Translating Cancer Genomes and Transcriptomes for Precision Oncology

Sameek Roychowdhury, MD, PhD1,2,3; Arul M. Chinnaiyan, MD, PhD4,5,6,7,8,9

Understanding the molecular landscape of cancer has facilitated the development of diagnostic, prognostic, and predictive biomarkers for clinical oncology. Developments in next-generation DNA sequencing technologies have increased the speed and reduced the cost of sequencing the nucleic acids of cancer cells. This has unlocked opportunities to characterize the genomic and transcriptomic landscapes of cancer for basic science research through projects like The Cancer Genome Atlas. The cancer genome includes DNA-based alterations, such as point mutations or gene duplications. The cancer transcriptome involves RNA-based alterations, including changes in messenger RNAs. Together, the genome and transcriptome can provide a comprehensive view of an individual patient’s cancer that is beginning to impact real-time clinical decision-making. The authors discuss several opportunities for translating this basic science knowledge into clinical practice, including a molecular classification of cancer, heritable risk of cancer, eligibility for targeted therapies, and the development of innovative, genomic-based clinical trials. In this review, key applications and new directions are outlined for translating the cancer genome and transcriptome into patient care in the clinic. CA Cancer J Clin 2015;00:000–000. © 2015 American Cancer Society.

Keywords: genomics, transcriptome, neoplasm, gene fusion, patient care

Introduction

The molecular classification of cancer has informed novel approaches for clinical practice in oncology, such as diagnosis, prognosis, and treatment decisions. In 2011, the National Research Council convened a committee to develop a framework for the precision taxonomy of human disease, using molecular classification as the foundation for advancing personalized or precision cancer medicine.1,2 Since 2003, new DNA sequencing technologies called next-generation sequencing (NGS) have increased the speed and reduced the cost of sequencing cancer by nearly 1 million-fold each.3 These technologies have advanced our understanding of various subtypes of cancer through several national and international, large-scale, basic science, cancer profiling efforts.4,5 Since 2011, clinicians have begun to translate genome and transcriptome sequencing approaches for patients with cancer in the clinic.6 In this review, we discuss the current and future impact of genome and transcriptome sequencing for patient care in clinical oncology.

Omic and Defining the Cancer Genome and Transcriptome

Through technology innovations that have miniaturized and parallelized laboratory tests to allow testing of thousands of molecules, we have entered the “omics era.” Omics refers to the collection and analysis of large data sets of biologic variables...
or phenomenon and is regularly applied to genes (genomics), proteins (proteomics), and their subunits, including nucleotides and amino acids (metabolomics). Before the omics era, researchers studied the traditional paradigm in which genes encode a sequence of nucleotides that is transcribed into a messenger RNA (mRNA). These mRNAs are processed by ribosomes, where the sequence information is translated into a protein with the sequential addition of transfer RNAs (tRNAs) and their associated amino acids. Collectively, proteins can affect biologic processes, such as metabolism, by functioning as enzymes or structural proteins or by regulating genes themselves through transcription or translation. Today, rather than studying a few genes, transcripts, or proteins, researchers can use new technologies to rapidly test or evaluate up to tens of thousands of data points.

The human genome includes 2 haploid sets of 23 chromosomes each comprised in total of 6 billion nucleotides. One haploid set of chromosomes encodes approximately 20,000 genes that are transcribed into RNAs, including the classical mRNAs, ribosomal RNAs (rRNAs), and tRNAs. Collectively, these RNAs constitute the transcriptome. The mRNA sequence provides the recipe for a protein and is translated through ribosome machinery (including rRNA and ribosomal proteins) by consecutively adding amino acids carried by tRNAs. Additional RNAs have been discovered that do not encode proteins, termed noncoding RNAs (ncRNAs). These ncRNAs include microRNAs (miRNAs) and long ncRNAs and have more recently been proven to have regulatory functions that affect gene expression and protein function. Before 2001, the scientific community did not have a complete road map or dictionary of the entire human genome. The Human Genome Project was initiated by the National Institutes of Health and coordinated the sequencing of the human genome over 14 years at a cost greater than 3 billion US dollars. The project provided the necessary reference for researchers to study the role of genetics in human diseases such as cancer. The project also inspired the development of new technologies that have changed the landscape of genomics research by accelerating the speed and reducing the cost of DNA sequencing by 1 million-fold.

**Historical Impact of Omics on Clinical Medicine**

The impact of omics data on precision medicine already can be seen in clinical practice today. Clinical practices include the application of data from genes, transcripts, and proteins toward diagnosis, disease monitoring, risk determination, counseling, and development of novel therapies. In an early application for metabolomics, Koenig et al described the value of assessing glycated hemoglobin (hgb A1c) as an everyday metabolic measure of long-term glucose levels for patients with glucose intolerance or diabetes. Factor V Leiden is a genetic risk factor that occurs in 5% of North American Caucasians as a heterozygote mutation; it is routinely applied toward risk assessment for thrombosis, and it has led to the avoidance of prothrombotic drugs or prophylaxis recommendations in high-risk situations for patients who are at increased risk. Genetics has also made an impact on the most common form of dementia, Alzheimer dementia, with the discovery of several genetic factors involved in this disease. Mutations in the gene for apolipoprotein E have been identified as a risk factor for late-onset Alzheimer dementia, and research efforts are underway to study other genetic factors. The genetic basis of cystic fibrosis was described in 1989, and from 3% to 4% of Caucasians are carriers for this disease. Early on, genetic testing was important for counseling and diagnosis; and, more recently, metabolic research has led to the development of novel therapies for cystic fibrosis. In cancer, omics research led to the application of chromosome karyotyping of leukemias that guide diagnosis, risk stratification, and therapy selection. The discovery of the Philadelphia chromosome in chronic myeloid leukemia and the subsequent characterization of the breakpoint cluster region-Abelson murine leukemia viral oncogene homolog 1 (BCR-ABL1) gene translocation would pave a path for the development of imatinib, the first tyrosine kinase inhibitor for cancer. The majority of these omics discoveries preceded the Human Genome Project (1989-2002), which has opened new doors for genomic medicine.

**NGS Technologies and Cancer**

DNA sequencing has consistently relied on using one strand of DNA as a template to synthesize the other strand by adding complementary nucleotides with the enzyme DNA polymerase. The first method to determine which nucleotides are added was described in 1977 and involved termination of the sequencing reaction with nucleotides that cannot allow further DNA synthesis (so-called Sanger or dyeoxy termination). This strategy relied on radiographic detection of radioactively labeled nucleotides and gel electrophoresis and, over time, was supplanted for practical reasons by fluorescent-labeled nucleotides and electrophoresis in small capillary tubes. Thus, the Human Genome Project was completed with capillary sequencers. In 2005, new technologies known as NGS miniaturized and parallelized the sequencing process to improve yield and reduce cost. All NGS methods start by fragmenting DNA into small segments, followed by the addition of adaptors that allow the fragment to be sequenced. The addition of specific nucleotides can be measured by light (Illumina, San Diego, CA) or pH changes (Ion Torrent; ThermoFisher Scientific, New York, NY). The resulting sequence data are comprised of
millions of pieces ranging in length from 50 to 250 base pairs that must be matched and assembled into a reference genome or map, and this is akin to a jigsaw puzzle. Base pairs that are different from the expected reference may be mutations. This field is called bioinformatics data analysis and uses high-performance computing to process this so-called big data set. In addition to DNA sequencing, RNA can be similarly sequenced by first converting RNA into complementary DNA using the reverse transcriptase enzyme and then following the same procedure for DNA. Today, there are third-generation technologies for NGS that have effectively decreased the cost of sequencing by 1 million-fold since the Human Genome Project. Technologies for NGS are further detailed and reviewed elsewhere.3

Cancer Genome (DNA) Sequencing

Two major collaborative efforts, The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium, have used NGS to profile the landscape of the 30 most common cancer types. Each of these efforts has collected primary tumors from surgical resections for tumor sequencing. In addition to these collaborative, multicenter projects, there are numerous independent research groups that have contributed to cancer genomic profiling for cancer subtypes with lower prevalence. In general, rather than sequencing the whole genome, these profiling projects have focused on 1% of the genome containing the approximate 20,000 known genes, also known as the whole exome. Exome sequencing uses probes or baits to allow NGS platforms to focus on DNA fragments containing the 20,000 known genes, thereby reducing the cost of sequencing by 100-fold. Despite the improvements in cost and throughput for NGS, whole-genome sequencing remains expensive for both basic and clinical research applications. Today, the majority of basic and clinical applications use a targeted or whole-exome approach.20

These large-scale cancer-profiling projects have revealed a landscape of the cancer genome that includes a diverse variety of genomic alterations, including point mutations, copy number variation, and translocations. These genomic alterations can affect a variety of cellular processes from cell signaling and metabolism to gene expression.21 Point mutations are single base pair substitutions that can change the function of a gene’s protein product. One example is the clinically relevant B-Raf proto-oncogene, serine/threonine kinase (BRAF oncogene) mutation V600E (an amino acid change from valine to glutamic acid at codon 600) that occurs commonly in melanoma and leads to constitutive activation compared with the wild-type gene.22-25 Copy number variation refers to either extra or missing copies of a gene. Tumors with copy number amplification have extra copies beyond the expected two genes, such as from 50 to 100 copies of human epidermal growth factor receptor 2 (HER2 or ERBB2), as seen in approximately 20% of breast cancer.26-28 Alternatively, loss of gene copies, also known as deletion, often occurs in tumor suppressor genes such as phosphatase and tensin homolog (PTEN) in prostate and other cancers.29 Translocations or gene rearrangements involve two genes that are brought together and have a new function as a chimera. These events can create new functions for the gene fusion or can inactivate functions. Gene fusions in cancer commonly involve kinases (enzymes that have phosphorylating activity) and transcription factors that are deregulated by the fusion event. With the advent of chromosomal karyotyping and banding in the 1960s, the first appearances of chromosomal rearrangements were uncovered in hematological malignancies and sarcomas because of the ability to easily obtain tumor tissue and metaphase chromosomes in these cancers. Visible chromosomal rearrangements in lymphomas, leukemias, and sarcomas facilitated the identification of novel oncogenes and tumor suppressors in cancer and characterization of their role in cancer biology and clinical applications.30 As an example, chronic myeloid leukemia, characterized by the Philadelphia chromosome and BCR-ABL1 gene rearrangement,31,32 would later become a model for understanding kinases and the application of targeted therapies for treatment of cancer.33 Meanwhile, outside of sarcomas and select solid tumors, there initially was a paucity of oncogenic gene fusions or translocations recognized in solid tumors. This was largely because of limited tissue access and technological restrictions; however, it was predicted that gene fusions would be recurrent genomic alterations in solid tumors.34 Since 2005, cancer genome and transcriptome sequencing has revealed additional, clinically relevant, novel gene fusions in solid tumors.35

Cancer Transcriptome Profiling

The transcriptome is comprised of “classical” RNAs (mRNA, rRNA, and tRNA) as well as multiple subtypes of noncoding RNA (miRNAs and long ncRNAs) that have been discovered to have novel regulatory functions in cell biology.36 Gene expression can be characterized using earlier microarray technology or the more recent transcriptome sequencing (RNAseq) methods. Transcriptome sequencing has significant advantages, including precise detail about base pairs and the ability to detect novel RNAs that cannot be detected on microarrays. For clinical applications of the cancer transcriptome, efforts have focused on using gene expression to classify cancer subtypes (that differ with regard to prognosis and response to specific treatments) and to detect gene fusions or rearrangements. In routine clinical practice, fluorescence in situ hybridization (FISH) and reverse transcriptase–polymerase chain reaction (RT-PCR) are used to detect gene rearrangements but are
limited by only testing for one gene at a time. Hence, the advantage of sequencing approaches is the ability to detect multiple gene rearrangements as well as novel ones. Although genome sequencing can detect fusions, whole-genome sequencing of cancer remains costly, and the cost for RNAseq is a fraction of that for whole-genome sequencing and has been applied with new bioinformatics approaches to detect fusions. By using a paired-end sequencing approach for RNA, gene fusions that are expressed at the transcript level can be detected. More recently, Stransky et al performed a comprehensive analysis of publicly available tumor RNA sequence (RNAseq) data for nearly 7000 cancers in TCGA to catalog a diverse landscape of known and novel candidate kinase gene fusions. Furthermore, Klijn et al performed RNAseq on 675 cancer cell lines and similarly cataloged kinase fusions. The application of RNAseq to detect novel, clinically relevant gene fusions in cancer is still in its infancy, and we anticipate that additional fusions will be discovered as the number of cancers profiled increases, especially in rare or previously uncharacterized cancer subtypes. Importantly, detection of novel gene fusions involving kinases has also led to novel treatment opportunities and therapeutic benefit with kinase inhibitors in patients with advanced cancer.

In pediatric B-cell acute lymphoblastic leukemia, Roberts et al recently identified kinase fusions involving genes, such as ABL1, Janus kinase 2 (JAK2), colony-stimulating factor 1 receptor (CSFR1), and neurotrophic tyrosine kinase receptor 3 (NTRK3), that have corresponding targeted therapies, which opens up new treatment hypotheses for patients with this type of leukemia to be tested in clinical trials.

Gene expression signatures can be used to classify cancer types into molecular subsets that have clinical relevance. For example, early studies applied transcriptome profiling of B-cell lymphoma using microarray technologies to further classify this disease into clusters of activated B-cell and germinal center B-cell subtypes. Activated B-cell lymphoma bears a poorer prognosis compared with germinal center B-cell lymphoma. In another pivotal study of breast cancer, Perou et al used microarray-based transcriptome profiling on primary breast cancer samples and classified this disease into five molecular subsets with biological and clinical relevance. Moving beyond microarray technologies, RNAseq has the potential to enable the study of other diverse components of the cancer transcriptome.

For example, in addition to the classical elements of the transcriptome including messenger (mRNA), ribosomal (rRNA), and transfer (tRNA) RNAs, multiple subtypes of RNA have been discovered with novel regulatory functions in cell biology. In fact, the majority of the transcriptome is comprised of nonprotein encoding RNAs (ncRNAs), including but not limited to miRNAs, small interfering RNAs, and long noncoding RNAs. Beyond the classical function for mRNAs that encode proteins, these novel RNAs can play multiple roles in cell biology, ranging from regulation of transcription, posttranscriptional events, gene silencing, translation, and protein-level function. Much like prototypical genes encoded by DNA, miRNAs are subject to genomic alterations, including mutation, deletion, amplification, and epigenetic modifications. Similarly, miRNAs can function as tumor suppressor or oncogenes. Small interfering RNAs are small RNAs that mediate a highly specific gene-silencing mechanism, which is conserved from nematodes and plants to mammalian biology, and have emerged as tools for biomedical research and potential strategies for gene-silencing therapies. Newly described long noncoding RNAs are ubiquitous in cancer, have diverse regulatory functions, and are only recently being systemically characterized.

Cancer Genomes and Transcriptomes: What Have We Learned?

Over 10,000 cases of cancer have undergone DNA sequencing and have been published through collaborative projects, and this has revealed diverse heterogeneity within and across cancer types classified by tissue of origin (eg, breast, lung). We have also learned that signatures of mutational patterns for point mutations can aid classification through association with an underlying mechanism, such as defects in DNA repair, radiation exposure, and tobacco exposure. As an example, several cancer types, such as lung and melanoma, have abundant point mutations because of carcinogen exposures of tobacco smoke and ultraviolet radiation, respectively (Fig. 1). In contrast, acute myeloid leukemia and prostate cancer generally have few point mutations and, rather, have more copy number variation and gene fusions. Ciriello et al retrospectively assessed over 3000 cases from the TCGA and evaluated 12 types of cancer as having predominantly point mutations (M class) and copy number variations (C class), while the majority are a mixture of both M and C class within a single disease type. Along with the genome, the cancer transcriptome has informed the classification of lymphoma and breast cancer into clinically relevant molecular subsets. Lung cancer is an ideal example in which genome and transcriptome profiling have affected measureable clinical outcomes by moving from histology to a genomics-based classification based on point mutations (BRAF V600E), copy number alterations (MET proto-oncogene, receptor tyrosine kinase [MET] amplification), and gene fusions (anaplastic lymphoma kinase [ALK] fusion), which lead to treatment with matching targeted therapies.

Moving forward to the clinic, there are several lessons to consider for translation. It has become clear that clinical
decision making for cancer will require a personalized approach based on an individual’s cancer, which is likely to be unique compared with other patients, even among cancers of the same histologic type. Second, with slightly more than just 10,000 cases analyzed, researchers have only detected the clinically relevant mutations that represent greater than 20% of common cancer types. Therefore, based on limited sampling, we have not yet uncovered clinically relevant mutations with an estimated prevalence less than 20% in common cancers or in rare cancers that simply have not yet been sequenced. Therefore, more cancer sequencing data are necessary to advance a comprehensive catalog of cancers. Third, the majority of cancer genome data available are based on primary tumors rather than metastatic or advanced cancers, which may have acquired additional mutations, are cancers that behave more aggressively, and display more heterogeneity because of selective pressures from therapy. With this knowledge in hand, how are we proceeding to apply cancer genome and transcriptome biomarkers in the clinic?

Types of Molecular Biomarkers Applied in the Clinic: Diagnostic, Prognostic, and Predictive

Research discoveries derived through cancer genome and transcriptome studies have the potential for clinical impact as biomarkers. There are three key types of biomarkers used for clinical decision making, including diagnostic, prognostic, and predictive biomarkers. Diagnostic biomarkers facilitate the identification of a cancer type or subtype. Prognostic biomarkers aid clinicians in determining the risk of relapse or disease progression after therapy, wherein patients with high risk are selected for aggressive screening or adjuvant therapy to prevent recurrence. Clinicians use predictive biomarkers to select one therapy over others, based on associations between biomarker results and the likelihood of response to certain therapies. In practice, predictive biomarkers often identify the molecular targets of relevance to targeted anticancer drugs.

Each type of biomarker could be assayed to detect changes in a tumor’s genome (DNA), transcriptome (RNA), proteome (protein), or by phenotypic characteristics (such as histopathologic classification). As examples of methods to detect these biomarkers, BRAF gene mutation testing for melanoma is a DNA-based predictive biomarker that can guide therapy selection. For RNA-based biomarkers, FISH methods are used for standard diagnostic subtyping of lymphoma to measure Epstein-Barr virus RNA expression. Immunohistochemistry is used to detect estrogen receptor protein in breast cancer and is an example of a biomarker that has both predictive and prognostic value. OncotypeDX testing for breast cancer (Genomic Health Inc, Redwood City, Calif) assesses the expression of 21 transcripts in women with lymph node-negative, estrogen receptor-positive breast cancer and is another example of a biomarker method that is both predictive and prognostic, facilitating the identification of patients after surgery who need further...
therapy (prognosis) and are most likely to benefit from adjuvant chemotherapy (predictive).65

However, before any biomarker can be translated into the clinic for use in standard practice, the clinical utility of the biomarker must be tested through clinical trials to establish its impact and association with clinical outcomes.66 The presence of ALK gene fusions in patients with metastatic lung cancer is an example of a predictive biomarker for clinical response to ALK inhibitors.67,68 Preceding the use of NGS, the clinical utility of ALK gene fusions was established in pivotal clinical trials using standard FISH methods to detect the gene fusion. However, it is costly to develop FISH and Sanger sequencing tests for single genes and their relevant mutations, and NGS has been translated to the clinic to cost-effectively broaden the number of genes and types of mutations tested.

Clinical Examples: Whole-Genome (DNA) Sequencing

Early efforts to use whole-genome sequencing for patients with cancer began as case-by-case research endeavors. In 2011, Welch et al applied whole-genome sequencing for a patient with acute promyelocytic leukemia, which is characterized by having a gene fusion involving the retinoic acid receptor α (RARA).69 Patients with acute promyelocytic leukemia (PML) and RARA fusions are typically very sensitive to therapy (predictive biomarker) using oral all-trans retinoic acid, which significantly improves long-term survival.70 However, this patient’s standard of care testing with cytogenetics and FISH did not detect the expected chromosome 15 and 17 translocation or the RARA fusion. Welch and team hypothesized that the RARA gene might be involved in a cryptic fusion that is not visible using standard cytogenetics or FISH methods and, thus, evaluated the leukemia using NGS. They chose whole-genome sequencing over exome sequencing because they were trying to detect potential chromosomal breakpoints that might not involve the exons tested by whole-exome methods. After 7 weeks of sequencing and analysis, they were able to identify a PML-RARA gene fusion, and this subsequently changed the course of treatment for the patient, who received all-trans retinoic acid therapy instead of allogeneic stem cell transplantation.

In another clinical application, whole-genome sequencing was able to identify a germline or heritable risk of cancer in a woman aged 37 years who had a personal history of ovarian cancer, breast cancer, and secondary therapy-related acute myeloid leukemia. Although the patient did not have a significant family history of cancer, her clinicians were suspicious of her multiple primary cancers. On the basis of the personal history, which suggested hereditary breast and ovarian cancer syndrome, gene testing for the breast and ovarian cancer susceptibility genes (BRCA1 and BRCA2) was completed, but no heritable genetic cause was identified. Link et al performed whole-genome sequencing of the patient’s skin biopsy and bone marrow leukemia sample and identified a deletion in the tumor protein 53 (TP53) gene, which can confer a heritable risk of cancer as part of the Li-Fraumeni syndrome.71 This case is illustrative of the advantages of NGS approaches for detecting multiple genomic alteration types when the etiology is not apparent based on clinical presentation or standard testing. As a comparison, conventional, comprehensive sequencing of BRCA1 and BRCA2 in 2011 would have cost $4000 alone. This information has substantial clinical impact for the patient’s family, where early screening measures to detect cancer are the standard of care for relatives carrying the same mutation.

More recently, Demeure et al evaluated a patient with papillary thyroid cancer whose disease was progressing despite thyroidectomy, radical neck dissections, and radioiodine treatment. They performed whole-genome sequencing of the patient’s thyroid tumor and identified a gene fusion involving echinoderm microtubule-associated protein-like 4 (EML4)-ALK, which is a targetable fusion observed in 3% to 5% of lung cancers but is not commonly found in thyroid cancer.72 This led to treatment with crizotinib, an oral ALK inhibitor approved for lung cancer, and stabilization of the patient’s tumor growth. This example illustrates the advantage of NGS approaches in finding uncommon genetic changes in patients who test negative for the common genetic changes.

These represent a sample of case reports demonstrating how whole-genome sequencing technologies can affect the care of patients by providing an individualized treatment or screening plan that could affect the patient and even the family.

Clinical Examples: Integrating Whole-Exome (DNA) and Transcriptome (RNA) Sequencing

Several groups have developed trials for clinical tumor sequencing to offer patients with advanced cancer new molecular diagnostic tests to classify their tumors, gain molecular eligibility for investigational therapies in trials, and track clinical outcomes.6,73-75 In 2011, we completed a pilot study that offered a combination of whole-genome, whole-exome, and transcriptome sequencing for patients with advanced cancer and returned clinically significant results within a clinically relevant time frame6 (Fig. 2). That study demonstrated the feasibility of offering cancer genomic testing and addressed some of the early logistical challenges related to informed consent, incidental findings, and interpretation. The study also demonstrated the need for multidisciplinary team effort required by oncologists,
genomics scientists, bioinformaticians, pathologists, and genetic counselors. Currently, the majority of NGS assays are focused on targeted DNA sequencing for from 25 to 300 gene panels, as discussed below; however, several academic cancer centers continue to study the merits of whole-exome (20,000 genes) and transcriptome sequencing. Nevertheless, there are several advantages to be gained by incorporating RNA sequencing concurrently with DNA sequencing, including data on gene expression, enhanced variant calling, splice variants, novel RNAs, noncoding RNAs, and gene fusions. For example, whole-transcriptome sequencing lead to the discovery of novel gene fusions involving fibroblast growth factor receptors (FGFRs) that occur in an estimated 5% to 7% of solid tumor cancers. FGFR signaling is an important pathway for cancer biology, and there are multiple inhibitors of FGFR in clinical development. The discovery of FGFR fusions subsequently led to the development of clinical trials of the tyrosine kinase inhibitors ponatinib and BGJ398 for the treatment of patients whose tumors have FGFR fusions (national clinical trials NCT02272998 and NCT02160041).

To expand on this pilot study of an integrative sequencing approach for clinical tumor sequencing, we collaborated with others on efforts to expand clinical cancer genomics in pediatric oncology and also to overcome the logistical barriers to launching a multicenter clinical study in adults. For pediatric oncology, we addressed issues for providing informed consent and assent to minors and their guardians and for obtaining permission for research biopsy and tumor sequencing. That study enrolled 102 patients with refractory cancer, performed exome and transcriptome sequencing of tumors, demonstrating feasibility for pediatric oncology, and was able to identify clinically relevant alterations in 46% of patients. In one example, an infant who was diagnosed with spindle cell sarcoma was found to have a novel translocation involving NTRK1 (neurotrophic tyrosine kinase...
receptor type 1), which lead to treatment with crizotinib, resulting in a partial response, followed by stable disease. To demonstrate feasibility for multicenter trials, our collaborative team designed a study for patients with advanced prostate cancer deployed across multiple clinical sites and evaluated 150 men with metastatic, castration-resistance prostate cancer.85 These patients had progressed after receiving standard antiandrogen hormonal therapies for prostate cancer. We consented patients to tumor biopsy and to tumor and germline testing with whole-exome and transcriptome sequencing approaches. The study has identified pathways with known mutations and also pathways that were not observed previously in prostate cancer, including WNT pathway signaling and somatic defects in DNA repair. This international study has demonstrated multicenter feasibility and has brought attention to additional considerations for scaling the volume of patients in the study related to testing (turnaround time, use of centralized testing, and quality control) and availability of therapies in clinical trials.

**Expanding Cancer Genomic Medicine**

**Clinical Sequencing Exploratory Research Program: Systematically Advancing Genomic Medicine**

To help address the need to systematically apply genomics to the practice of medicine beyond a case-by-case basis, the National Human Genome Research Institute established and funded the Clinical Sequencing Exploratory Research (CSER) program to study and provide guidelines for bringing genomics to clinical practice (cser-consortium.org; accessed July 20, 2015). The CSER Consortium includes sites that study adult and pediatric cancer, cardiovascular disease, and hereditary diseases. The projects include expertise in clinical specialties, laboratory scientists, bioinformaticians, clinical genetics, legal experts, bioethicists, and patient advocates. CSER has implemented working groups to evaluate specialized issues related to return of results, the electronic medical record, genetic counseling, informed consent, outcomes, pediatrics, phenotype measures, standards for sequencing, and cancer.

**Physician and Patient Attitudes on Genomic Testing**

As academic cancer centers begin to deploy clinical trials to evaluate how to deliver genomic medicine, it is vital to assess how both physicians and patients view genomic testing approaches. Gray et al surveyed 160 physicians at an academic cancer center about the use of somatic testing and their genomic confidence. Interestingly, 22% of physicians reported “low confidence in their genomic knowledge,” and the authors suggested a need for guidelines and education to support the understanding of genomic tests for physicians.86 Miller et al completed structured interviews with 17 physicians about genomic testing in practice, and they similarly observed a need for decision tools and education to aid physicians.87 Gray et al and Blanchette et al completed studies by interview and questionnaire, respectively, of patients with cancer about genomic testing and observed that patients were interested and motivated to undergo genomic testing and potentially improving their cancer care, but they also learned that some patients had concerns about incidental findings, discrimination, and a need for more information or genetic counseling.88,89 Studies such as these and through the CSER program are critical to address barriers for translating cancer genomic medicine into the larger clinical oncology and patient community.

**Clinical Interpretation of Tumor Sequencing Results**

Interpretation of mutations is critical for translation of genomic testing results into the clinic. There are many software tools to predict or model the potential impact of mutations in basic science research,90 but there is a need for expert clinical annotation of specific mutations that provides the exact level of clinical or preclinical evidence a physician will need for decision making. When physicians receive genomic test results, they are faced with mutations in a large number of genes across multiple pathways, and this is a significant obstacle for busy practicing oncologists who cannot keep up with the vast volumes of data that are emerging. Not all mutations are driver mutations that confer a selective advantage (to the cancer) for its survival, growth, or spread, and most are so-called passenger mutations or variants of unknown significance. Several databases and websites have been developed as clinical decision support tools to aid in the interpretation of mutations, including MyCancerGenome (mycancergenome.org; accessed July 20, 2015), Knowledge Base for Precision Oncology (pct.mdanderson.org; accessed July 20, 2015), and Cancer Driver Log (candl.osu.edu; accessed July 20, 2015). Moving forward, the CSER Tumor Working Group and the ClinGen Somatic Working Group are working together to provide overarching infrastructure and leadership to support a more comprehensive database and framework for clinical interpretation of somatic mutations.

**Genomic Tests That Inform Clinical Decision Making**

Bringing NGS-based cancer genomic testing up to clinical-grade standards to support clinical decision making equates to understanding and following standards for molecular diagnostics. Although NGS is relatively new to the molecular pathology and diagnostics community, several groups
have already offered guidelines to address quality for NGS-based cancer genomic testing. Assays must undergo analytic validation that includes determination of the assay’s sensitivity and specificity for detecting mutations using standards. Clinical validation of the assay refers to the broader application of the assay on clinical samples, such as formalin-fixed, paraffin-embedded or frozen tumors, and association of the test results with real-world clinical diagnoses. This often includes confirmation of the test result by another assay, such as Sanger sequencing, PCR, or FISH. All of this is performed in a clinical-grade laboratory that has been inspected by a certifying body, such as a state Department of Health or the College of American Pathologists, to ensure that the laboratories have standard operating procedures, training, and quality-assurance programs in place to deliver quality tests. Demonstrating clinical utility of the assay is separate from building the tests with analytic validity and is subsequently completed through clinical trials that may look at clinical outcomes retrospectively or prospectively.

### Cancer Gene Panels and Case Reports

Developing NGS-based genomic tests in clinical-grade laboratories has considerable constraints, which include ensuring a rapid turnaround time, keeping costs of the assay down, and limiting the complexity (size) of the assay for analysis. As a consequence, many commercial and academic laboratories have developed targeted gene panels focused on from 25 to 400 genes that are known to be important for cancer biology or disease management (Table 1). Because of the ease of testing for gene panels, thousands of patients have undergone genomic testing with cancer gene panels in the United States since 2012. Ou et al reported a patient with lung cancer who had a novel ROS proto-oncogene 1, receptor tyrosine kinase (ROS1) gene fusion identified based on an NGS test that was missed by standard FISH or PCR approaches because it involved a novel fusion partner with TMEM106B (transmembrane protein 106B). Unfortunately, testing was not initiated until the patient had progressed on standard chemotherapy and died from disease before receiving an ROS1 inhibitor. Chalmer et al reported that a patient who had a myeloid neoplasm with eosinophilia had a novel gene fusion involving PDGFRα (platelet-derived growth factor receptor, α polypeptide) and subsequently benefited from therapy with imatinib. Once again, standard FISH testing missed this particular gene fusion, because it is designed to detect only a specific fusion. Ali et al observed a patient who had metastatic kidney cancer with a tuberous sclerosis 1 (TSC1) mutation, which is predicted to result in activation of mechanistic target of rapamycin (mTOR) signaling and who responded clinically and benefited from mTOR inhibitors.

### Unmet Needs: Cancers of Unknown Primary

Cancers of unknown primary (CUPs) represent from 2% to 3% of adult cancers—up to 80,000 cases per year in the United States—and are defined by the inability to identify the anatomic organ or tissue of origin in these patients using traditional radiologic imaging and immunohistochemical assessment of the tumor. Without knowledge of the tissue of origin, it is challenging for oncologists to select the appropriate treatment for these patients, and overall survival outcomes for these patients are poor. In clinical practice, the focus has been to identify the most probable tissue of origin based on clinical presentation and available pathologic data, recognize favorable subsets of cancer when possible, and choose therapies that match the suspected disease. Up to 20% of CUPs may be favorable subsets.

### Table 1. Commercial Targeted DNA Pan-Cancer Next-Generation Sequencing Assays

<table>
<thead>
<tr>
<th>VENDOR</th>
<th>ASSAY NAME</th>
<th>NO. OF GENES</th>
<th>RESULTS</th>
<th>ESTIMATED TURNOVER TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foundation Medicine (Cambridge, MA)</td>
<td>Foundation One</td>
<td>315</td>
<td>SNVs, CNVs, fusions</td>
<td>12-14 days</td>
</tr>
<tr>
<td>University of Washington (Seattle, WA)</td>
<td>UW-Oncoplex</td>
<td>234</td>
<td>SNVs, CNVs, fusions</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Paradigm (Ann Arbor, MI)</td>
<td>PCDx</td>
<td>114</td>
<td>SNVs, CNVs, fusions</td>
<td>4-5 days</td>
</tr>
<tr>
<td>Genomics and Pathology Services, Washington University School of Medicine (St. Louis, MO)</td>
<td>Solid Tumor Gene Set</td>
<td>48</td>
<td>Hot-spot mutations, 6 fusions</td>
<td>3 weeks</td>
</tr>
<tr>
<td>ARUP Laboratories (Salt Lake City, UT)</td>
<td>Solid Tumor Mutation Panel</td>
<td>48</td>
<td>Hot-spot mutations</td>
<td>14 days</td>
</tr>
<tr>
<td>Caris Life Sciences (Irvine, TX)</td>
<td>MI Profile</td>
<td>46</td>
<td>Hot-spot mutations</td>
<td>14 days</td>
</tr>
<tr>
<td>Knight Diagnostic Laboratories (Portland, OR)</td>
<td>GeneTrails Solid Tumor Panel</td>
<td>37</td>
<td>Hot-spot mutations</td>
<td>10-14 days</td>
</tr>
</tbody>
</table>

CNVs indicates copy number variations; SNVs, single nucleotide variations or point mutations. Gene content is subject to change with additional content added over time.
including prostate cancer, ovarian cancer, breast cancer, germ cell tumors, or neuroendocrine cancers, all of which have established effective therapies. However, most CUP patients have tumors with poorly differentiated histology, limited markers, and no clear evidence of primary tumor origin. Consequently, empiric chemotherapy has been the standard of care but generally produces poor survival outcomes. Several assays have been developed to classify the tissue of origin on the basis of mRNA or miRNA expression signatures and may aid in choosing chemotherapy. More recently, NGS-based genomic testing assays have revealed potentially actionable genomic alterations in patients with CUP that could more directly guide selection of a targeted therapy.

**Design of Clinical Trials**

Genomics-based classification of cancer has changed the outlook on how clinical trials are designed. Traditionally, an investigational agent is developed for a specific cancer type, such as lung or breast cancer. Because these cancers can now be characterized and split into different molecular subsets, there is a rationale for enrolling patients for treatment based on molecular eligibility. As an example, for early trials of BRAF inhibitors in melanoma, the initial phase 1 trial included patients with any solid tumor to determine dosing and assess toxicity, and the investigators observed that patients with **BRAF** V600E-activating mutations were more likely to respond. In the subsequent expansion phase, only patients with **BRAF** mutations were enrolled, and 80% of patients had a response based on the overall response rate. Similarly, molecular eligibility with **ALK** gene fusions was a requirement for entry into early trials of **ALK** inhibitors in lung cancer.

A special challenge for conducting such clinical trials is that the number of eligible patients with molecular eligibility is dramatically reduced, and the traditional approaches for statistical trial design may not be feasible. For example, a clinical trial for a mutation with prevalence of 1% in a common cancer type will be difficult to accrue and complete. In contrast, this may in fact represent an advantage, as a molecularly enriched trial may be more likely to have patients who respond to therapy and may display a greater magnitude of response. Traditionally, patients are randomized to receive the targeted therapy and either the previous best standard therapy or a placebo, but there may be insufficient numbers of patients to complete these trials. Alternative endpoints, such as the response rate and the magnitude of response, may be necessary for rare mutations or rare cancers, such as the nonrandomized trial for imatinib as a tyrosine-protein kinase **KIT** inhibitor in patients with gastrointestinal stromal cell tumors. Meeting the demands of trial accrual for patients with rare mutations may be partially accomplished via screening across many clinical sites and multicenter trials through networks like the National Cancer Institute (NCI) and its cooperative groups. One limitation for multicenter trials is the substantially increased regulatory cost, and reduced funding, for institutions to open trials for potentially enrolling only 0 to 2 patients per year and difficulty in implementing complex correlative studies within those trials. Nevertheless, there are several examples in this regard for lung cancer. The Lung Cancer Mutation Consortium developed over a dozen pathway-based trials for lung cancer across 16 clinical sites (golcmc.com; accessed July 20, 2015). Selected trials may ultimately be more feasible than others toward meeting accrual goals based on gene or mutation prevalence. More recently, the LungMap trial for squamous cell carcinoma similarly included pathway-based trials for **FGFR**, phosphoinositide 3-kinase, cyclin-dependent kinase pathways, and an immunotherapy arm for patients who lack an actionable driver mutation (lung-map.org; accessed July 20, 2015).

Another emerging approach for trials is exclusively mutation-based and pathway-based eligibility and, thus, is truly tumor-site agnostic. One example is a so-called “basket trial” for patients with any solid tumor that has alterations in **FGFRs**, including point mutations, amplifications, or fusions. In this phase 2 study, patients receive an oral pan-**FGFR** inhibitor (ponatinib), and the endpoints are to identify clinical responders in disease or mutation subsets to guide future trials and drug development (NCT02272998). The NCI has recently laid out a strategic plan for precision medicine trials, including the Molecular Analysis for Therapy Choice (MATCH) program. The NCI-MATCH trials can be opened at any NCI-designated cancer center and entail centralized tumor testing for each patient at one of four genomic testing laboratories and multiple trials each with eligibility for a targeted therapy that is mutation-based. While a majority of trials are tumor-site agnostic, some trials could be focused on a mutation pathway in a disease group, such as mTOR signaling in genitourinary cancers.

**Learning From Exceptional Responders in Trials**

In addition to prospective trials that match patients to targeted therapies based on the mutations in their cancer, other efforts from clinical trials are learning from rare patients who experience an exceptional response to a therapy, but the mechanism for that response is unknown. In this approach, clinicians make phenotypic observations in rare patients who have complete responses to a therapy for their metastatic disease, and retrospective genomic and transcriptome sequencing of the patient’s archival tumor can potentially reveal the underlying biology. For example, Iyer et al evaluated a patient with metastatic bladder cancer who had...
a sustained complete response to an mTOR inhibitor clinical trial, while the majority of patients in that trial did not respond. Those authors performed whole-genome sequencing on the patient’s archival tumor and identified a point mutation in TSC1, a tumor-suppressor gene that negatively regulates mTOR signaling. Subsequently, they identified other patients who had TSC1 mutations who were more likely to have response to mTOR inhibition. In another example, Wagle et al observed an exceptional response in a patient with metastatic urothelial cancer who was receiving an mTOR inhibitor and identified alterations in another gene, specifically, activating point mutations in MTOR. Each of these exceptional responder evaluations has now identified mutations that represent new treatment hypotheses for the development of mTOR inhibitors in patients with mutations in the MTOR and TSC1 genes. On a national level, the NCI has established an Exceptional Responders Initiative to use genomic sequencing to facilitate drug development for advanced cancer. The mission of the NCI Exceptional Responders Initiative is to identify and confirm patients who have had remarkable responses to systemic therapy and use genomic technologies to characterize their tumors to study the molecular mechanisms underlying why these patients benefit from systemic therapy, particularly chemotherapies.

Challenges and Opportunities

Although the new framework for precision medicine in cancer has great promise, the full realization of this approach has several challenges as well as opportunities.

 Beyond the molecular characterization of cancer through whole and transcriptome sequencing, there are additional complexities for cancer biology to consider. First, basic science research on epigenetic alterations are continuing to reveal how methylation regulates gene expression and can contribute to our view of individual cancers. Second, new omics approaches for the proteome and the metabolome can reveal additional layers to cancer biology. As new “omics” approaches became more practical, we should consider new standards for integrating data analysis and transparency in clinical trials, as recommended by the Institute of Medicine’s recent assessment of translational omics. Third, additional aspects of cancer biology, including tumor heterogeneity, mechanisms of drug resistance, the tumor microenvironment, and stem cell properties of cancer, can influence how patients respond.
to therapy. Together, these challenges can be met with new opportunities created by investment in science. Beyond genomics-guided therapy, we envision combination therapies with other modalities, including immunotherapy, oncolytic viruses, and stem cell/metabolism-targeting inhibitors. For immunotherapy, genomic sequencing approaches can be applied to identify the burden of neotumor antigen or molecular defects in DNA repair, such as mismatch repair genes, which may predict response to novel therapies that inhibit immune-regulatory checkpoints to boost the immune response against cancer.119-121 These challenges can be met through a collaborative network of innovative clinical trials with systematic collection of tumor tissue, clinical data, and transparency (Fig. 3).

Future Directions
Integrative profiling through DNA and RNA sequencing opens new doors for both basic and clinical cancer research. Molecular classification of cancer based on genomic and transcriptome alterations may reveal novel biomarkers for diagnosis, prognosis, and predicting response to therapies. Translating the cancer genome and transcriptome for patients will require continued multidisciplinary collaboration between oncologists, pathologists, basic scientists, and computational biologists (Table 2). Additional resources and funding are necessary to support the ongoing profiling efforts for basic genomics research, tumor sequencing in the clinic, and data-sharing networks to enable precision cancer medicine.

References
14. Koenig RJ, Peterson CM, Jones RL, Sauder C, Lehrman M, Cerami A. Correla-


