Clinical Chemistry 59:1
000–000 (2013)

Reviews

Translating Genomics to the Clinic: Implications of Cancer Heterogeneity

Nardin Samuel[1,2] and Thomas J. Hudson[1,2,3*]

BACKGROUND: Sequencing of cancer genomes has become a pivotal method for uncovering and understanding the deregulated cellular processes driving tumor initiation and progression. Whole-genome sequencing is evolving toward becoming less costly and more feasible on a large scale; consequently, thousands of tumors are being analyzed with these technologies. Interpreting these data in the context of tumor complexity poses a challenge for cancer genomics.

CONTENT: The sequencing of large numbers of tumors has revealed novel insights into oncogenic mechanisms. In particular, we highlight the remarkable insight into the pathogenesis of breast cancers that has been gained through comprehensive and integrated sequencing analysis. The analysis and interpretation of sequencing data, however, must be considered in the context of heterogeneity within and among tumor samples. Only by adequately accounting for the underlying complexity of cancer genomes will the potential of genome sequencing be understood and subsequently translated into improved management of patients.

SUMMARY: The paradigm of personalized medicine holds promise if patient tumors are thoroughly studied as unique and heterogeneous entities and clinical decisions are made accordingly. Associated challenges will be ameliorated by continued collaborative efforts among research centers that coordinate the sharing of mutation, intervention, and outcomes data to assist in the interpretation of genomic data and to support clinical decision-making.

The notion that chromosome defects cause cancer was first proposed in the early 1900s *(1)*. Definitive evidence that alterations in chromosomal subunits play a role in cancer was obtained from studies of genes that control cancer growth, and discoveries of the transforming capacity of retroviruses in the 1970s established that virtually all cancers arise from the disruption of genetic material. The seminal finding that cancer is a disease of the genome initiated the pursuit of cancer-causing genes, which continues unrelentingly to this day. In contrast to earlier approaches that discovered cancer genes one at a time, however, the advent of large-scale surveys of the genome for systematically identifying cancer genes became a turning point in cancer research. Systematic and scalable methods for nucleic acid analysis—genome sequencing—have enabled the study and characterization of thousands of cancer genomes to date *(2)*.

Effectively translating cancer research into clinical practice is a major challenge for the field. Data obtained from sequencing tumor genomes are producing an explosion in the content and complexity of cancer mutation databases. Interpreting the immense volume of data generated from sequencing studies is benefiting from the continuing development of bioinformatics approaches that facilitate identification of relevant genetic changes. Consequently, many new cancer genes and potential targets for drug development have been identified. Today, the pursuit of diagnostic markers and therapeutic interventions is far more informed by knowledge of a cancer's cellular mechanisms than it was in the pregenomics era; thus, the time between a scientific discovery and its clinical application is becoming much shorter *(3)*.

Despite the major strides in furthering our understanding of the cancer genome, definitive treatments and cures remain elusive when the tumor mass cannot be surgically resected. Among the many causes of tumor progression and disease recurrence in the setting of chemotherapy or radiation therapy, one of the major ones is tumor heterogeneity. The concept of tumor heterogeneity encompasses multiple classifications, including genetic heterogeneity among samples from patients with the same tumor type and the presence of genetically distinct subpopulations of tumor cells within an individual patient's tumor. The latter, termed "intratumoral heterogeneity," is highly relevant to studies of cancer genomics. The complexity underlying intratumoral heterogeneity complicates can-

[1] Ontario Institute for Cancer Research, Toronto, Ontario, Canada; Departments of [2] Molecular Genetics and [3] Medical Biophysics, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada.
* Address correspondence to this author at: Ontario Institute for Cancer Research, 101 College St., Suite 800, Toronto, Ontario, M5G 0A3 Canada. Fax 416-673-6650; e-mail tom.hudson@oicr.on.ca.
Received August 3, 2012; accepted October 22, 2012.
Previously published online at DOI: 10.1373/clinchem.2012.184580

cer genome sequencing efforts because of the inherent difficulties in distinguishing cells with different genetic characteristics (subclones) derived from a malignant parent clone and distinguishing the relevant mutations among these clones.

In this review, we use the example of breast cancer to outline not only some aspects of the recent progress in cancer genomics but also perspectives for the clinical application of genomics research for personalized medicine, given the immense complexity and heterogeneity of cancer genomes.

Next-Generation Sequencing Enables Comprehensive Study of Cancer Genomes

The advent of next-generation sequencing (NGS)[4] technologies, which grew exponentially in the decade after publication of the first iteration of the human genome sequence *(4)*, has provided substantial insights into new genes and the biological processes that underlie cancer pathogenesis. These insights are outlined below. NGS technologies "parallelize" sequencing processes via high-throughput means to produce millions of short sequencing "reads" from amplified DNA clones *(5)*. NGS is also referred to as "massively parallel sequencing," because the reaction steps occur in parallel with the detection steps and millions of reactions occur simultaneously *(6)*. This parallelism makes it possible to read the same segment of a DNA sequence repeatedly to increase confidence in the sequence obtained for the targeted genomic segment. This multiple sampling of a genomic segment is referred to as the "coverage" of the sequencing run.

Before the NGS era, much progress had been made toward identifying mutated cancer genes and cellular processes. Array-based genome analysis techniques facilitated the identification of clinically important cancer-related copy number aberrations, such as deletions and amplifications. First-generation capillary-based sequencing, or Sanger sequencing, was the main technology used in the initial sequencing of the human genome and was used to detect important alterations in the coding sequences of cancer genes. Sanger sequencing, however, was deemed too labor-intensive and costly to allow extensive genome analyses of large numbers of tumors. The costs of NGS technologies have been decreasing rapidly, however *(7)*, and large-scale global sequencing of entire cancer genomes, coding sequences (exomes), expressed mRNA transcripts (transcriptomes), and regulatory chemical modifications to DNA (epigenomes) has become feasible and affordable. NGS technologies are now shaping cancer research. The costs and times associated with sequencing runs are expected to continue to decline, thereby rendering whole-genome sequencing of thousands of cancer samples for unbiased detection of genetic aberrations an achievable reality.

In addition to recent improvements in NGS platforms, new "third-generation" sequencing technologies have also emerged. These technologies detect the binding of nucleoside triphosphates to the polymerase in real time [PacBio (Pacific Biosciences) *(8)*, nanopore (Oxford Nanopore Technologies)] and allow the sequencing of nucleic acids from single molecules, thereby circumventing prior DNA amplification and labeling in the library-preparation steps. Another, newer approach, Ion Torrent technology, is based on detecting hydrogen ions that are released as the nucleotides are incorporated into a growing DNA strand *(9)*. Although the advantages of third-generation sequencing methods include low instrument and run costs, longer read lengths, and shorter run times, they have their limitations and have not yet entered the mainstream for sequencing entire cancer genomes *(10)*.

Although the costs associated with whole-genome sequencing have declined substantially since the introduction of NGS approaches, whole-genome sequencing of large numbers of tumor samples initially had not been a routine part of cancer genome research. The substantial physical coverage associated with surveying the 3×10^9 bases of an entire genome requires 90 Gb of sequence to obtain the minimum 30-fold coverage that is required *(11)*. Considerably higher sequence coverage (up to 75-fold) can be achieved by targeted exome sequencing, given that exons constitute approximately 1% of the genome *(12)*. That fact has allowed nearly countless studies of only the coding fractions of cancer genomes (or exomes) and their publication.

Efforts to construct complete catalogues of the mutational spectrum obtained from sequencing the genomes of numerous cancer types require substantial coordination and collaboration by multiple institutions. That is necessary to garner conclusions with the most impact, standardize research protocols and approaches, and reduce duplication of effort. Indeed, these goals and mandates are central to such international collaborative cancer research efforts as the International Cancer Genome Consortium *(2, 13)* and the Cancer Genome Atlas *(14)*. The goals of these consortia are to elucidate the full mutational profiles of cancer genomes and to make sense of these data by coupling them to transcriptomes, epigenomes, and clinical correlates to identify prognostic markers or potential therapeutic targets.

[4] Nonstandard abbreviations: NGS, next-generation sequencing; AML, acute myeloid leukemia.

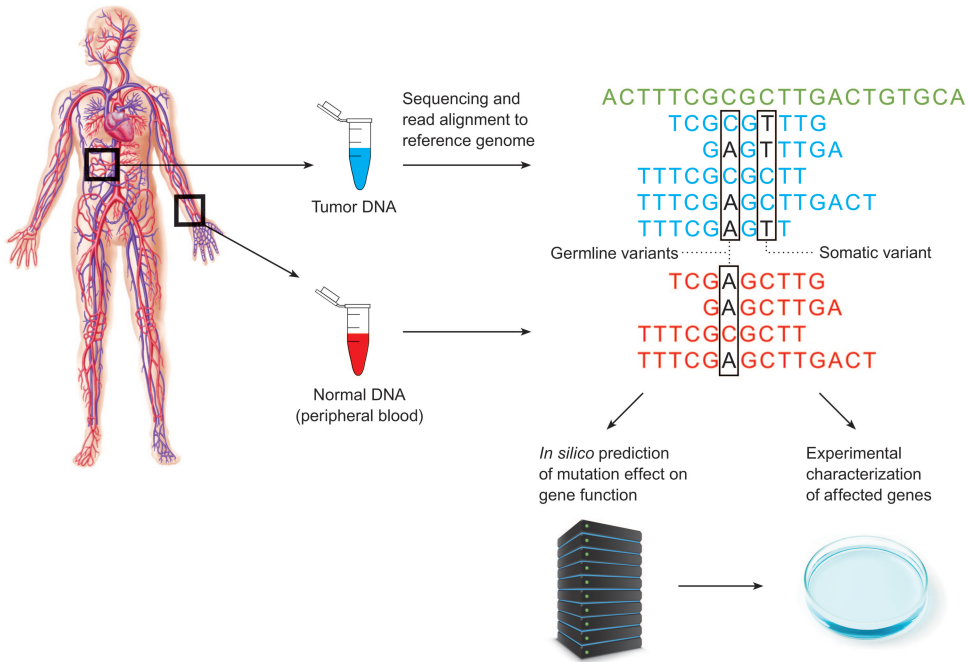

Fig. 1. Whole-genome analysis of tumor and matched normal DNA.

The basic pipeline of analysis involves isolation of DNA (or RNA) from tumor tissue, as well as germline DNA from normal tissue, such as peripheral blood (other nonmalignant tissue can be used). DNA is then analyzed with NGS or third-generation sequencing platforms, and computational algorithms are used to map sequencing reads to the human reference genome (green). Variants observed in both tumor and normal DNA are germline variants, whereas those observed only in the tumor sample are inferred to be somatic, or acquired, variants. The figure depicts DNA derived from normal cells to be heterozygous at the varying base. Consequently, some alleles are the same as the reference sequence (C), and others are variant (A). Functional consequences of somatic mutations on protein function can be predicted by *in silico* modeling and/or studied experimentally.

Finally, although NGS has the power to uncover the full spectrum of mutations in a cancer genome, what cannot be ignored is that although sequencing yields immense amounts of mutation information, only a small fraction of the mutations detected may be relevant and contribute to the tumor process. That is, it is important to distinguish mutations that occur in a tumor but play no role in the neoplastic process (so-called passenger mutations) from driver mutations that influence cancer progression and evolution. Although there are bioinformatics-based approaches that can assist in identifying probable driver genes, extensive experimental validation is the benchmark for establishing an oncogenic role for a putative cancer gene.

Insights into the Cancer Genome Gained from NGS

Although large-scale sequencing of tumor genomes is providing comprehensive catalogues of mutations in cancer genes and other mutational processes, translating these findings into clinically useful parameters and therapeutic targets will require a greater understanding of the cancer cellular processes that are deregulated by such mutations. It is therefore imperative to harness the information of mutations in cancer genomes to yield mechanistic insights into malignant processes.

Whole-genome NGS surveys of tumor samples and matched normal samples from the same individual afford immense power for exploring the cancer genome. Although inherited germline variants can undoubtedly play a role in an individual's susceptibility to cancer, somatic mutations (i.e., newly acquired genetic aberrations) constitute the majority of genetic mutations that cause cancer *(15)*. Assessing the somatic-mutation burden in a tumor sample by identifying cancer-specific mutations that are absent in the germline enables vast investigation of that cancer's genome (Fig. 1). This approach was first established via sequencing of a tumor sample from a patient with acute myeloid leukemia (AML) and a matched normal sam-

ple *(16)*. This unbiased survey of the genome identified mutations in 2 well-known AML-associated genes, as well as 8 other somatic mutations that had not previously been detected in AML. This study set the stage for the use of whole-genome sequencing as a robust and powerful tool for studying the cancer genome. Since this study was undertaken, numerous tumors have been analyzed by whole-genome sequencing, and many new cancer genes have been identified. Comprehensive sequencing of cancer genomes has shed light on some fascinating cellular processes that cancer cells depend on to maintain their malignant phenotype. Because not all driver mutations in a cancer are "druggable" or can otherwise be targeted therapeutically, an improved understanding of the cellular mechanisms and pathways that are deregulated can certainly aid in developing novel therapeutic strategies that would abrogate oncogenic signaling. The following representative set of examples illustrates some of the interesting insights that are possible through the sequencing of cancer genomes.

HIGHLY COMPREHENSIVE INVESTIGATION OF BREAST CANCER GENOMES

A recent torrent of data from analyses of breast carcinoma genomes has led to remarkable insights into the landscape of this complex disease; these studies represent the most comprehensive genome-sequencing analyses of any tumor type to date. A fascinating observation from comparing these studies is that several have generated an in-depth analysis of one or more features of cancer genomes, which we briefly discuss below.

Breast cancer genomes often bear distinctive mutation signatures. To gauge the DNA damage and repair processes that underlie the acquisition of somatic mutations in breast cancer, one study generated whole-genome coverage for 21 breast cancers *(17)*. The mutational signatures describe the relative contribution of each of 6 classes of base substitutions (C>A, C>G, C>T, T>A, T>C, and T>G) to the complete spectrum of mutations observed. Five distinct mutational signatures were observed from the repertoire of somatic mutations in the breast cancer genomes. The results indicate that 5 independent mutation processes steer the different mutation patterns observed in the 21 breast cancers. One mutational signature, signature B, which is characterized by C>T, C>G, and C>A substitutions at TpCpX trinucleotides (where X is any base and C is the base mutated to T, G, or A), was responsible for a substantial majority of the mutations in some cancer samples. Among its other findings, this study is the first to find that the genomes of breast carcinomas bear regional clusters of hypermutation, termed "kataegis," and that these regions display a prevalence of signature B mutations. Intriguingly, kataegis is associated with chromosomal rearrangements, indicating that these processes may occur in close temporal proximity.

Subclonal diversification over the lifetime of breast cancers leads to most mutations being found in just a fraction of cells. In the same study, tumor clonal evolution was investigated by sequencing 1 sample to a substantially higher depth (188-fold coverage). An analysis that applied a read depth–based algorithm revealed that point mutations and chromosomal rearrangements accumulate throughout the lifetime of a tumor *(18)*. Mutations in driver genes are presumed to occur first in the evolution of the tumor and to lead subsequently to genomic instability. Tumor subclones are later formed through kataegis and/or catastrophic chromosome shattering and aberrant rejoining, a phenomenon referred to as "chromothripsis" *(19)*. Rate-limiting mutations successively shape the dominant subclone, thereby causing malignant growth. This clone can constitute up to half the tumor bulk *(18)*.

Most recognized driver mutations in breast cancer occur at low frequency, even within groups of patients having the same cancer subtype. Variation in the somatic-mutation burden among tumors was reported for a study of 100 breast tumors, along with 9 new cancer genes that had low-frequency mutations *(20)*. Moreover, a subset of the breast cancers analyzed bore no recognizable driver mutations. A substantial number of low-frequency mutations were also observed in a study of estrogen receptor–positive breast cancers *(21)*. These findings underscore the challenge of detecting and validating new driver mutations and then correlating them to parameters such as expression and clinical features.

The paradigm of targeted therapies holds promise when distinct subgroups are defined. A study of mutations across breast cancer subtypes identified recurrent aberrations in the *CBFB*[5] (core-binding factor, beta subunit) gene, deletions in its partner *RUNX1* (runt-related transcription factor 1), and fusion of the *MAGI3* (membrane associated guanylate kinase, WW

[5] Human genes: *CBFB*, core-binding factor, beta subunit; *RUNX1*, runt-related transcription factor 1; *MAGI3*, membrane associated guanylate kinase, WW and PDZ domain containing 3; *AKT3*, v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma); *MAGI3-AKT3*, fusion of genes *MAGI3* and *AKT3*; *KRAS*, v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog; *CTNNB1*, catenin (cadherin-associated protein), beta 1, 88kDa; *SMARCA4*, SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4; *CREBBP*, CREB-binding protein; *WNT*, wingless-type MMTV integration site family; *BRAF*, v-raf murine sarcoma viral oncogene homolog B1.

and PDZ domain containing 3) gene with the *AKT3* [v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)] gene (i.e., *MAGI3-AKT3*), which is most prevalent in triple-negative breast cancer *(22)*. This study also demonstrated that AKT inhibitors might be a promising therapeutic intervention for *MAGI3-AKT3* fusion–positive triple-negative breast cancers *(22)*. Similarly, sequencing analyses of estrogen receptor–positive breast cancers revealed distinct subtypes of this disease, each bearing a unique mutational signature and deregulated cellular pathways *(21)*. Therefore, targeted therapies tailored to the mutational features of each subtype of this disease are likely to be more beneficial than treating all estrogen receptor–positive tumors as a single entity.

Taken together, the evidence produced in the current era of large-scale genomic analyses with NGS has led to remarkable insights into the pathogenesis of numerous major histologic types of tumors. It is important to consider, however, that genes can act in specific contexts, depending on the varying germline of the host and the microenvironment; thus, the functional impact of mutations should be studied in the context of the molecular subtypes of tumors, when possible. Many driver mutations are detected at low frequency, however. Such findings indicate that molecular subtyping may be highly complex and that the molecular consequences of these mutations will need to be understood in the context of other drivers. Although the time needed to translate such findings into consequential clinical applications certainly depends on the steps required to analyze and interpret these findings and then to introduce them to the clinic, the large-scale analysis of cancer genomes is an indispensable first step toward this goal.

Tumor Complexity and Heterogeneity from a Genomic Perspective

As tumor samples continue to be sequenced and mutations in cancer genes are identified, the challenges of tumor heterogeneity will need to be addressed to make meaningful sense of cancer genome data. As outlined above, only 7 years have passed since the introduction of the first NGS platform. Countless studies have been undertaken to analyze and characterize cancer genomes, and these studies have been fruitful in uncovering a large number of aberrant molecular events that shape cancer genomes. The major factor of tumor heterogeneity has not always been addressed adequately, however, and that has important implications for the clinical application of genomic findings. The heterogeneity of cancer genomes spans numerous levels of complexity, as we describe below.

Table 1. Terms and definitions.

Intertumoral heterogeneity	Differences in genetic and molecular characteristics of tumors from different patients
Intratumoral heterogeneity	Differences in genetic and molecular characteristics of tumor cells derived from an individual patient's tumor
Tumor subclone	Subpopulation of tumor cells present in a bulk tumor that differs in mutational spectrum and genetic characteristics from other clones
Tumor cellularity	The proportion of tumor cells in a bulk tumor among the populations of other, nonmalignant cell types, such as normal-tissue cells, fibroblasts, lymphocytes, and others
Personalized medicine	A paradigm of tailoring clinical decisions and therapies to individuals on the basis of the characteristics of the individual's disease, such as its genetic features
Clonal diversification	Ongoing evolution of tumor cells to give rise to additional subclones that possess new mutations, some of which may confer heightened growth advantages or potential to seed metastases

INTERTUMORAL HETEROGENEITY

Intertumoral heterogeneity, or heterogeneity among patient tumors, is defined by the molecular and histologic subtypes within a given tumor type (Table 1). As we have illustrated above, intertumoral heterogeneity is markedly prominent in breast cancers and therefore highly relevant for this disease (as well as other cancers), because tumor subtypes may display different rates and spectrums of mutations.

Whereas certain mutations are characteristic of nearly all samples taken from a tumor of a given type—such as *KRAS* (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog) mutations in pancreatic cancer, which occur in >90% of pancreatic tumors *(23)*—certain mutations may define tumor subtypes. This subtype-specific mutation patterning has been observed for many tumor types, including childhood medulloblastoma. In this disease, mutations that affect genes in cellular pathways such as histone methylation occur in only 2 of 4 discrete medulloblastoma subgroups *(24)*. Mutations in genes associated with the *CTNNB1* [catenin (cadherin-associated protein), beta 1, 88kDa] gene—such as the *SMARCA4* (SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4) and *CREBBP* (CREB-binding protein) genes—reportedly occur in one of the 4 subgroups: tumors in the *WNT* (wingless-type MMTV integration site family) subgroup *(24)*. This finding highlights the importance of delineating

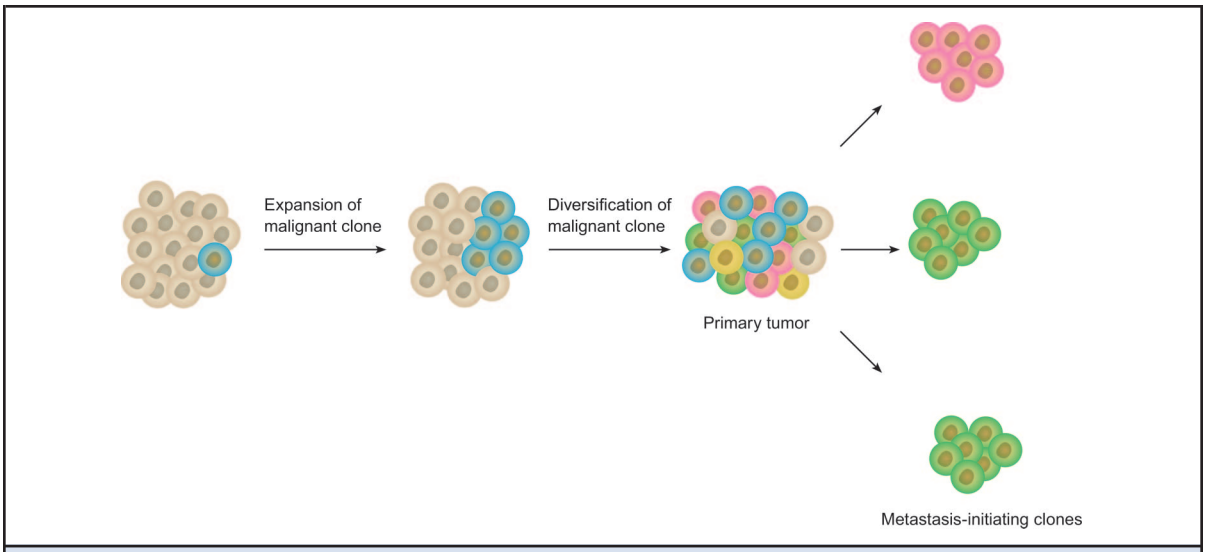

Fig. 2. Clonal evolution of heterogeneous tumor cells.
A malignant parent tumor clone (blue) amid normal cells (beige) expands and can undergo mutational processes to acquire new mutations and subsequently give rise to additional clones (pink, green, yellow). A fraction of the tumor clones (green, pink) can seed metastases in nearby and/or distant organs.

different histologic and molecular subgroups when possible, because the mutations governing tumorigenesis differ and research and subsequent therapy must be guided accordingly. Given this level of heterogeneity, classification based on pathology alone is insufficient.

The numerous landmark sequencing studies of breast cancer genomes described in the previous section have revealed intertumoral heterogeneity to be an important feature of this disease. Breast cancer is more complex and heterogeneous than the discrete clinical subtypes that have historically been defined. A recent survey of nearly 2000 breast cancer samples demonstrated that surveying substantially large cohorts of this tumor type can uncover novel subtypes that downstream investigations and therapy-development studies reveal to be molecularly distinct diseases *(25)*. A discovery cohort of primary breast tumors used for sequencing and transcriptome analysis uncovered novel subgroups, each of which was associated with distinct clinical outcomes. The findings were substantiated in an equally large validation cohort. Additionally, low-frequency mutations occur in breast cancers—as in other cancers—and most recurring mutations are rare *(20, 21)*. Marked intertumoral heterogeneity was revealed in a survey of 104 cases of primary triple-negative breast cancer *(26)*. Tumors in this group of breast cancers appear to differ substantially from each other with respect to the number of mutations detected, the type of mutations, and the clonality of tumor samples *(26)*. These findings demonstrate that for breast cancer (and likely for the majority of cancers) no 2 tumors are mutationally identical and that stratifying tumors by mutational class is far more complex than previously thought.

INTRATUMORAL HETEROGENEITY

An additional level of heterogeneity to consider in cancer genomics research efforts is intratumoral heterogeneity, which can be defined as the presence of nonidentical cellular clones or subclones of tumor cells within a bulk tumor (Table 1). This heterogeneity can be attributed, in part, to the ongoing and parallel evolution of cancer cells. A parent clone (or the index malignant cell in a tumor) can give rise to numerous subclones that differ in their mutational architecture and growth potential (Fig. 2). Some subclones may have a considerable growth advantage and come to dominate the bulk of the tumor volume, whereas a minor subpopulation of clones may be capable of seeding metastases at sites in other organs. The bulk tumor therefore consists of subpopulations of distinct cellular clones, the composition and characteristics of which may have substantial implications for biomarker discovery and application, and for drug development.

The intratumoral genetic heterogeneity of cancers has been known for decades. The initial studies used fluorescence in situ hybridization techniques to uncover different intratumoral patterns of structural chromosomal aberrations. Early observations of intra-

tumoral heterogeneity in human cancers were frequently for breast cancers *(27)*, gliomas *(28)*, bone and soft-tissue sarcomas *(29)*, and pancreatic cancers *(30)*. Recent analyses of clear-cell renal cell carcinoma have illuminated intratumoral heterogeneity in the context of cancer genome sequencing *(31)*. Intratumoral heterogeneity can arise by numerous scenarios, which can affect the interpretation of cancer genome sequencing data. The first consideration, as demonstrated in the study of this disease by Gerlinger and colleagues, is that biopsy samples from patient tumors may not adequately represent the cellular architecture and composition of the bulk of the tumor *(31)*. Consequently, sampling small tumor regions that are not representative of the entire disease may hamper sequencing studies of cancer genomes that are aimed at biomarker identification. This possibility has implications for many of the sequencing studies that have already been undertaken, because the lack of adequate sampling may yield only a partial understanding of any genomic data that are produced.

Intratumoral heterogeneity is also relevant in the common instance in which only a minor subclone is capable of seeding metastasis. Therapeutically targeting mutations in a dominant subclone that has been identified by NGS sequencing may treat the primary tumor effectively, but it may not impede metastasis. This bicompartmental attribute of cancer, in which the primary disease is biologically divergent from the metastatic disease, was recently demonstrated for medulloblastoma. Clonal genetic changes in metastatic lesions of medulloblastoma occur only in a rare subpopulation of clones in the primary tumors. This finding bolsters the notion that only minor subclones can give rise to metastases, but it also emphasizes the importance of accounting for this bicompartmentalization when developing therapies *(32)*. Similarly, as metastatic clones evolve to acquire additional mutations and encounter new microenvironmental niches, the driver mutations in distant metastatic lesions may actually become distinct from those of the primary tumor. That has been shown for pancreatic cancer in a study that sequenced the DNA of primary tumors and metastases to annotate and describe clonal relationships among metastases and primary tumors. The results revealed genetic heterogeneity among metastasis-initiating cells, as well as distinct driver mutations in these cells that were not detected in the primary tumors *(33)*.

In addition, mutations in genes of minor subclones that occur in the tumor but do not initially confer a selective growth advantage may actually become drivers when the nature of the selective pressure is altered, as when chemotherapy is initiated. In AML, for example, a whole-genome sequencing study of both primary-tumor and relapse genomes identified pronounced patterns of clonal evolution in the patients. For example, a founding clone in the primary tumor could gain mutations to become the relapse clone, or a subclone of the founding clone that had not been eradicated by therapy might gain additional mutations and thereby expand the lineage at relapse *(34)*. Chemotherapy did not ablate the founding malignant clone in any of the cases of this study, yet it altered the selective pressures, inducing new mutations and expanding the clonal evolution that engendered disease relapse.

Increasing the degree of sequencing coverage in NGS to assess and characterize both nonmalignant cell populations and tumor subclones can facilitate the recognition of tumor subclones. A comprehensive analysis of the temporal clonal evolution of breast cancer genomes sequenced to approximately 40-fold coverage found only a 5% chance of identifying a clonal mutation occurring in 25% of the tumor samples *(17)*. These results underscore the fact that the existence of intratumoral heterogeneity requires that inferences of clonality require that whole genomes of tumors must be sequenced at a much greater depth than that used to identify base-level mutations present in the majority of subclones (Fig. 3). Additionally, integrated computational approaches are indispensable for inferring clonality from sequencing data. One recently published algorithm, ABSOLUTE, demonstrates that quantifying copy number changes and point mutations identified in sequencing studies at an absolute level, rather than at a relative level, allows inferences about the subclonal architecture of tumors *(35)*. As cancer genomes continue to be sequenced, such algorithms will aid substantially in interpreting sequencing data and in drawing inferences regarding the important feature of intratumoral heterogeneity.

The development of single-cell sequencing approaches is another avenue for aiding in our quantitative understanding of intratumoral heterogeneity. Characterizing the genomic features of individual cells—rather than a mixed population of tumor cells (or adjacent nontumor cells)—helps in resolving the mixtures of genetically distinct cells in a bulk tumor. Additionally, this approach may provide genetic information about most malignant tumor cells, which could represent a minority of the tumor cell population *(36)*. The first study of so-called single-nucleus sequencing used single nuclei from breast cancers and performed low-coverage sequencing to characterize intratumoral DNA copy number variation *(37)*. Since this study was undertaken, additional studies of single-cell exome sequencing of human tumors (specifically, clear-cell renal cell carcinoma and a myeloproliferative neoplasm) have explored the potential capability of single-cell genomics *(38, 39)*. Single-cell genomic analyses, cou-

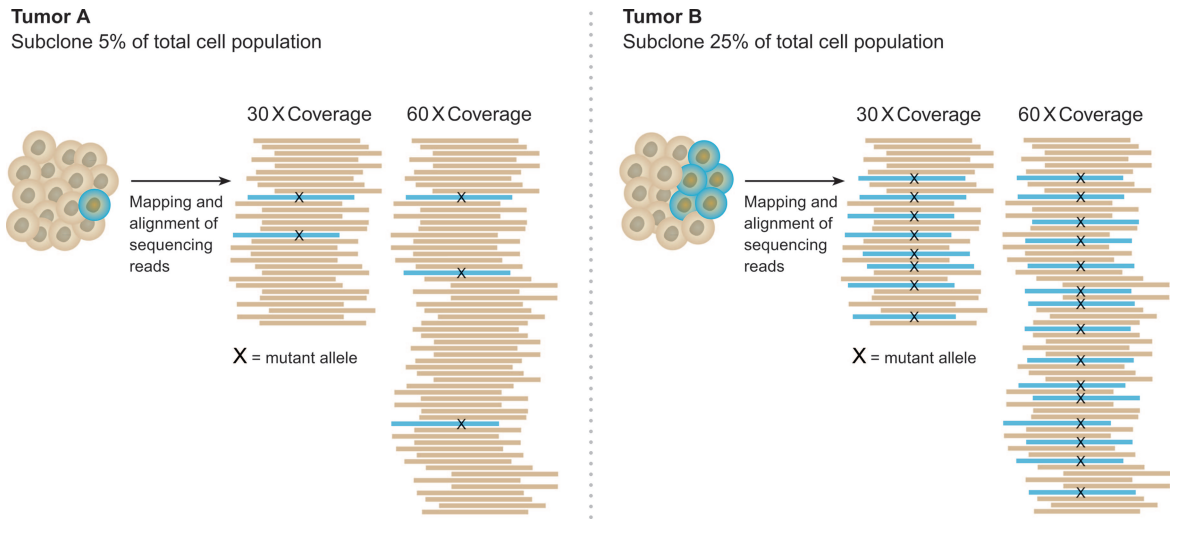

Fig. 3. Identification of minor tumor subclones by increasing sequencing depth.

In the simple situation in which 1 minor tumor subclone (blue) coexists with a dominant population of major clones (beige), a sequencing coverage of 30-fold (30×) may not detect the presence of a mutant allele, because most of the sequencing results will likely be germline in origin. Increasing coverage (e.g., to 60×) may produce sufficient reads to detect mutations derived from the minor subclone; however, robust analysis is required to ensure that the detected variant is not due to sequencing error. As depicted, the larger the proportion of the minor subclone, the better the odds are for detecting cancer-specific mutations.

pled with transcriptomic and proteomic profiling at the single-cell level, will undoubtedly provide a deeper view of the genetic diversity within tumors.

TUMOR CELLULARITY

Lastly, bulk tumors almost always consist of tumor cells exclusively, but they are usually intermixed with some fraction of nonmalignant cells, such as fibroblasts or lymphocytes. This cellular heterogeneity within a tumor is referred to as "tumor cellularity" (Table 1). Bulk tumors of pancreatic ductal adenocarcinomas, for example, are highly heterogeneous, with up to 90% of the cells consisting of nonmalignant cells commonly referred to as "stromal cells" *(40)*. Consequently, DNA and RNA isolated from bulk tumors typically are mixtures of tumor and normal DNA and RNA, which pose a challenge for detecting somatic mutations (i.e., the false-negative rate will increase because the effective coverage of reads of mutations will decrease). Although other approaches, such as immortalizing tumor cells in culture or producing xenografts, can help to decrease infiltration by populations of normal cells, these methods have their limitations. When the tumor can be surgically resected, laser capture microdissection is an ideal approach for extracting tumor cells from bulk tumors. Additionally, a substantially high level of sequencing coverage may be necessary to detect tumor-specific aberrations in the presence of contamination by nontumor nucleic acids *(35)*. Finally, algorithms have been developed to infer tumor purity from sequencing data, and these algorithms have been validated with high accuracy *(35, 41)*.

Clinical Perspective and Considerations

The ultimate goal of cancer research is to better comprehend the biological processes of cancer so that the many dimensions of patient care—from diagnosis to prognostic markers to therapy—can be improved. Unquestionably, cancer genomics can play a role toward these ends; however, prudent and well-designed studies are necessary for the power of cancer genomics to be realized to its fullest potential. In addition, consideration of the patient should be paramount at each step of development, from research to the clinic. For example, as we describe above, patient samples obtained by biopsies from a single region of a tumor are insufficient for inferring the complete clonal architecture of that tumor.

In addition to the role of cancer genomics in the discovery and understanding of cancer-causing mutations, genome sequencing has immense utility in the clinical setting as a diagnostic and prognostic tool. That is the promise of the new era of personalized medicine, in which therapies are directed at the specific characteristics of a patient's disease (Table 1). Yet, before per-

sonalized medicine can be fully implemented, it is necessary to make sense of the cancer genomic data generated via the myriad cancer genome sequencing studies and how these findings can be implemented to change and enhance patient care effectively. One major consideration is the time and costs associated with implementing cancer genomics approaches into the clinical setting. For example, although conducting whole-genome sequencing may not be practical for patients in an oncology clinic, it may be useful for sequencing a defined panel of genes. If mutations are found, they could be clinically informative and therefore guide therapeutic decisions. In addition, the clinical implementation of whole-genome sequencing has numerous bioethical implications, including those involving the disclosure of incidental findings from a patient's genomic data, among others *(42)*.

An optimal accord between cancer genomics tools and clinical practice essentially involves genetic analysis of patient tumors, interpreting the sequencing data, and making clinical decisions accordingly. This paradigm has been implemented in the Genome Pathway Strategy, a collaboration between the Princess Margaret Hospital – University Health Network and the Ontario Institute for Cancer Research. Through this initiative, patient tumors are molecularly profiled via mutation genotyping and the use of third-generation sequencing platforms. Patients participating in the Genome Pathway Strategy are enrolled in 5 cancer centers in Ontario, Canada. The study is recording the impact of molecular profiling on treatment decisions for these patients *(43)*. Systematic bridging of cancer genomics and clinical decisions will require integrated clinical studies to verify that prognostic and treatment decisions can be optimized by considering the full spectrum of laboratory and clinical information, including the complex genomic features that distinguish each patient's disease.

The presence of tumor heterogeneity is a major challenge in the development and translation of effective therapies for the clinic. For example, mutation in the *BRAF* (v-raf murine sarcoma viral oncogene homolog B1) gene encoding the $BRAF^{V600E}$ variant account for >90% of BRAF mutations in melanoma. Initial clinical trials of BRAF kinase inhibitors showed remarkable responses in patients whose tumors bore this mutation *(44, 45)*; however, a significant proportion of the patients experienced a response duration of only months before disease progression resumed *(44)*. Follow-up studies have revealed that the efficacy of BRAF inhibitors for melanoma depends on the intratumoral heterogeneity and the presence of genetically distinct subclones in the bulk tumor *(46)*. This finding suggests that individual melanomas are polyclonal, consisting of cells that bear the $BRAF^{V600E}$ mutation and some cells that do not, both of which have a metastatic capability *(46)*. This example highlights the importance of quantifying the driver mutations that could become the targets of personalized therapy and the extent to which they represent mutations in a diverse population of polyclonal tumor cells. In the era of personalized medicine, evaluating a tumor's heterogeneity is an imperative consideration if the genomic context of a tumor is to be used to guide therapy. Precision therapies may be effective only when they are coupled with a comprehensive understanding of the clonal and genetic architecture of a cancer and the underlying genetic characteristics of the subclones constituting the tumor.

Looking Ahead

Cancer genome sequencing is revolutionizing this era of cancer research and has led to an enormously rapid rate of discoveries and insights into the mutational landscape of cancer. Yet, it is becoming a challenge to maintain a balance between the magnitude and the rapidity at which these data are being generated, and our ability to analyze and interpret it effectively. The underlying complexity of cancer genomes is a factor contributing to why definitive cures for this disease remain elusive and why this complexity cannot be ignored.

Looking ahead, it is certain that comprehensive catalogues of cancer genomes will continue to be generated. These catalogues will enable more of the genes that drive cancer to be identified and will facilitate the identification of meaningful clinical correlations. The heterogeneity of cancer genomes is a principal factor to consider when interpreting such genomic data, however, because tumors comprise complex populations of malignant cells. Beyond genetics, the elucidation of oncogenic mechanisms governed by different transcriptional or methylation profiles in tumors will also be important for advancing our understanding of cancer biology and in supporting clinical decisions related to the diagnosis, prognosis, and prediction of responses to therapies *(47)*. The full road map of cancer will be complete only when all roads have been discovered and analyzed. That will be possible only by regarding cancer as the heterogeneous entity that it is.

Given that the extensive sequencing efforts carried out to date have revealed that mutations driving cancer are infrequent, there will remain a continuing need to establish and maintain large repositories of cancer genome databases that catalogue these mutations. These databases should also include, when possible, information on disease types, genome data sets, therapies, and outcomes. Such databases can be achieved only through coordinated and effective networking among cancer research centers *(48)*.

Finally, understanding and tackling cancer through the avenue of cancer genomics will require additional analyses outside of the genetic landscape of cancers. Such efforts include integrating genomics with functional data to uncover the deregulated cellular mechanisms and vulnerabilities of cancer cells and then to target them accordingly. The consequences of mutations with respect to tumor biology, metastatic potential, and sensitivity to targeting agents will need to be resolved. Through systematic and collaborative efforts on these fronts, a clear understanding of cancer is well within our reach.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: None declared.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: N. Samuel, Frank Fletcher Memorial Fund and Ontario Institute for Cancer Research; T.J. Hudson, Ontario Institute for Cancer.
Expert Testimony: None declared.
Patents: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References

1. Boveri T. The origin of malignant tumors [English translation]. Baltimore: Williams & Wilkins; 1929.
2. International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabé RR, et al. International network of cancer genome projects. Nature 2010;464:993–8.
3. Chin L, Hahn WC, Getz G, Meyerson M. Making sense of cancer genomic data. Genes Dev 2011; 25:534–55.
4. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409: 860–921.
5. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 2010;11: 685–9.
6. Schweiger MR, Kerick M, Timmermann B, Isau M. The power of NGS technologies to delineate the genome organization in cancer: from mutations to structural variations and epigenetic alterations. Cancer Metastasis Rev 2011;30:199–210.
7. National Human Genome Research Institute. DNA sequencing costs. http://www.genome.gov/sequencingcosts/ (Accessed July 2012).
8. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science 2009;323:133–8.
9. Pourmand N, Karhanek M, Persson HH, Webb CD, Lee TH, Zahradnikova A, Davis RW. Direct electrical detection of DNA synthesis. Proc Natl Acad Sci U S A 2006;103:6466–70.
10. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet 2010;19:R227–40.
11. Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. Nature 2010;464:999–1005.
12. Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch MJ, D'Ascenzo M, et al. Whole exome capture in solution with 3 Gbp of data. Genome Biol 2010;11:R62.
13. Jennings J, Hudson TJ. Reflections on the founding of the International Cancer Genome Consortium. Clin Chem 2013;59:XXX–XXX.
14. Collins FS, Barker AD. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. Sci Am 2007; 296:50–7.
15. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009;458:719–24.
16. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 2008;456:66–72.
17. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. Cell 2012;149:979–93.
18. Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW, et al. The life history of 21 breast cancers. Cell 2012;149:994–1007.
19. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell 2011;144:27–40.
20. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature 2012;486:400–4.
21. Ellis MJ, Ding L, Shen D, Luo J, Suman VJ, Wallis JW, et al. Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature 2012;486:353–60.
22. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 2012;486:405–9.
23. Smit VT, Boot AJ, Smits AM, Fleuren GJ, Cornelisse CJ, Bos JL. KRAS codon 12 mutations occur very frequently in pancreatic adenocarcinomas. Nucleic Acids Res 1988;16:7773–82.
24. Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L, et al. Novel mutations target distinct subgroups of medulloblastoma. Nature 2012;488:43–8.
25. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. Nature 2012;486:346–52.
26. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012;486:395–9.
27. Gray JW, Collins C, Henderson IC, Isola J, Kallioniemi A, Kallioniemi OP, et al. Molecular cytogenetics of human breast cancer. Cold Spring Harb Symp Quant Biol 1994;59:645–52.
28. Steilen-Gimbel H, Henn W, Kolles H, Moringlane JR, Feiden W, Steudel WI, Zang KD. Early proliferation enhancement by monosomy 10 and intratumor heterogeneity in malignant human gliomas as revealed by smear preparations from biopsies. Genes Chromosomes Cancer 1996;16:180–4.
29. Orndal C, Rydholm A, Willen H, Mitelman F, Mandahl N. Cytogenetic intratumor heterogeneity in soft tissue tumors. Cancer Genet Cytogenet 1994;78:127–37.
30. Gorunova L, Hoglund M, Andren-Sandberg A, Dawiskiba S, Jin Y, Mitelman F, Johansson B. Cytogenetic analysis of pancreatic carcinomas: intratumor heterogeneity and nonrandom pattern of chromosome aberrations. Genes Chromosomes Cancer 1998;23:81–99.
31. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366: 883–92.
32. Wu X, Northcott PA, Dubuc A, Dupuy AJ, Shih DJ, Witt H, et al. Clonal selection drives genetic divergence of metastatic medulloblastoma. Nature 2012;482:529–33.
33. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature 2010;467:1109–13.
34. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature 2012;481:506–10.

35. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012;30:413–21.
36. Navin N, Hicks J. Future medical applications of single-cell sequencing in cancer. Genome Med 2011;3:31.
37. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature 2011;472:90–4.
38. Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 2012;148:886–95.
39. Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X, et al. Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. Cell 2012;148:873–85.
40. Samuel N, Hudson TJ. The molecular and cellular heterogeneity of pancreatic ductal adenocarcinoma. Nat Rev Gastroenterol Hepatol 2011;9:77–87.
41. Su X, Zhang L, Zhang J, Meric-Bernstam F, Weinstein JN. PurityEst: estimating purity of human tumor samples using next-generation sequencing data. Bioinformatics 2012;28:2265–6.
42. Knoppers BM. Consent to 'personal' genomics and privacy. Direct-to-consumer genetic tests and population genome research challenge traditional notions of privacy and consent. EMBO Rep 2010;11:416–9.
43. Tran B, Dancey JE, Kamel-Reid S, McPherson JD, Bedard PL, Brown AM, et al. Cancer genomics: technology, discovery, and translation. J Clin Oncol 2012;30:647–60.
44. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011;364:2507–16.
45. Flaherty KT, Puzanov I, Kim KB, Ribas A, McArthur GA, Sosman JA, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. N Engl J Med 2010;363:809–19.
46. Yancovitz M, Litterman A, Yoon J, Ng E, Shapiro RL, Berman RS, et al. Intra- and inter-tumor heterogeneity of $BRAF^{V600E}$ mutations in primary and metastatic melanoma. PLoS One 2012;7:e29336.
47. Heyn H, Esteller M. DNA methylation profiling in the clinic: applications and challenges. Nat Rev Genet 2012;13:679–92.
48. Dancey JE, Bedard PL, Onetto N, Hudson TJ. The genetic basis for cancer treatment decisions. Cell 2012;148:409–20.