being used in the clinic not only for research but also commercially for clinical use. Techniques are also available that allow high-throughput evaluation of the epigenome, microRNAs, and proteins and analytic approaches that integrate information from multiple profiling methods. This review will address how NGS is being used for discovery-based research and its potential clinical application in breast cancer.

## FIRST GENERATION: GENE EXPRESSION PROFILING

Until recently genomic profiling focused on the evaluation of gene expression or the translation of information encoded in genomic DNA into an RNA transcript.[6] RNA transcripts include mRNAs that are translated into proteins and various other RNAs (e.g., transfer RNA, ribosomal RNA, micro RNA, noncoding RNA) that have important biologic functions. Perou et al first identified "intrinsic" breast cancer subtypes by evaluating variation in gene expression patterns using hierarchical clustering in a set of 65 breast cancers from 42 individuals using complementary DNA microarrays representing 8,102 human genes.[7] The subtypes were recapitulated in other datasets and shown to be clinically relevant with distinct clinical outcomes.[8] It was hypothesized that these subtypes had distinctive gene expression profiles because they originated from different cell types, including luminal epithelial cells (the cells that line the duct and give rise to the majority of breast cancers) and basal epithelial cells of the normal mammary gland (characterized by expression of cytokeratins 5/6 and 17), hence the terms "luminal" and "basal" subtypes. More recent data indicate that the basal tumors arise from luminal progenitor cells rather than the basal myoepithelial cells of the mammary gland.[9,10] The intrinsic gene panel was subsequently reduced to a panel of 50 genes (called the PAM50) detectable by qRT-PCR, with 10 genes selected for each centroid used to define four intrinsic subtypes, including luminal A, luminal B, basal, and HER2-enriched (plus a "normal" subtype, which reflects an inadequate biopsy specimen containing predominantly normal breast tissue).[11] The PAM50 and several multiparameter gene expression assays have been approved for clinical use, including some that have been recommended by expert panels for clinical decision making.[12]

From the Departments of Medicine, Obstetrics, Gynecology and Women's Health, Pathology, Genetics, Pediatrics, and Developmental & Molecular Biology, Albert Einstein College of Medicine, Bronx, NY; Montefiore Medical Center, Bronx, NY.

Authors' disclosures of potential conflicts of interest are found at the end of this article.

Corresponding author: Joseph A. Sparano, MD, Department of Oncology, Montefiore Medical Center, Albert Einstein College of Medicine, 1825 Eastchester Road, Bronx, NY 10461; email: jsparano@montefiore.org.

© 2013 by American Society of Clinical Oncology.

## NEXT GENERATION: BEYOND GENE EXPRESSION

NGS relies on the use of high-throughput, massively parallel sequencing and bioinformatic approaches to analyze massive datasets, including a variety of approaches for genome sequencing and gene expression. When applied to RNA (RNA-Seq), NGS not only provides absolute expression levels (as the precise number of transcripts can be counted) but also allows the identification of alternatively spliced isoforms, mutant transcripts, and novel transcripts arising from fusion genes. Terms commonly used in NGS studies are summarized in Table 1, and the processes typically used are illustrated in Fig. 1.

The technology platforms for NGS are developing at a rapid rate, which is inexorably leading to faster, cheaper, and more accurate sequencing data. Currently, the field is dominated by machines from Roche, Illumina (HiSeq2000), and ABI (SOLiD).[5] Each of these platforms is capable of performing whole genome sequencing, whole exome sequencing, RNA-Seq, and methylation analysis on a time scale between approximately 10 hours (Roche) and 11 days (Illumina). These platforms are being supplemented by smaller machines from the same manufacturers that may be run more quickly but handle smaller amounts of sequencing (i.e., not sufficient for whole genome analysis). Of the possible approaches, whole genome sequencing that sequences all of the DNA base pairs in the genome is the most intensive. Whole exome sequencing, in which libraries are generated from the transcribed exons of the genome (encoding proteins, microRNAs, and other RNAs), offers a less intensive approach, which still likely captures the majority of the interpretable information. RNA-Seq analysis sequences cDNA copies of the sample RNA component, providing detailed information on gene expression levels, splicing, and mutations. DNA methylation can also be analyzed using Methyl-Seq, providing detailed coverage of the distribution of methylated CpG islands throughout the genome, which play key roles in controlling gene expression.

Although the precise technical details of conducting these assays vary between platforms, the principles are quite similar in each case. Short, single-stranded DNA sequences are generated from the starting material, which can be genomic DNA (whole genome), hybridization-captured DNA exons (exome sequencing), cDNA (RNA-Seq) or bisulfite-treated DNA (Methyl-Seq). Small DNA adapters are ligated to each strand of the library, and the samples are then amplified to provide enough representations of each individual strand for sequencing. Amplification can be performed on beads (Illumina and SOLiD) or on a glass slide (Illumina). Each strand can then be sequenced by sequential addition of the four DNA bases (A, C, G, and T).

In all cases, the output of the experiment is an extremely large amount of raw DNA sequence reads (of between 60 and 400 bp, depending on the platform), which must be processed further to be interpreted. This computationally intensive step requires mapping of each individual read to the correct location on the reference genome. A variety of commercial and open-source software packages are available to perform this function and subsequent steps of the analysis. In the case of cancer, the goal is usually to identify sequence variants between the tumor sample and the reference genome or nontumor DNA from the same patient, if available. Comparing normal and tumor DNA from the same individual is particularly advantageous, as it excludes the influence of the large number of single nucleotide polymorphisms, which differ between individuals in a species. Sequence alterations—which can include point mutations, insertions, deletions, or translocations—can be identified using the analysis packages. Comparison of differences in mutation patterns, amplification/deletion, and translocation in individual tumor samples may be visually represented by Circos plots.[13]

Because tumors have a high mutation rate, any sequencing experiment will identify a large number of sequence variations—only some of which make a contribution to the disease pathology. Distinguishing causative (driver) from random (passenger) mutations can be challenging for variants and genes that have not been previously reported as playing a role in cancer. For such novel mutations, some insight can be gained from software tools such as SIFT, which can predict whether a particular change in a protein sequence is likely to have a function-altering effect. Both the sequencing technology platforms and the relevant analysis strategies have been described in detail elsewhere.[5]

Whole genome and exome sequencing and the analysis of the data generated are important research tools, but the sheer volume of data generated makes it difficult to efficiently deploy these approaches in a routine clinical setting. Instead, it is likely that targeted resequencing of small portions of the genome, known to be frequently altered in cancer is likely to prove more useful. Approaches such as AmpliSeq (Life Technologies), which uses ultra-high multiplex PCR to amplify all of the exons of 400 cancer-relevant genes from small amounts of starting material, will generate clinically actionable data in the near future.

### KEY POINTS

- Next-generation sequencing (NGS) refers to not only evaluation of gene expression, but also DNA single nucleotide variants, small insertions and deletions, structural alterations, and copy number alterations.
- When applied to RNA, NGS provides information on expression levels, alternatively spliced isoforms, mutant transcripts, and novel transcripts arising from fusion genes.
- NGS has facilitated more refined disease classification and target discovery.
- *PIK3CA* and *TP53* are the most commonly mutated genes that vary in frequency by intrinsic subtype.
- Clinical application of NGS for therapeutic targeting is currently limited by a dearth of available drugs targeting actionable alterations.

**TABLE 1. Glossary of Terms Commonly Used in Next-Generation Sequencing Studies (Listed Alphabetically)**

| Term | Definition |
|---|---|
| Circos plot | A circular ideogram figure that visually represents differences in genomic structure between individual specimens |
| Codon | A series of three adjacent bases in one polynucleotide chain of a DNA or RNA molecule, which codes for a specific amino acid |
| Copy number alterations | Alterations in gene copy number, which are typically somatically acquired |
| Copy number variants | Germ-line variation in the number of copies of a particular gene, which varies between individuals |
| Exome | The part of the genome formed by exons; the exome of the human genome consists of roughly 180,000 exons constituting approximately 1% of the total genome or about 30 megabases of DNA |
| Exon | Any nucleotide sequence encoded by a gene that remains present within the final mature RNA product of that gene after introns have been removed by RNA splicing; these include exons that are translated into protein and the untranslated region flanking them, as well as exons encoding microRNAs and noncoding RNAs. |
| Indel | Mutation resulting in a co-localized insertion or deletion and a net gain or loss in nucleotides |
| Intron | Any nucleotide sequence within DNA or RNA that is not encoded into protein; introns are removed by RNA splicing while the final mature RNA product of a gene is being generated |
| Library | A collection of DNA fragments prepared from a more complex sample |
| RNA-Seq | Use of high-throughput sequencing technologies to sequence cDNA to get information about a sample's RNA content |
| Single nucleotide variant | DNA sequence variation when a single nucleotide (A, C, T, G) differs between members of a biological species |

## NGS IN BREAST CANCER

All cancers carry somatic mutations in their genome that are dominated by point mutations, amplifications, translocations, and complex rearrangements.[14] The genomic landscape of breast cancer is characterized by a handful of commonly mutated gene "mountains" and a much larger number of gene "hills" that are mutated at a low frequency.[15] Driver mutations have been defined in at least 40 genes in breast cancer, and somatic genomic rearrangements resulting in oncogenic fusion transcripts are also common—some of which have been postulated to play a key role in disease progression.[16-18] The results of selected NGS studies in breast cancer are summarized in Table 2 and described below, which includes studies correlating genomics with clinical outcomes for the purpose of classification or prediction, with specific breast cancer phenotypes for target discovery, or with genomics integrated with epigenomic, microRNA, and proteomic data to gain a deeper understanding of how specific genes may be regulated and influence protein expression.

## USE OF NGS FOR DISEASE CLASSIFICATION: METABRIC

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) study was a joint effort by research teams from Canada and the United Kingdom to genomically classify breast tumors. Approximately 2,000 clinically annotated, fresh-frozen breast cancer specimens from patients with operable breast cancer were assembled from tumor banks in the United Kingdom and Canada; this included 997 tumors analyzed in a discovery group and 995 in a validation group.[19] Nearly all patients with estrogen receptor (ER)-positive and/or lymph node–negative disease did not receive adjuvant chemotherapy, whereas patients with either lymph node–positive or ER-negative disease did. Matched DNA and RNA were extracted from each specimen and subject to copy number and genotype analysis on the Affymetrix SNP 6.0 platform, transcriptional profiling on the Illumina HT-12 v3 platform, and *TP53* mutations by Sanger sequencing in a cohort of 820 patients. This allowed evaluation of somatically acquired CNAs, germline copy number variants, and whether genomic variants acted in cis (impacts its own expression) or in trans (impacts expression of other genes in genome, defined as outside at least a contiguous three megabase window).

There were several important observations in this study. First, a number of known driver mutations (e.g., *ZNF704, PTEN, MYC, CCND1, MDM2, ERBB2, CCNE1*) and putative driver mutations (e.g., *MDM1, MDM4, CDK3, CDK4, CAMK1D, PI4KB, NCOR1*) were identified. There were also important deletions identified, including known deletions (e.g., *PTEN*) and novel deletions (e.g., *PPP2R2A, MTAP*, and *MAP2K4*). *PPP2R2A* deletions were noted in luminal B cancers and have likewise been reported in clear cell and ovarian and endometrioid cancers. *MTAP* is often codeleted with *CDKN2A* and *CDKN2B* tumor suppressor genes in a variety of cancers. *MAP2K4* mutations were noted in ER-positive tumors and felt to be consistent with a tumor suppressor gene. A deletion event on chromosome 5 was also noted in basal-like tumors. Second, the effect of specific genes acting in trans by influencing expression of distant sites of the genome identified known aberration hot spots that could be grouped into pathway modules (e.g., *ERBB2, MYC*) and novel loci, including T-cell receptor loci on chromosome 7 (*TRG*) and 14 (*TRA*). These upregulate mRNAs and are highly enriched from T-cell activation and proliferation, dendritic cell presentation, and leukocyte activation, thereby indicating an adaptive immune response associated with tumor-infiltrating lymphocytes. Third, joint clustering of copy number and gene expression data revealed that 10 integrative clusters characterized by well-defined copy number aberrations that split many of the intrinsic subtypes were associated with variable clinical outcomes and were

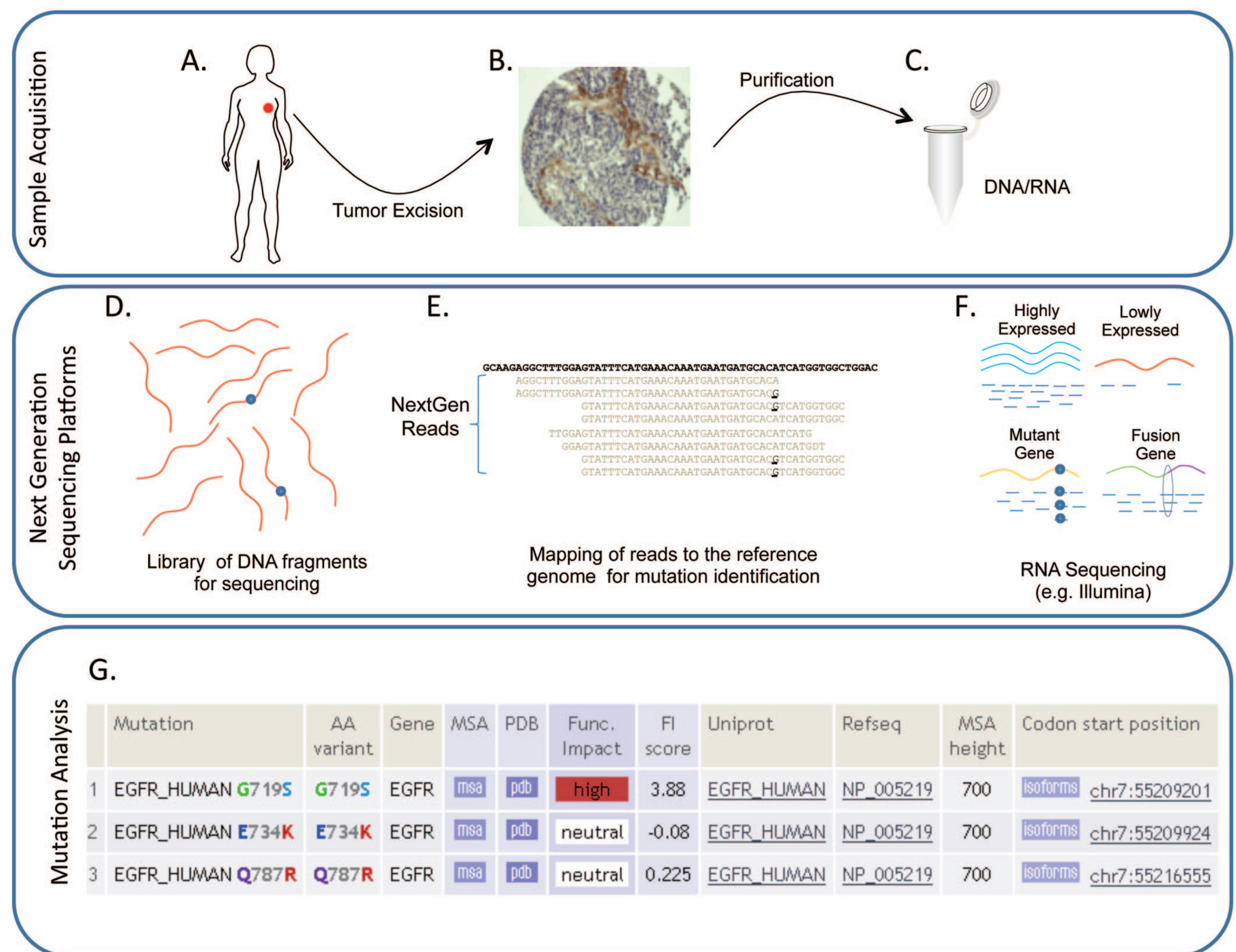


**FIG 1. Summary of the steps involved in various next-generation sequencing approaches. (A-C) A tumor sample is acquired and processed to purify either DNA (for genome/exome/methylation analysis) or RNA (for gene expression analysis. (D) A library of short DNA fragments is prepared from the sample. For RNA-Seq analysis, the RNA from the sample is converted to cDNA. Some fragments may contain mutations (circles). (E) Sequencing reads are mapped to the reference human genome. Here, part of the reference sequence of the *PIK3CA* gene is shown, with several aligned reads. Some reads have an A to G point mutation at position 3140 of the *PIK3CA* coding sequence. This mutant encodes the oncogenic *H1047R PIK3CA* mutation. (F) RNA-Seq analysis reads are mapped similarly to E. This approach provides data on gene expression level as well as gene mutations and fusions. (G) Sample output from mutationassessor.org showing an analysis of three epidermal growth factor receptor mutations. Two of the mutations are predicted to have a neutral effect on protein function, while the *G719S* mutation is predicted to have a high effect.**

consistent in discovery and validation sets. This indicates heterogeneity even within "intrinsic subtypes" revealed by integration of information regarding somatic CNAs, including subgroups of ER-positive tumors that were high risk (characterized by 11q13/14 cis-acting alterations), favorable risk (characterized by a paucity of copy number and cis-acting alterations), and intermediate risk (characterized by 17q23/20q and 8p12 cis-acting subgroups). In addition, this report demonstrated for the first time that genomic copy number loss at T-cell receptor loci drives a trans-acting immune response in an otherwise genomically quiescent subgroup of ER-positive and ER-negative tumors associated with a good prognosis, whereas copy number loss at 5q in a group of basal-like cancers drives a trans-acting transcriptional control of genomic and chromosomal instability associated with a poor prognosis.

## USE OF NGS TO IDENTIFY PREDICTORS OF SENSITIVITY AND RESISTANCE TO ENDOCRINE THERAPY

Aromatase inhibitors (AIs) are commonly used as adjuvant endocrine therapy in postmenopausal women with ER-

**TABLE 2. Summary of Selected Next-Generation Sequencing Studies in Breast Cancer**

| Reference | No. Patients | Patient Population | Clinical Outcomes | Profiling Methods | Key Findings |
|---|---|---|---|---|---|
| **Curtis et al[22]** | 1,992 | All types | Yes | CNV | 10 subtypes identified that correlate with clinical outcomes |
| | | | | CNA | |
| | | | | Gene transcription | Deletions in *PPR2A, MTAP, MAP2K4* common |
| **Ellis et al[25]** | 77 | ER-positive | Yes | Whole genome sequencing | 18 significantly mutated genes identified |
| | | | | Whole exome sequencing | *MAP3K1* associated with luminal A subtype |
| | | | | | *GATA3* mutations associated with response to neoadjuvant aromatase inhibitor therapy |
| **TCGA[26]** | 825 | All types | No | DNA copy number and CNA | *TP53* and *PIK3CA* mutations most common, with others substantially less common |
| | | | | DNA methylation | Luminal A subtype characterized by mutations in *PIK3CA, GATA3,* and *MAPK3K*1 |
| | | | | mRNA expression | Basal-like subtype similar to serous ovarian cancer |
| | | | | microRNA expression | |
| | | | | Protein arrays | |
| **Shah et al[28]** | 104 | Triple-negative | No | RNA-Seq | Wide and continuous spectrum of genomic evolution |
| | | | | | *TP53, PIK3CA, PTEN* mutations clonally dominant |

Abbreviations: CNV, copy number variants; CNA, copy number alterations; ER, estrogen receptor; TCGA, The Cancer Genome Atlas.

positive disease. Response to preoperative AI therapy has been shown to be a short-term surrogate reflecting favorable prognosis with AI therapy alone. Response may be assessed by an algorithm that reflects the extent of residual disease after a 16-week course of therapy and Ki67 expression, or Ki67 expression alone after a 2-week course of therapy.[20] Ellis et al performed whole genome analysis of 46 samples and exome sequencing in 31 additional samples from postmenopausal women with ER-positive breast cancers treated with preoperative AI therapy, including 29 samples with Ki67 levels above 10% after a 16-week course of AI therapy (and thus considered to have AI-resistant disease) and 48 with KI67 levels of 10% or less (indicating AI-sensitive disease).[21] There were several important observations from this study. First, the background mutation rate was about twofold higher for AI-resistant tumors than for sensitive tumors. Second, 18 significantly mutated genes were identified, including genes previously identified in breast cancer (e.g., *PIK3CA, TP53, GATA3, CDH1, RB1, MLL3, MAP3K1, CDKN1B*) and novel genes not previously observed, including five previously seen in hematopoietic cancers (e.g., *RUNX1, CBFB, MYH9, MLL3,* and *SF3B1*). Third, certain mutations were associated with specific subtypes, including an association between *MAP3K1* mutations and luminal A tumors and *TP53* mutations with luminal B tumors. Fourth, *GATA3* mutations correlated with AI-induced suppression of proliferation and hence sensitivity to AI therapy.

## INTEGRATING GENOMIC, EPIGENOMIC, AND PROTIEN DATA: THE CANCER GENOME ATLAS (TCGA)

TCGA included tumor and germ-line DNA samples obtained from 825 patients, which included evaluation by Affymetrix 6.0 SNP arrays (733 patients) and Agilent mRNA expression microarrays (547 patients), plus Illumina Infinium DNA methylation chips (802 patients), miRNA sequencing (697 patients), whole exome sequencing (507 patients), and reverse-phase protein arrays (403 patients).[22] This included 466 tumors evaluated for gene expression, methylation, miRNA, and exome sequencing, and 348 who also had protein arrays. Correlation was not performed with clinical outcomes because of the short follow-up time (median 17 months) and resultant small number of events.

Nearly all genes previously implicated in breast cancer were identified (*PIK3CA, PTEN, AKT1, TP53, GATA3, CDH1, RB1, MLL2, MAP3K1, CDKN1B*), plus a number of novel genes recently identified in other studies (*TXB1, RUNX1, CBFB, AFF2, PIK3R1, PTPN22, PTPRD, NF1, SF3B1, CCND3*). The overall mutation rate was lowest in the luminal A subtype and highest in the basal-like and HER2-enriched intrinsic subtypes. The most commonly mutated genes included *TP53* (37%), *PIK3CA* (36%), *GATA3* (11%), *MAP3K1* (8%), *MLL3* (7%), and *CHD1* (7%), with 17 other mutations occurring in 1% to 4% (Fig. 2A). The distribution of mutations varied by subtype (Fig. 2B), with PIK3CA mutations occurring more commonly in luminal A/B and HER2-enriched than in basal subtypes (45%/29% and 39% vs. 9%), and *TP53* mutations dominating in basal (80%) and HER2-enriched subtype (72%) compared with luminal B (29%) and luminal A (12%) subtypes. This pattern is similar to other reports, with *PIK3CA* and *TP53* mutations predominating and consistently varying in frequency by subtype and with other mutations relatively uncommon. The types of mutations also differed by intrinsic subtypes, including differences in *TP53* mutations between basal-like (nonsense and frame shift) and luminal tumors (missense). Approximately 9% of 507 cases evaluated revealed germ-line predisposing

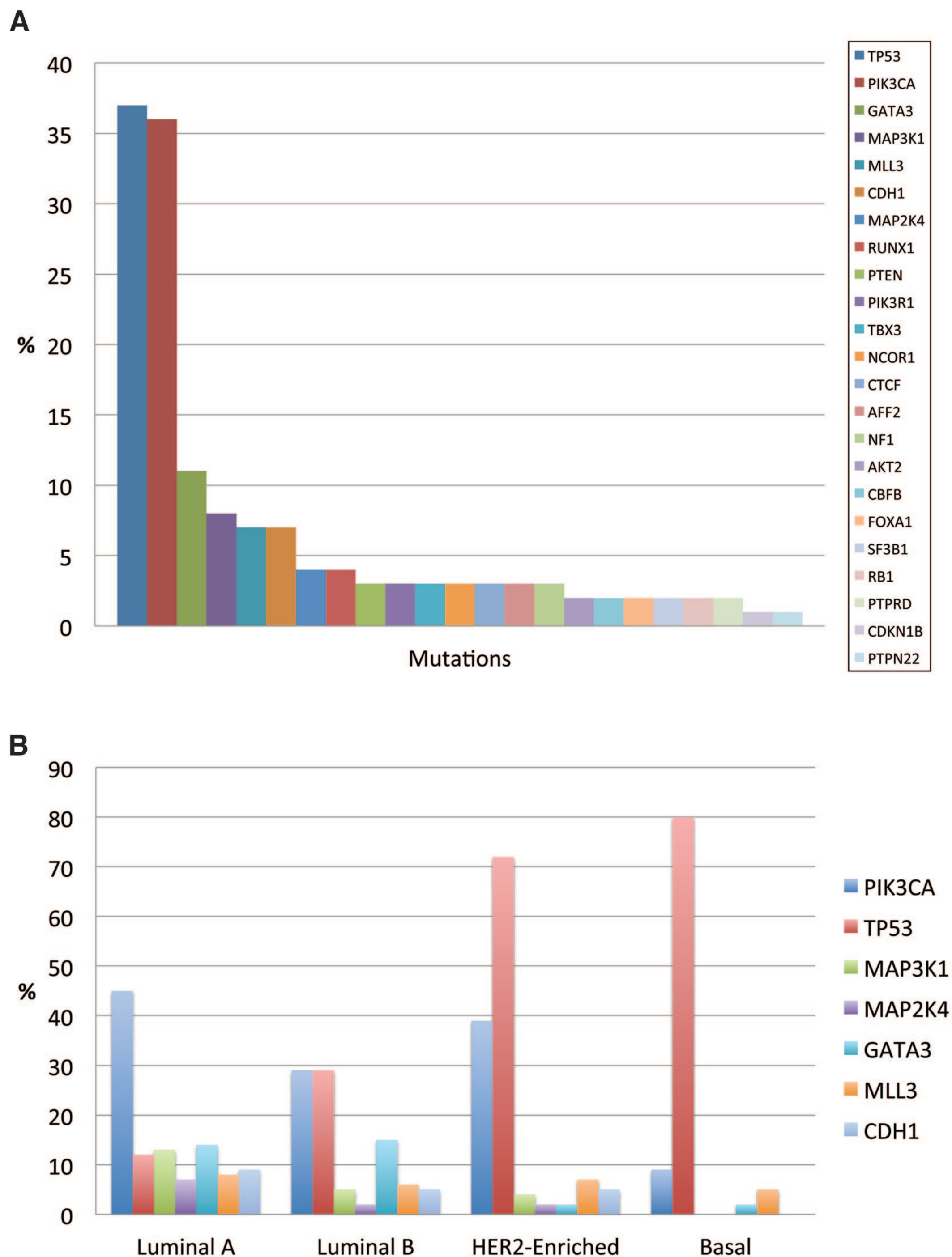


**FIG 2. Distribution of genomic alterations observed in The Cancer Genome Atlas study in breast cancer, included all tumors (2a) and by intrinsic subtype the seven most common mutations occurring in at least 5% (2b).**

variants (e.g., *ATM, BRCA1, BRCA2, BRIP1, CHEK2, NBN, PTEN, RAD51C*, *TP53*). Similar to other reports, copy number changes correlated with some intrinsic subtypes, including loss of 5q and gain of 10p in basal-like cancers and gain of 1q and/or 16q loss in luminal tumors.

A unique feature of TCGA was the comprehensive nature of the molecular profiling, which included evaluation of miRNA, methylation, and proteins. Clustering analysis revealed seven subtypes by miRNA that did not correlate with mutational status or PAM50, with the exception of two miRNA groups that showed overlap with basal-like subtype and *TP53* mutations. Five distinct DNA methylation groups were identified, including at the extremes a hypermethylated phenotype enriched for luminal B subtype and a

hypomethylated phenotype that overlapped with basal-like tumors enriched for *TP53* mutations. Protein analysis identified seven subtypes that were highly concordant with the mRNA intrinsic subtypes, especially basal-like and HER2-enriched subtypes. A multiplatform data matrix analysis revealed that the information content from copy number aberrations, miRNAs, and methylation is captured at the level of gene expression and protein expression and activity.

TCGA permitted a detailed analysis of each of the intrinsic mRNA-defined subtypes. Luminal tumors exhibited the most heterogeneous gene expression, mutational spectrum, and copy number changes and were associated with several characteristics. First, although *PIK3CA* mutations were common in luminal tumors, markers typical of an activated PI3K pathway (e.g., pAKT, pS6, p4EBP1 protein) were not present. In contrast, *PIK3CA* mutations were usually associated with pathway activation in basal and HER2-enriched tumors. Second, luminal tumors also were frequently associated with *MAP3K1* and *MAP2K4* mutations—two contiguous steps in the p38-JNK pathway—whereas they occurred infrequently in other subtypes. Third, in comparison to luminal B tumors, luminal A tumors were associated with higher frequency of intact *TP53* and *Rb1* tumor suppressor genes. HER2-enriched tumors showed high aneuploidy, somatic mutation rate, and DNA amplification of other potential targets (e.g., *FGFR, EGFR, CDK4, cyclin D1*). The basal subtype showed a high degree of *TP53* mutations and high PI3K pathway activity despite a low PI3K mutation rate (because of *PTEN* and *INPP4B l*oss and/or amplification of *PIK3CA*). Similar to serous ovarian carcinoma, DNA repair deficits (ATM mutations, BRCA1 and 2 inactivation, RB1 loss, and cycle E activation), genomic instability, and increased activity of the *HIF1-alpha/ARNT*, *MYC* and *FOXM1* pathways were also common.

**TABLE 3. Summary of Key Findings from Next-Generation Sequencing Studies in Breast Cancer**

| Disease Classification, Prognosis, and Prediction |
|---|
| Breast cancer is a heterogeneous disease with variable genomic complexity. |
| Gene expression classifies disease into "intrinsic" subtypes (luminal A, luminal B, HER2-enriched, basal). |
| Evaluation of somatically acquired CNAs and germ-line CNVs permit further sub-classification. |
| Basal-like breast cancers are genotypically similar to serous ovarian carcinoma. |
| Some mutations are predictive of response to aromatase inhibitors (*GATA3*). |
| **Target Discovery** |
| Somatic mutations are common and dominated by point mutations, duplications, translocations, and rearrangements. |
| Rearrangements commonly result in fusion genes, some of which produce oncogenic proteins. |
| Most common currently targetable mutations are in the PI3K-mTOR-AKT pathway. |
| Not all *PI3KCA* mutations lead to pathway activation (especially in luminal disease). |
| **Promise** |
| Next-generation sequencing is available for clinical use due to improved methodology and declining costs. |
| Refined disease classification may assist in guiding standard therapy. |
| Up to approximately 20% of genomic alterations may be potentially actionable targets. |
| **Pitfalls** |
| Few breast cancers are "addicted" to driver alterations. |
| Inactivation of tumor suppressor genes (e.g., *TP53*, *RB1*) are difficult to therapeutically target. |
| Preliminary clinical trials show that fewer than about 30% of screened patients receive a genomically directed therapy. |
| Logistical barriers include need for representative biopsy and current dearth of available targeted agents. |
| Scientific barriers include tumor heterogeneity, intrinsic and acquired resistance, and need for combinatorial therapy. |

Abbreviations: can, copy number alterations; CNV, copy number variants.

## GENOTYPIC ANALYSIS OF PHENOTYPICALLY DEFINED TRIPLE-NEGATIVE BREAST CANCER

The METABRIC and TCGA studies focused on using NGS to sub-classify intrinsic breast cancer subtypes. However, phenotypic classification based on patterns of ER, progesterone receptor, and HER2/*neu* expression is typically used in clinical practice for clinical decision making. Triple-negative breast cancer (TNBC) is a phenotypic subset that accounts for approximately 15% of all breast cancers, occurs more commonly in younger women and black or Hispanic women, and is characterized by a higher risk of relapse, earlier time to relapse, proclivity for recurrence in visceral organs and the central nervous system, and absence of specific targeted therapy.[23] Shah et al described an analysis of 104 patients with TNBC subjected to RNA-Seq and deep resequencing measurements of allelic abundance for more than 2,400 somatic mutations.[24] Approximately 20% of tumors had potentially clinically actionable somatic aberrations, including *BRAF* V600E, high-level *EGFR* amplifications, and *ERBB2* and *ERBB3* mutations. The distribution of somatic mutation abundance varied in a continuous distribution and was unrelated to CNA or tumor cellularity. In another report, Banerji et al identified a recurrent *MAGI3-AKT3* fusion in TNBC that led to constitutive activation of AKT kinase that was abolished by a competitive AKT small-molecule inhibitor.[25]

## CLINICAL TRIALS INTEGRATING HIGH-THROUGHPUT GENOMIC ANALYSIS INTO CLINICAL CARE

Andre et al tested the ability of array comparative genomic hybridization (CGH) and Sanger sequencing to provide improved therapeutic direction in the SAFIR1 trial.[26] Patients with metastatic breast cancer underwent biopsy of metastatic sites for genomic analysis, followed by genotype-directed therapy after progression on standard therapy. At the time of their initial preliminary report, of the 423 patients who

consented, 59% eventually had their tumor sequenced, resulting in 26 patients (6% of consented group) who received a genome-directed treatment and eight patients (2% of consented group) having evidence of clinical benefit from the treatment. It is anticipated that up to 30% may eventually receive a genome-directed therapy in this ongoing trial. In the SAFIR2 trial, patients with HER2-negative metastatic breast cancer will undergo biopsy of a metastatic disease site and if, after completing six to eight cycles of standard chemotherapy, they are responding or showing stable disease with a potentially actionable genomic alteration, they will be randomly assigned to receive either standard care or a treatment targeted to the alteration.

In addition to matching the right drug or combinations of drugs with the right patient, a significant limitation of using NGS to direct therapy may be tumor heterogeneity from the mutator phenotype and other factors contributing to heterogeneity and the emergence of drug resistance.[27-28] For example, evaluation of the primary and metastatic sites in two patients with renal cell carcinoma revealed that biopsy of a single metastatic site was not representative of the genomic landscape of all tumor sites and that tumor adaptation contributed to therapeutic failure.[29]

## CONCLUSION

Key points about the knowledge gained thus far from applying NGS to clinical breast cancer specimens are summarized in Table 3. Potentially promising applications of NGS include target discovery, refined and more accurate disease classification, and improved therapeutic direction. However, there remain significant pitfalls including logistical obstacles to having an effective drug available to target every oncogenic alteration, and a regulatory and cancer care delivery system that would allow matching the right patient with the right drug, as exemplified by the SAFIR1 trial. In addition to the logistical obstacles, other pitfalls include a dearth of tumors that are "addicted" to oncogenic mutations, innate tumor heterogeneity at presentation resulting in the need for combinatorial therapy directed at multiple aberrant pathways, and the rapid emergence of resistance to therapy even when appropriately applied.

## ACKNOWLEDGMENT

Supported in part by grants from the National Institutes of Health (P30–13330), Susan G. Komen for the Cure (KG100888), and the American Cancer Society (RSG-Tris-borate-EDTA [TBE]-123001).

## Disclosures of Potential Conflicts of Interest

*Relationships are considered self-held and compensated unless otherwise noted. Relationships marked "L" indicate leadership positions. Relationships marked "I" are those held by an immediate family member; those marked "B" are held by the author and an immediate family member. Relationships marked "U" are uncompensated.*

**Employment or Leadership Position:** None. **Consultant or Advisory Role:** Joseph A. Sparano, Pfizer. **Stock Ownership:** None. **Honoraria:** Joseph A. Sparano, Bristol-Myers Squibb. **Research Funding:** Joseph A. Sparano, Merck. **Expert Testimony:** None. **Other Remuneration:** None.

## References

1. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 1975;94:441-448.
2. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977;74:5463-5467.
3. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409:860-921.
4. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science.* 2001;291:1304-1351.
5. Soon WW, Hariharan M, Snyder MP. High-throughput sequencing for biology and medicine. *Mol Syst Biol*. 2013;9:640
6. Sparano JA, Fazzari M, Kenny PA. Clinical application of gene expression profiling in breast cancer. *Surg Oncol Clin N Am.* 2010;19:581-606.
7. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature.* 2000;406:747-752.
8. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A.* 2003;100:8418-8423.
9. Gastaldi S, Sassi F, Accornero P, et al. Met signaling regulates growth, repopulating potential and basal cell-fate commitment of mammary luminal progenitors: implications for basal-like breast cancer. *Oncogene*. Epub 2012 May 7.
10. Molyneux G, Geyer FC, Magnay FA, et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell.* 2010;7:403-417.
11. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol.* 2009;27:1160-1167.
12. Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol.* 2007;25:5287-5312.
13. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19:1639-1645.
14. Hillmer AM, Yao F, Inaki K, et al. Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes. *Genome Res.* 2011;21:665-675.
15. Wood LD, Parsons DW, Jones S, et al. The genomic landscapes of human breast and colorectal cancers. *Science.* 2007;318:1108-1113.
16. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature.* 2012;486:400-404.

17. Stephens PJ, McBride DJ, Lin ML, et al. Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature.* 2009;462: 1005-1010.
18. Inaki K, Hillmer AM, Ukil L, et al. Transcriptional consequences of genomic structural aberrations in breast cancer. *Genome Res.* 2011;21: 676-687.
19. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012;486:346-352.
20. Ellis MJ, Suman VJ, Hoog J, et al. Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2 to 3 breast cancer: clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype-ACOSOG Z1031. *J Clin Oncol.* 2011;29:2342-2349.
21. Ellis MJ, Ding L, Shen D, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature.* 2012;486: 353-360.
22. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490:61-70.
23. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med.* 2010;363:1938-1948.
24. Shah SP, Roth A, Goya R, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature.* 2012;486: 395-399.
25. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature.* 2012;486:405-409.
26. Andre F, Bachelot T, Campone M, et al. Array CGH and DNA sequencing to personalize therapy for metastatic breast cancer: a prospective National trial (UNICANCER SAFIR-01). *European Society of Medical Oncology.* 2012;23(suppl 9):7-30 (abstr 3426).
27. Loeb LA. Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat Rev Cancer.* 2011;11:450-457.
28. Ng CK, Pemberton HN, Reis-Filho JS. Breast cancer intratumor genetic heterogeneity: Causes and implications. *Expert Rev Anticancer Ther.* 2012;12:1021-1032.
29. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med.* 2012;366:883-892.