INNOVATION

Implementing personalized cancer genomics in clinical trials

Richard Simon and Sameek Roychowdhury

Abstract | The recent surge in high-throughput sequencing of cancer genomes has supported an expanding molecular classification of cancer. These studies have identified putative predictive biomarkers signifying aberrant oncogene pathway activation and may provide a rationale for matching patients with molecularly targeted therapies in clinical trials. Here, we discuss some of the challenges of adapting these data for rare cancers or molecular subsets of certain cancers, which will require aligning the availability of investigational agents, rapid turnaround of clinical grade sequencing, molecular eligibility and reconsidering clinical trial design and end points.

New technology for next-generation sequencing[1,2] (NGS) has enabled a systematic cataloguing of cancer genomes through national and international genomics projects. For example, the Cancer Genome Atlas and the International Cancer Genomics Consortium have identified recurrent point mutations, translocations and potentially new therapeutic targets in more than 20 and 50 cancer subtypes, respectively[3,4]. These projects leverage new technologies and make these data available to the wider cancer research community. Meanwhile, academic cancer centres and companies are seeking to translate this immense data set and technology for clinical applications. Fortunately, emerging technologies such as desktop sequencers and targeted gene capture have made these efforts both feasible and affordable for clinical cancer research. However, utilizing sequencing technology prospectively for the treatment of patients involves important new challenges. In this article, we discuss the challenges of implementing cancer sequencing in clinical oncology.

Molecular taxonomy for cancer medicine

In late 2011, the National Research Council (NRC) of the US National Academy of Sciences released a publication supporting the need to build and utilize a "new taxonomy of human disease" to facilitate precision medicine[5]. Potential applications for putative cancer biomarkers include predictive, prognostic and pharmacogenomic biomarkers (FIG. 1) that provide decision-making support for answering questions about which therapy to choose, who should be given treatment and what dose of drug to use. However, and in contrast to the extensive number of hits one finds for the terms "cancer" and "biomarker" in a search on the PubMed database, the true number of clinically applied predictive biomarkers is staggeringly small.

Three crucial steps for developing a clinical biomarker involve establishing its analytical validity, clinical validity and clinical utility for a well-defined indication, and this has only been performed for a small number of biomarkers in oncology. Analytical validation means establishing that the test measures what it claims to measure, and does so accurately with adequate sensitivity and specificity. Analytical validity refers not just to the hardware platform used for sequencing but to the entire process of sequencing a sample, including sample preparation, performing the sequencing assay and the computational pipeline for assembling the sequence read-outs and calling variants. The clinical validity of a predictive biomarker establishes that the biomarker correlates with a specific clinical response. Clinical utility means that measuring the biomarker and using it for decision-making is beneficial to patients relative to the standard of care. The end point for establishing clinical utility is generally survival or progression-free survival, whereas the end point used for establishing clinical validity is often tumour or clinical response. Analytical validation typically does not require samples from patients enrolled in clinical trials, but establishing clinical validation and clinical utility requires that patients are studied in clinical trials in which the relationships among the biomarker, treatment and outcome are determined.

Thus, NGS can be utilized to identify known genomic targets that have evidence of 'driver status' and indicate sensitivity to targeted therapies, as established in preclinical models or clinical studies. Driver mutations confer a growth advantage on cancer cells and are positively selected at some point in the development of that cancer. In TABLE 1, we highlight genomic aberrations, including point mutations, amplifications and rearrangements, that represent putative targets for novel molecularly targeted therapies. Clinical trials need to test the clinical validity and utility of each of these putative predictive genomic biomarkers to match patients with targeted therapies. A major challenge of such trials is that most putative genomic aberrations occur across a range of cancer subtypes, yet they typically only have a low frequency within a disease group defined by a specific tissue of origin. Clinical investigators will be tasked with accruing patients with a broad range of diseases, including rare cancers, into their trials according to a molecular classification strategy. This will require a cost-effective approach to screen or test patients using molecular diagnostics for predictive biomarkers. The challenges for the implementation of personalized cancer genomics in clinical trials are summarized in BOX 1.

Molecular diagnostics

To implement molecular diagnostics for clinical trials, several important points need to be considered. These include assay design, costs, tissue samples, analytical test validity, clinical laboratory implementation, the availability of results and data analysis.

Assay design. Current tumour sequencing strategies (FIG. 2) enable the evaluation of several genes simultaneously, and range from sequencing panels focused on ten genes to whole-genome sequencing. Establishing the analytical validity of a sequencing diagnostic, however, is considerably more difficult for whole-exome or whole-genome approaches, in which the average depth of coverage is much lower than for targeted gene sequencing of a defined panel of genes. The establishment and maintenance of the analytical validity of a diagnostic test requires substantial effort. In a clinical laboratory, this entails initial and periodic testing on reference standards to demonstrate the reproducibility of the assay.

Although some platforms utilize mass spectrometry-based assays, most clinical cancer sequencing efforts are converting to NGS-based approaches, given the decreasing costs and the potential for expansion to

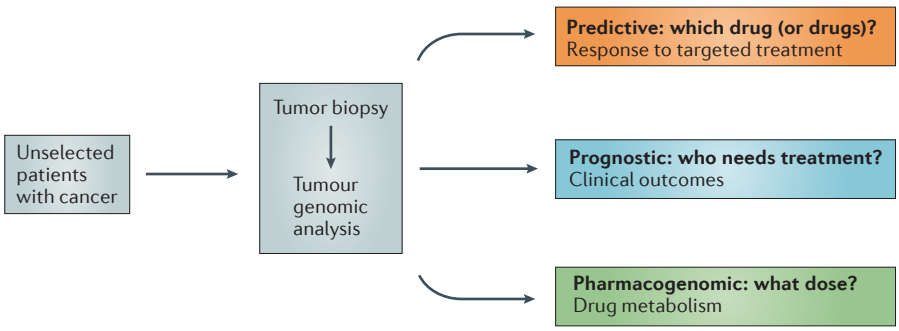

Figure 1 | **Development and application of biomarkers for oncology.** Genomic sequencing and other omics-based strategies have the potential to identify candidate biomarkers in clinical oncology. Clinical trial design is dictated by the type of biomarker being testing or developed. Predictive biomarkers inform the investigator of a potential clinical response to a given therapy. Prognostic biomarkers provide information on the risk of disease progression or relapse. Pharmacogenomic biomarkers relay data on how a patient may respond to a drug with respect to toxicity or efficacy.

cover more of the genome. Pre-NGS panels include Oncomap (33 genes) and SNAPshot (13 genes)[6,7] assays, and focus on selected oncogenes with known hotspots[6,7]. These assays are well suited for detecting a limited number of mutations in oncogenes, and the hotspots investigated are usually well annotated with regard to functional relevance. However, this approach does not comprehensively identify the large number of different mutations that can often be found in tumour suppressor genes, or mutation classes such as copy number alterations or rearrangements.

By contrast, NGS provides an unbiased testing strategy that does not solely focus on hotspots but also identifies variants whose functional status is not documented in as much detail. For example, hybridization-based capture is a technological innovation that focuses on NGS of a targeted list of genes or exons through the enrichment of DNA regions via complementary oligonucleotide baits. Targeted capture of exons for panels of 100–200 genes identifies relevant hotspots but can also enable the determination of copy number alterations such as deletions and amplifications[8–10]. Rearrangements or gene fusions involving kinases are also crucial. Known rearrangements can be detected through exon capture supplemented with the capture of introns involved in breakpoints[9,10].

TABLE 2 illustrates the potential advantages and disadvantages of the various NGS-based molecular diagnostic approaches. The main practical considerations include the number of genes sequenced, the turnaround time, ease of scalability, ensuring analytical validity, potential for research discoveries and the cost of testing per patient. A multiplier based on the number of genes tested and the amount or depth of sequencing (depth of coverage) affects the cost per patient.

Generally, sequencing is conducted with a planned average depth of coverage. The actual coverage depends, however, on the local genome structure. For example, the coverage often decreases in the region of repetitive sequences. Furthermore, the cost is affected by the complexity of larger numbers of genes and the amount of sequencing performed. For example, sequencing whole exomes, whole genomes and transcriptomes involves additional sample preparation, reagents and analysis; also, it may not be practical for a laboratory that is not supported by an existing sequencing and bioinformatics infrastructure.

The challenge of establishing the analytical validity of thousands of genes, each with numerous potential variants, is daunting. However, as there is a limited fraction of genes and mutations that can be actionable as biomarkers for clinical trials, analytical validation should focus on a selected set of genes and mutations, as determined by an institution's complement of clinical trials. The number of variants for a particular tumour suppressor may be numerous, whereas clinical actionability would be restricted to those mutations for which there is evidence of relevance from the literature, those resulting in a stop codon or a frame shift or other types of mutations that are specifically defined as one of the eligibility criteria of a particular clinical trial.

An alternative approach for establishing analytical validity would be to perform NGS on a platform that is validated for selected variants and to validate other variants found by NGS using an orthogonal platform such as Sanger sequencing[11]. Any additional NGS data produced could be considered exploratory or correlative research and may not be used prospectively for clinical decision-making, but instead used retrospectively for hypothesis-generating research. Thus, with regard to the range of assay designs available, the prospective clinical mission may be accomplished through the targeted sequencing of 100–500 genes, whereas a retrospective biomarker discovery mission could benefit from a broader scope of sequencing, including exome, transcriptome and whole-genome sequencing.

Tissue quality and tumour content. The quality of molecular tumour assessment is limited by several factors, including the tissue quality, the tumour content within a sample, the depth of sequencing and the effectiveness of the computational pipeline. Tissue quality relates to the age of the tissue, fixation time and the size of the specimen. Fresh frozen tissues have several important advantages over formalin-fixed paraffin-embedded (FFPE) diagnostic specimens for prospective clinical trials. This is because fresh frozen tissue is more likely to accurately reflect the current stage of the disease, particularly if there have been numerous prior cytotoxic therapies that have imparted selective pressure for the development of resistance. The DNA and RNA quality of fresh frozen tissue is generally also better than of FFPE material. Moreover, the examination of fresh frozen tissue ensures comparable tissue quality with post-progression samples. However, several groups have developed protocols to work effectively with FFPE material, thereby enabling important retrospective biomarker discovery and research.

The tumour content is a reflection of the underlying admixture of tumour cells, adjacent normal tissue and other stromal components, and thereby affects the sensitivity for detecting tumour variants. For example, for a tumour sample with 50% tumour content, the sensitivity of detecting heterozygous alterations is limited to 25% of the sequencing bandwidth applied to that sample. Laser capture microdissection of tissues to separate tumour tissue from adjacent normal tissue has the potential to overcome the challenges posed by low tumour content, but this technique is time-intensive, may lead to additional nucleic acid degradation during handling and yields only limited amounts of DNA. Such small quantities of tissue or nucleic acid can

Table 1 | **Genomic alterations as putative predictive biomarkers for cancer therapy**

Genes	Pathways	Aberration type	Disease examples	Putative or proven drugs
PIK3CA[51,52], *PIK3R1* (REF. 53), *PIK3R2*, *AKT1*, *AKT2* and *AKT3* (REFS 54,55)	Phosphoinositide 3-kinase (PI3K)	Mutation or amplification	Breast, colorectal and endometrial cancer	• PI3K inhibitors • AKT inhibitors
PTEN[56]	PI3K	Deletion	Numerous cancers	• PI3K inhibitors
MTOR[57], *TSC1*[58] and *TSC2* (REF. 59)	mTOR	Mutation	Tuberous sclerosis and Bladder cancer	• mTOR inhibitors
RAS family (*HRAS*, *NRAS*, *KRAS*), *BRAF*[60] and *MEK1*	RAS–MEK	Mutation, rearrangement or amplification	Numerous cancers, including melanoma and prostate cancer	• RAF inhibitors • MEK inhibitors • PI3K inhibitors
Fibroblast growth factor receptor 1 (*FGFR1*), *FGFR2*, *FGFR3*, *FGFR4* (REF. 36)	FGFR	Mutation, amplification or rearrangement	Myeloma, sarcoma and bladder, breast, ovarian, lung, endometrial and myeloid cancers	• FGFR inhibitors • FGFR antibodies
Epidermal growth factor receptor (*EGFR*)	EGFR	Mutation, deletion or amplification	Lung and gastrointestinal cancer	• EGFR inhibitors • EGFR antibodies
ERBB2 (REF. 61)	ERBB2	Amplification or mutation	Breast, bladder, gastric and lung cancer	• ERBB2 inhibitors • ERBB2 antibodies
SMO[62,63] and *PTCH1* (REF. 64)	Hedgehog	Mutation	Basal cell carcinoma	• Hedgehog inhibitor
MET[65]	MET	Amplification or mutation	Bladder, gastric and renal cancer	• MET inhibitors • MET antibodies
JAK1, *JAK2*, *JAK3* (REF. 66), *STAT1*, *STAT3*	JAK–STAT	Mutation or rearrangement	Leukaemia and lymphoma	• JAK–STAT inhibitors • STAT decoys
Discoidin domain-containing receptor 2 (*DDR2*)	RTK	Mutation	Lung cancer	• Some tyrosine kinase inhibitors
Erythropoietin receptor (*EPOR*)	JAK–STAT	Rearrangement	Leukaemia	• JAK–STAT inhibitors
Interleukin-7 receptor (*IL7R*)	JAK–STAT	Mutation	Leukaemia	• JAK–STAT inhibitors
Cyclin-dependent kinases (*CDKs*;[67] *CDK4*, *CDK6*, *CDK8*), *CDKN2A* and cyclin D1 (*CCND1*)	CDK	Amplification, mutation, deletion or rearrangement	Sarcoma, colorectal cancer, melanoma and lymphoma	• CDK inhibitors
ABL1	ABL	Rearrangement	Leukaemia	• ABL inhibitors
Retinoic acid receptor-α (*RARA*)	RARα	Rearrangement	Leukaemia	• All-trans retinoic acid
Aurora kinase A (*AURKA*)[68]	Aurora kinases	Amplification	Prostate cancer and breast cancer	• Aurora kinase inhibitors
Androgen receptor (*AR*)[69]	Androgen	Mutation, amplification or splice variant	Prostate cancer	• Androgen synthesis inhibitors • Androgen receptor inhibitors
FLT3[70]	FLT3	Mutation or deletion	Leukaemia	• FLT3 inhibitors
MET	MET–HGF	Mutation or amplification	Lung cancer and gastric cancer	• MET inhibitors
Myeloproliferative leukaemia (*MPL*)	THPO, JAK–STAT	Mutation	Myeloproliferative neoplasms	• JAK–STAT inhibitors
MDM2 (REF. 71)	MDM2	Amplification	Sarcoma and adrenal carcinoma	• MDM2 antagonist
KIT[72]	KIT	Mutation	GIST, mastocytosis, leukaemia	• KIT inhibitors
PDGFRA and *PDGFRB*	PDGFR	Deletion, rearrangement or amplification	Haematological cancer, GIST, sarcoma and brain cancer	• PDGFR inhibitors
Anaplastic lymphoma kinase (*ALK*)[9,37,73,74]	ALK	Rearrangement or mutation	Lung cancer and neuroblastoma	• ALK inhibitors
RET	RET	Rearrangement or mutation	Lung cancer and thyroid cancer	• RET inhibitors
ROS1 (REF. 75)	ROS1	Rearrangement	Lung cancer and cholangiocarcinoma	• ROS1 inhibitors
NOTCH1 and *NOTCH2*	Notch	Rearrangement or mutation	Leukaemia and breast cancer	• Notch signalling pathway inhibitors

CDKN2A, cyclin-dependent kinase inhibitor 2A; ERBB2, also known as HER2; GIST, gastrointestinal stromal tumour; FLT3, FMS-like tyrosine kinase 3; HGF, hepatocyte growth factor; JAK, Janus kinase; MEK, MAPK/ERK (mitogen-activated protein kinase/extracellular signal-regulated kinase) kinase; mTOR, mammalian target of rapamycin; PDGFR, platelet-derived growth factor receptor; PIK3R1, PI3K regulatory subunit 1; PIK3CA, PI3K catalytic subunit-α; PTCH1, Patched 1; PTEN, phosphatase and tensin homolog; RTK, receptor tyrosine kinase; SMO, Smoothened; STAT, signal transducer and activator of transcription; THPO, thrombopoietin; TSC1, tuberous sclerosis 1 protein.

negatively impact the sensitivity of NGS, as small amounts of DNA require additional rounds of enzymatic amplification, which results in the sequencing of a greater number of duplicate DNA fragments and may affect the sensitivity of detecting mutations that are present in only a small percentage of the DNA. The depth of sequencing is directly proportional to the sensitivity of detecting mutations in heterogeneous tissue samples and thus can be adjusted for samples with limited tumour content. However, the depth of sequencing also influences throughput and cost.

Currently, there are no formal standards for the minimum depth of sequencing. Such standards are likely to be developed, however, as laboratories and companies are required to determine the sensitivity and specificity of their assays through comparison with reference samples. Reference standards may be quality-controlled DNA samples from a secondary vendor or samples from other clinical sequencing laboratories with expected genomic variants, and they can be tested to determine a laboratory's proficiency for detecting the variant as well as defining the limits of detection[11]. The required depth of coverage will depend on several factors including the NGS platform, expected tissue types, tumour purity and the types of variants tested. As the depth of coverage may depend strongly on local sequence context, the accuracy of detecting certain variants can be limited.

Certification and turnaround time. Laboratories generating clinical tumour sequencing data with the intent to use these for therapeutic decision-making will need to consider good laboratory practices and turnaround time of results. The Clinical Laboratory Improvements Amendment (CLIA) was established in the United States to ensure a high quality of laboratory testing, including accuracy, reliability and timeliness of clinical test results. Regulations should also take into consideration the training and experience of personnel, equipment and reagents utilized, organization of laboratory operations, quality control measures and proficiency testing, maintenance and troubleshooting of equipment and procedures as well as the interpretation of results. Several professional societies have issued or are generating guidelines for the implementation of NGS, including the College of American Pathologists (CAP; see the 29 September 2011 press release on the CAP website), the US National Cancer Institute (NCI; see the Cancer Diagnosis Program on the NCI website), the US Food and Drug Administration (FDA), a genome centre working group[11], the American Society of Human Genetics (ASHG) and the American College of Medical Genetics and Genomics (ACMG).

Most NGS applications for clinical oncology are initially developed for research or investigational purposes only and then proceed to CLIA certification. Several clinical research centres and consulting biotechnology companies are running CLIA-certified laboratories for their NGS-based cancer diagnostic tests[9,12]. At present, there are no guidelines for the recommended depth of sequencing or for computational tools. A key outcome of a recent Institute of Medicine report on translational omics-based testing was the importance of a laboratory's consistent and transparent use of methods for computational analysis from the beginning to the end of a given clinical trial[13]. In other words, once the computational method has been selected for analysing data prospectively in a given clinical trial, it should remain 'locked down'. However, once the study was completed, the data could still be retrospectively analysed anew with novel methods as they become available.

Establishing analytical validity will be particularly challenging for exome or whole-genome sequencing. In contrast to focused cancer panels (comprising 200 genes), whole-exome (comprising 20,000 genes) and whole-genome sequencing is substantially more costly and will therefore have a limited depth of coverage, thus increasing the risk of false positives. The probability of a false-positive variant call at a given location depends on the depth of sequencing and prior probability of a variant at that location. Conversely, false negatives are due to poor sensitivity and can negatively affect enrolment in clinical trials recruiting patients with rare cancers or mutations.

The turnaround time for results is a pragmatic consideration for patients undergoing molecular profiling. For patients with advanced cancer who are being considered for treatment with investigational agents, a turnaround time of 4–8 weeks is too long and not practical for clinical decision-making. Routine testing for single genes currently takes 1–2 weeks and we expect that NGS methodologies will enable turnaround times of under 14 days. As technology platforms continue to improve, permitting shorter run times and increased sequencing throughput, we expect that the rate-limiting step will not be sequencing but sample logistics, data analysis and interpretation.

Box 1 | **Barriers for clinical translation**

Molecular diagnostics
- Choice of assay and design
- Cost
- Tissue quality
- Tumour content
- Analytical validity
- Clinical Laboratory Improvements Amendment (CLIA) certification
- Turnaround time
- Bioinformatics analysis

Clinical implementation
- Tissue acquisition
- Heterogeneity
- Expert interpretation
- Pathway versus tissue of origin
- Availability of broad panels of drugs
- Trial design and end points
- Clinical validity and utility

Bioinformatics analysis. The Institute of Medicine recently completed an evaluation of 'translational omics', examining high-throughput technologies generating genomic, proteomic and metabolomic data sets and their translation for clinical research[13]. This committee was charged with making recommendations for the application of omics-based tests in clinical trials. It identified agencies and peer review organizations (journals, institutions, funding agencies and the FDA) to establish standards and regulations for omics-based testing in clinical trials. This included the identification and approval, by the FDA, of both sequencing assays and their associated analysis software tools as potential investigational devices. Furthermore, the committee identified the public availability and transparency of raw data as a means to enable the external validation of omics-based trials.

For NGS-based diagnostics, the basic computational challenges include establishing validated computational pipelines that are sufficient for determining the analytical validity of variant calls. The computational pipeline needs to allow secure storage of large data files as well as access to high-performance computing to enable rapid analysis of data. The accuracy of variant detection is strongly influenced by the quality of the computational pipeline in terms of mapping the short readouts to a reference genome and in making variant calls based on that mapping. Currently, several methods are available for these purposes[14–17]. To date, however,

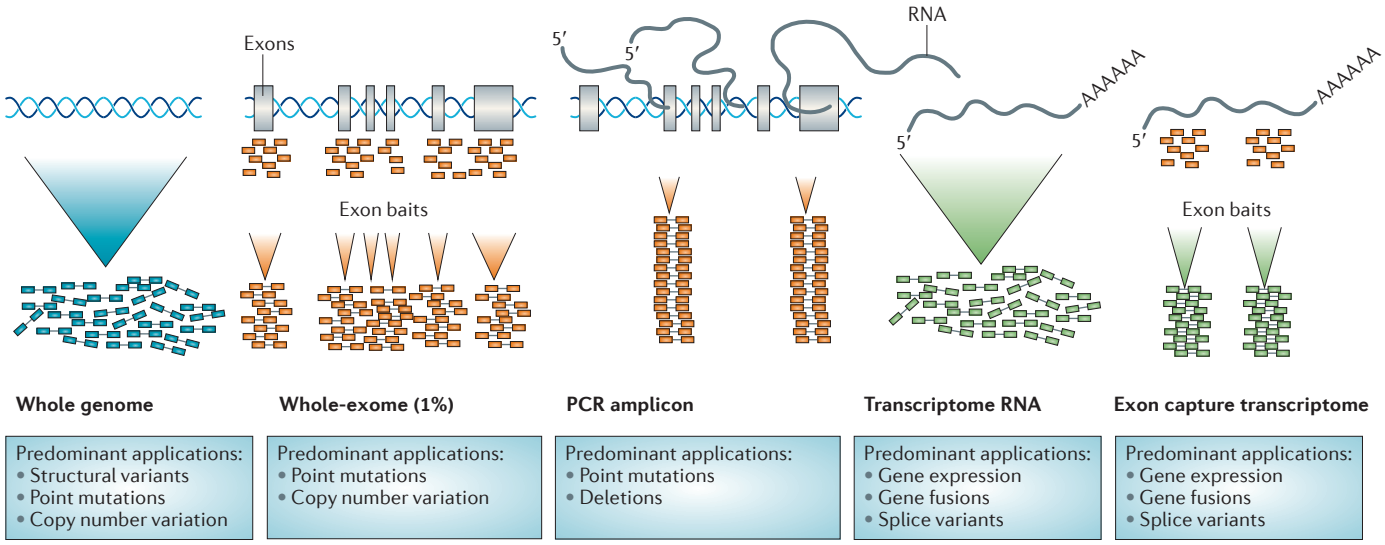

Figure 2 | **Strategies for next-generation sequencing in cancer.** This schematic demonstrates the potential strategies for the application of next-generation sequencing (NGS) in clinical oncology testing and research. Whole-genome sequencing evaluates the entire genome and includes both gene-coding and non-coding regions. Exome sequencing uses baits to hybridize and capture corresponding regions of the genome, focusing on the coding regions of the genome. Exome sequencing can include the whole exome (about 20,000 genes), comprising just over 1% of the genome; alternatively, it can focus on a panel of genes (hundreds of genes or more). Amplicon-based sequencing utilizes PCR amplification to isolate a smaller region for sequencing. Transcriptome sequencing[2,42] or RNAseq evaluates the expressed RNA and can be used to measure gene expression, splice variants and nominate candidate gene fusions. Similar to exome sequencing, complementary baits can be used to hybridize and capture portions of the transcriptome to focus on selected genes of interest.

there has not been a rigorous comparative study of the right computational tool for clinical sequencing. An optimal model will consider estimates of tumour purity, ploidy, tumour subclones, local genomic context and sequencing depth[18]. Finally, real-time medical informatics systems that integrate the sites and personnel involved with tumour sequencing, variant interpretation and patient treatment must be accomplished through the use of electronic medical records.

To facilitate the interpretation of individual patient data for potential clinical trials or expert interpretation through tumour boards, detected variants must be annotated with regard to their clinical 'actionability'. Some useful databases have emerged — such as My Cancer Genome — that facilitate the assessment of the actionability of known genomic alterations. Bioinformatics systems for managing individualized cancer genomics data in the context of various alterations and drugs are now being developed by utilizing the expertise of clinical professionals in their respective fields (for example, such systems are being developed by CollabrX and GeneInsight).

For mutational hotspots, the functional relevance of the alteration is generally apparent but there may be insufficient clinical or preclinical data to indicate whether or not the alteration is actionable for treatment with a specific investigational drug. For example, an activating V600E mutation in *BRAF* for a patient with colorectal cancer may have appeared to be actionable for treatment with a BRAF inhibitor before it was found that such tumours tend to be unresponsive because of the upregulation of epidermal growth factor receptor (EGFR) signalling pathways[19]. An alternative for these patients would be one or a few clinical trials offering dual treatment with BRAF and EGFR inhibitors. Thus, the annotation of variants must include up-to-date data on functional impact, previous clinical knowledge and current trials.

For non-hotspot alterations in genes that are considered to be biologically relevant, however, it may be difficult to know whether the alteration is involved in deregulating a particular pathway and whether it is clinically important. Such assessments can be supported by algorithms that predict the effects of genomic alterations on protein function, as well as the development of improved systems biology and experimental models for relating genomic alterations to drug effectiveness[20]. For example, assessing whether a newly discovered alteration may be functionally relevant rests heavily on how many times it has been reported in an international database of mutations associated with cancer (such as the Catalogue of Somatic Mutations in Cancer; COSMIC)[21], whether the alteration is in the phosphorylation loop of an oncogenic kinase or whether it alters the reading frame of a tumour suppressor gene. Nonetheless, novel variants of unproven biological significance cannot be utilized for clinical decision-making regarding therapies.

An important limitation in the synthesis of cancer genomics data for clinical trials is that expertise is required across diverse areas in statistical genomics, clinically applied bioinformatics, cancer biology, experimental therapeutics and clinical oncology, and assembling all of this expertise into one location is challenging because it is distributed across several departments or institutes. Although research laboratories and academic centres develop and apply their preferred tools and approaches, we anticipate that the growth of genomics-driven clinical trials will create a demand or force for the standardization of NGS assays and methodologies across centres that are sufficiently reproducible and can thus produce validated data for clinical use.

Table 2 | **Spectrum of molecular diagnostics for clinical oncology trials***

Methods or assays (genes tested)	Technology	Platforms	Academic centres or companies	Speed; discovery opportunity	Burden‡; CLIA certification challenges	§Cost per patient
Pre-NGS						
• Sequenom[15] • SnapSHOT[26] (20–50)	• PCR • Mass spectrometry	• MassArray	• Karmanos Cancer Institute • Dana-Farber Cancer Institute (DFCI)/Broad Institute of MIT and Harvard	12 weeks; limited scope	Low; medium	$500–1,500
NGS						
• Gene panel by PCR amplicon[5,12] (50–100)	• NGS	• Personal Genome Machine • MiSeq	• Oregon Health Sciences University • Baylor College of Medicine • Washington University in St Louis • Fox Chase Cancer Center	2 weeks; medium	Medium; medium	$500–1,500
• Gene panel by targeted capture[49,60] (200–10,000)	• NGS	• HiSeq2000 • HiSeq2500 • Proton	• Foundation Medicine • Ohio State University • University of Washington • University of Michigan	2–8 weeks; high	High; high	$500–1,500
• Exome and/or transcriptome (RNAseq) • Transcriptome (RNAseq)[8,35] (20,000)	• NGS	• HiSeq2000 • HiSeq2500 • Proton	• Baylor College of Medicine • DFCI/Broad Institute of MIT and Harvard • Ohio State University • University of Michigan • Washington University in St Louis	4–12 weeks; very high	Very high; very high	$5,000–10,000
• Whole-genome and transcriptome (RNAseq) • Transcriptome (RNAseq)[10,76,77] (20,000+)	• NGS	• HiSeq2000 • HiSeq2500 • Proton	• Translational Genomics Research Institute • Washington University in St Louis	4–12 weeks; very high	Very high; very high	$5,000–20,000

*Ongoing efforts for integrating molecular diagnostics for clinical trials are catalogued above, including pre-next-generation sequencing (NGS) and new NGS strategies. Sequencing strategies for clinical oncology can be divided into targeted approaches focusing on 50–500 clinically or biologically significant genes or broad approaches evaluating 20,000 genes or more (or the entire genome). These strategies have advantages and disadvantages with respect to the equipment needed, the speed of the results (turnaround time), opportunities for research discovery, complexity of data analysis, the challenges of operating in a Clinical Laboratory Improvements Amendment (CLIA)-certified laboratory and costs. This table reflects the NGS technologies that are currently in use for CLIA-certified testing in cancer. ‡Refers to computational burden. §The costs per patient (reagents only; in US$) reflect early 2013 estimates but may vary by economies of scale. MIT, Massachusetts Institute of Technology.

Clinical implementation

There are several points to consider for developing prospective clinical trials that incorporate NGS strategies. These include issues associated with tissue collection, tumour heterogeneity, the clinical interpretation for a given mutation, patient eligibility in pathway-oriented versus tissue-of-origin-oriented clinical trials as well as the selection of clinical end points.

Tissue acquisition. The age and method of preservation of clinical samples are important practical considerations in the clinical implementation of NGS, as discussed above. Although procedures for the evaluation of FFPE samples have enabled effective genomic sequencing, assay reliability is still affected by differential tissue quality and the age of the specimen, for both DNA and RNA sequencing[22,23]. In addition, although there may be concordance in mutations from primary cancer sites and metastatic disease sites, older archival samples may not reflect the current stage of the disease. Thus, although the sequencing of FFPE tissue has been effective in many cases for retrospective studies, in prospective clinical trials it is desirable to evaluate the current stage of the disease, particularly in patients who have undergone extensive pretreatment.

Research biopsy samples are increasingly being incorporated and accepted as components of clinical trials as they add correlative research value to study drug metabolism, drug effect and cancer biology[24]. Clinically, this added value supports the goal of developing predictive biomarkers for clinical decision-making and an enhanced understanding of cancer biology related to tumour resistance, evolution and heterogeneity. Thus, for genomics-driven clinical trials, the use of research biopsy samples may guide

the enrichment or eligibility for clinical trials of targeted therapies based on the presence of putative driving mutations (FIG. 3). Furthermore, a repeat research biopsy, particularly at the time of disease progression, will contribute to our understanding of the mechanisms of secondary or acquired resistance (FIG. 3a) and may also inform clinical decision-making.

Last, NGS strategies require the collection of normal or germline DNA through a blood, buccal, saliva or skin punch sample. Germline DNA shows normal variation in the form of thousands of single nucleotide polymorphisms that are not causative of disease and therefore need to be distinguished from acquired mutations in the DNA of tumour cells from the same individual. Similarly, by comparing the depth of sequencing of tumour versus germline DNA, one can determine how many copies of each gene are present in the tumour; that is, two copies (normal), extra copies (amplification) or the loss of one or both copies (deletion). Meanwhile, the assessment of germline DNA may also reveal known heritable cancer syndromes that are associated with up to 5–10% of cancers, and this information can be valuable to patients and their families. Blood samples are minimally invasive, easy to store and are thus efficient for patients with solid tumours. For patients with leukaemia, tumour DNA may be assessed in bone marrow or peripheral blood samples, and a second source of germline DNA is necessary. Although saliva is easy to collect from these patients, it may not be optimal as saliva contains mostly lymphocytes (thus contamination from leukaemia is possible) as well as some bacteria. Buccal swabs contain few lymphocytes but can also be contaminated by oral bacteria. Therefore, skin punch samples should be obtained for patients with leukaemia.

Heterogeneity. Tumour heterogeneity is a challenge in the application of personalized genomics. Heterogeneity may refer to subclones within a given population of tumour tissue or between spatially separated tumour sites. Exome sequencing of tissue from several metastatic sites in patients with renal cancers demonstrated site-to-site heterogeneity with diverging branch mutations but also common trunk mutations[25]. This and other studies support the notion of a heterogeneous population of cancer cells that exist in an 'ecosystem' with different selective pressures. Together, these studies support a model of a diverse tumour environment, in which selective pressures such as targeted therapy apply 'bottlenecks' and may fail because of emerging resistance in a tumour clone[26]. However, the 'founder mutations' that occur earliest in the carcinogenesis of the tumour subsequently exist in all subclones and metastatic sites, and so regimens that effectively target genomic alterations with high-variant frequencies may provide substantial tumour responses.

The oncogene addiction hypothesis states that several of the genomic alterations in later stages of disease develop in the context of the early founder mutations and are only viable in that context. This hypothesis also suggests that treatment regimens that are highly potent at targeting founder mutations may produce greater antitumour effects than might be expected based on tumour heterogeneity. However, there are other theoretical considerations that do not lead to such optimistic predictions. For example, it is possible that although all tumour subclones contain the early genomic alterations that are responsible for carcinogenesis, by the time they are treated these subclones are no longer dependent on the early alterations. In some cases, such as shown for the ABL fusion kinase in chronic myelogenous leukaemia, resistance can be a result of acquired mutations in the drug-binding site[27]. One thing that has become apparent is that owing to the diversity of cancer, a multi-agent approach will be required to target the multiple driving pathways[19,28–31].

Thus, a major application for NGS in clinical trials is the study of drug resistance. One strategy to address the role of site-to-site tissue heterogeneity in the response and resistance to therapy involves carrying out prospective biopsy procedures or retrospective autopsy series. An important limitation of multi-site biopsy samples of tumour metastases is the risk posed by invasive procedures. One strategy would be to evaluate selected patients simultaneously with one visceral biopsy site (for example, liver or lung) and biopsy samples from superficial sites (for example, skin or palpable lymph nodes). Such studies will help to determine which cancer types are prone to heterogeneity. To assess intratumoural heterogeneity, one approach has been ultradeep sequencing of tumours to detect low-frequency mutations in admixed subclones and normal tissue[32,33]. As the routine sampling of research biopsy samples is incorporated into clinical trial protocols, we anticipate that serial tumour sampling during disease progression will shed light on the role of tumour heterogeneity in drug resistance through the detection of emerging tumour subclones that appear at a low frequency before a given therapy but are subsequently involved in the development of resistance.

Expert interpretation. To use personalized cancer genomics data for patient selection in clinical trials, the importance of the mutations identified needs to be assessed. Putative driver mutations and passenger mutations need to be distinguished, and targeted therapies need to be identified that inhibit the deregulation of pathways by the driver mutations. Driver mutations provide a selective advantage for cancer, whereas the role of passenger mutations may not yet be clear. Putative genomic targets may have existing clinical relevance for a targeted therapy (for example, the *EGFR* mutation and the EGFR inhibitor erlotinib (Tarceva; Roche/Genentech)) or a potential treatment hypothesis that is being addressed through a clinical trial (for example, a phosphoinositide 3-kinase (PI3K) catalytic subunit-α (*PIK3CA*) mutation and a PI3K inhibitor).

Several databases have emerged to tackle the volumes of data for genome projects and drug screens[34]. Cancer genomics-oriented databases such as COSMIC catalogue mutations but do not provide an interpretation of specific mutations for clinical application. The Therapeutic Targets Database, DrugBank and the Pharmacogenomics KnowledgeBase (PharmGKB) are starting points for cataloguing gene mutations and matching drugs but owing to shortfalls with regard to the level of evidence required for potential drugs and in the updating of current investigational agents, these databases remain one step away from having utility for clinical oncologists who are practicing medicine or running clinical trials.

As patients and clinicians pursue clinical cancer sequencing, one must also consider the format in which this vast amount of data is reported. For example, reports from Foundation Medicine's FoundationOne test[9], the University of Michigan's MI-ONCOSEQ programme[35] and Oregon Health Science University's Cancer Panel[29] follow a strategy that summarizes the most important abnormal results or mutations. Owing to the limited real-time information available about individual patients, Foundation Medicine and MI-ONCOSEQ provide a short clinical summary of putative trials that the patient might be eligible for, but they do not report important details on whether the patient meets the eligibility criteria, whether the trial is actually open, whether the trial is

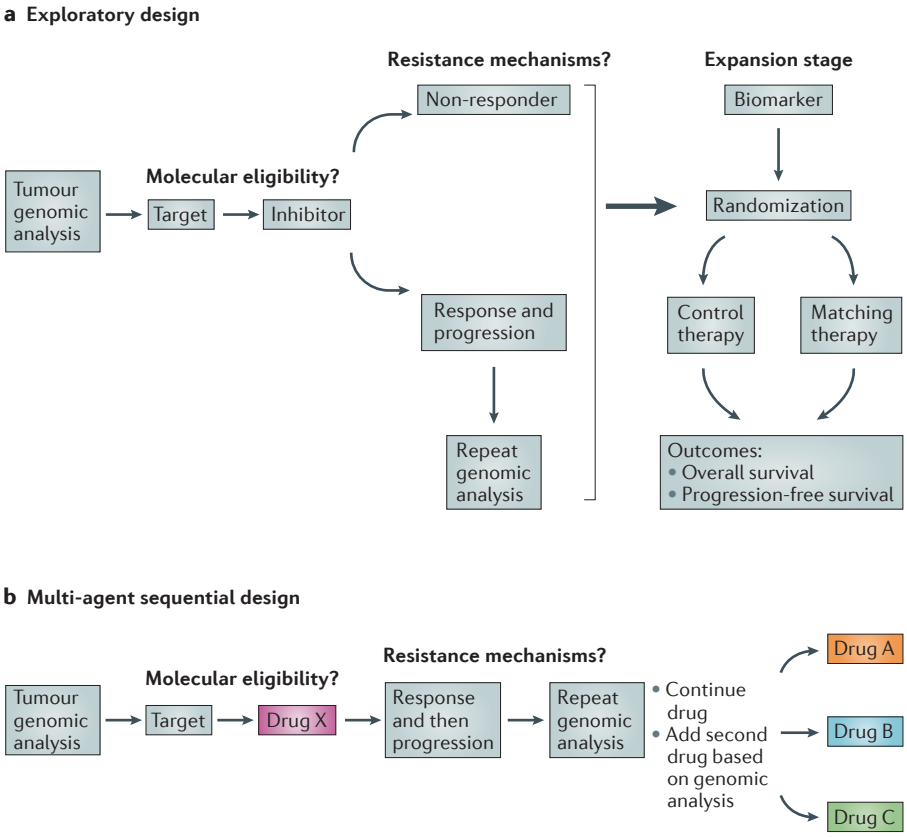

Figure 3 | **Prospective clinical trial designs incorporating genomic sequencing.** The following prospective clinical trial designs are structured to allow the study and treatment of patients with metastatic cancer and rare molecular alterations that cannot be studied in one disease owing to the rarity of the tumour or genetic alteration. These clinical trial designs accept patients with any histology of cancer but with a common mutation and require the sampling of fresh tumour biopsy tissue for research. These designs utilize next-generation sequencing to characterize the mechanisms of drug resistance. **a** | The first design is a molecular 'two-stage' exploratory design for patients who have currently untreatable tumours (regardless of the tissue of origin) and have defined actionable mutations that match the investigational drug. As the initial study is not randomized, the end points are objective tumour response, magnitude of response, durability of response and an understanding of the mechanism of drug resistance. The analysis will aggregate over tissue types but will be sensitive to the possibility that tissue type and mutations may have major effects on the response to therapy. The second stage — the expansion stage — utilizes the putative biomarkers from the exploratory stage to initiate a trial across several centres to enrol and randomize patients against standard therapies using traditional end points. **b** | The second design incorporates the sequential addition of a second therapy as a combination therapy. The second agent is determined based on a molecular analysis of patients with tumour progression after single-agent therapy. The combination is predicted to have additive or synergistic effects. This requires prerequisite knowledge about candidate mechanisms of acquired drug resistance to the initial therapy.

geographically feasible (usually defined as under two hours of travel between home and the clinical centre) and whether the trial's eligibility criteria will accept the patient. Thus, clinicians are often left with the task of additional interpretation, decision-making and navigating through potential trials. Several academic groups and companies are building decision tools that can aid the interpretation of these results. My Cancer Genome, based at Vanderbilt University, Tennessee, USA, is an emerging online resource that partially catalogues several types of cancer and may be a valuable resource for clinical interpretation.

At present, clinical oncologists must simply keep up with an ever-expanding list of putative genomic targets and novel investigational agents in clinical trials. To remain up to date with cancer genomics literature, several groups — for example, at the Translational Genomics Research Institute, University of Michigan, USA, and the Dana-Farber Cancer Institute, Boston, Massachusetts, USA — have implemented expert panels such as the 'Sequencing Tumour Board' or 'Molecular Rounds'. These tumour boards typically include 12–15 faculty members to share expertise in cancer genomics, bioinformatics, pathology, clinical genetics, bioethics and clinical oncology as well as experimental therapeutics. This model ensures that a group of specialists can interpret cancer genomics data for patients, and so the decisions do not rest with individual doctors. However, although a tumour board can facilitate the interpretation of tumour characterization data in the context of an individual patient, it may not serve as a final decision tool to dictate therapies but instead facilitate the triaging of patients into potential clinical trials with matching molecular eligibility. For example, a clinical trial of a novel fibroblast growth factor receptor (FGFR) inhibitor would include molecular eligibility for patients with predefined *FGFR* mutations, amplifications or translocations[36].

Pathway versus tissue of origin. Apart from the identification of putative genomic targets, disease-specific eligibility limitations are the main challenge for matching patients with therapies on the basis of a clinical cancer sequencing strategy. Clinical oncologists, pathologists, cancer centres, regulatory agencies (such as the FDA, institutional review boards and clinical trials offices) and the pharmaceutical industry currently organize and specialize the development of experimental therapies through a tissue-of-origin (histology) classification: for example, by appointing specialists with expertise in breast cancer, colon cancer or melanoma. Thus, most Phase II or III trials are conducted in patients who are selected according to histological criteria, and access to matching molecularly targeted therapies is limited for individual patients unless they have both a matching mutation and matching histological criteria. For example, a patient with colorectal cancer and a novel anaplastic lymphoma kinase (*ALK*) rearrangement would not be eligible for an ongoing ALK inhibitor study developed for treating lung cancer[37]. Conversely, it is not cost-effective for centres or the pharmaceutical industry to initiate trials for each investigational agent separately for each type of cancer because these trials are unlikely to accrue enough patients.

Another example is the rare cancer gastrointestinal stromal cell tumour (GIST) syndrome, in which 80% of patients have activating mutations in KIT, but a rare subset (1%) of patients have activating *BRAF* mutations. Physicians may want to offer an existing therapy that targets *BRAF*, such as vemurafenib (Zelboraf; Plexxikon/Roche), which is FDA-approved and could be administered 'off-label'. However, because the drug has no proven efficacy in GIST syndrome, it will probably not be covered by the patient's insurance; also, the data on response (or lack of response) are usually not as effectively captured as they would be in a clinical trial. A third option is to pursue Phase I trials that typically allow the participation of patients with different cancer subtypes. However, as the goal of Phase I studies is to determine the proper dose and assess dose-limiting toxicity, it can be unclear whether a given patient who does not respond is a true non-responder or has simply not received the right dosage.

How can we study rare mutations that may appear at a low frequency across many types of cancer, when the expected number of cases is small and will not allow a traditional randomized controlled trial to be carried out? One approach is to establish multicentre networks in which molecular characterization may facilitate the eligibility of patients for one of several existing clinical trials, each of which is developing a drug targeting a specific pathway. For example, the Lung Cancer Mutation Consortium presents a model for coordinating more than eight mutation-enriched trials — which could not otherwise be implemented separately — across 16 leading cancer centres. However, the Lung Cancer Mutation Consortium focuses on a common cancer with several molecular subsets — lung cancer — where it is possible to effectively screen and enrol enough patients even if the prevalence of a particular mutation is low. As discussed above, clinical trials of agents targeting a molecular aberration that has a low prevalence in a less common type of cancer may not be feasible in the context of cancer type-specific studies. Therefore, although there are increasing numbers of clinical trials that require molecular or mutation-based eligibility, these mainly focus on the more common cancer subtypes.

There are several national and international efforts in development for the large-scale profiling of patients with advanced cancer. In France, the SHIVA trial centralizes the screening of patient tumours for genomic alterations that putatively match existing approved therapies and compares the outcome of patients randomized to receive therapy based on profiling versus conventional therapy[38]. The recently founded multidisciplinary collaborative Melanoma Dream Team supports a study that is evaluating patients who have metastatic melanoma with wild-type *BRAF* (this encompasses about 50% of patients), and aims to match patients with mutations other than *BRAF* to a set of available drugs[39]. The NCI's Division of Cancer Treatment and Diagnosis in the United States is developing a pilot study for a nationwide system that connects tumour sequencing centres and clinical sites for the development of investigational agents for patients with a broad range of solid tumours.

> “Apart from the identification of putative genomic targets, disease-specific eligibility limitations are the main challenge for matching patients with therapies on the basis of a clinical cancer sequencing strategy.”

Despite these profiling efforts, at present there is no comprehensive cadre of clinical trials with mutation-based eligibility for any cancer subtypes; consequently, many patients and oncologists consider off-label therapy when there is a drug or drug combination for which there is a biological rationale based on data from another cancer type[40–43]. Although case reports are becoming ubiquitous, they do not usually report on patients who receive a particular treatment and do not respond, leading to suboptimal data collection. We therefore advocate the establishment of a registry or study that catalogues patients who are treated off-label and collects their data in a centralized manner.

Trial design and end points. Traditional biomarker-defined Phase III trials have sought to validate the clinical utility of common biomarkers in specific disease settings. An example of this is a clinical trial in breast cancer, in which patients with *HER2* (also known as *ERBB2* or neu) amplifications were randomized to receive either HER2-directed therapy plus chemotherapy or chemotherapy alone, and end points of progression-free or overall survival were measured[44–46]. The trend in oncological drug development has been to use this set-up even in Phase II studies. However, as discussed above, it is difficult to apply this model to patients with rare diseases or rare molecular subsets of certain types of cancer, for whom enrolment would be very low. Without a randomized control group and prospective protocol-defined patient follow-up and data collection, progression-free survival is very difficult to interpret with respect to evaluating the effect of a treatment because the rate of progression of different types of cancer varies substantially. Without a prospective clinical trial design that selects a treatment based on tumour characterization, and without prospectively planned and standardized criteria for evaluation and data collection, progress in understanding the relationships between a genomic alteration and the response to specific drugs will be slow.

Patients who are selected according to their molecular variants provide a sound basis for clinical trial design, but the design and analysis of clinical trials cannot be based on the assumption that all patients with similar mutations will respond similarly to a drug regardless of the primary site of the original tumour or additional molecular variants in their tumours. The pooling of such rare variants across tissue types may accrue enough patients to permit a randomized design and thus enable the use of progression-free survival as an end point for evaluating drug efficacy in the experimental versus control arm. If, however, the progression-free survival times vary widely based on the origin of the tumour tissue, then this may be of limited value.

Without a randomized control group, the only reliable clinical end points for assessing antitumour effects are the objective tumour response, the magnitude of responses and the duration of responses. These end points may be acceptable for molecularly enriched trials as one expects that overall responses (~30–40%) are higher than the 10–15% response rates that are typically observed in Phase I trials[47,48]. The early development of imatinib (Gleevec; Novartis) for GIST is a prime example of a non-randomized Phase II study that was conducted to establish that mutation-directed treatment in a new tissue type may achieve objective tumour responses[49]. In this study, investigators administered imatinib at two different doses to patients with refractory GIST and measured overall response rates of 33% and 43%. Thus, the aim of such Phase II studies can be to enhance our insight into the

Glossary

Amplicon-based sequencing
The use of PCR to selectively amplify small genomic regions for sequencing.

Clinical Laboratory Improvements Amendment
(CLIA). A US regulatory standard that applies to all clinical laboratory testing.

Depth of coverage
Also known as sequencing depth; the number of times a genome position has been sequenced to ensure data accuracy.

Driver mutations
Mutations that are implicated in cancer biology and provide a growth advantage at some point during the development of cancer, causing positive selection for the mutation.

Next-generation sequencing
(NGS). Also referred to as high-throughput sequencing and massively parallel sequencing. Refers to technologies that parallelize DNA sequencing effectively to produce millions of sequences in a rapid and cost-effective manner.

Orthogonal platform
A second DNA sequencing technology used to confirm data obtained through next-generation sequencing.

Passenger mutations
Mutations that do not contribute to cancer biology and do not appear to provide a growth advantage, but are carried along with driver mutations.

Transcriptome sequencing
Also referred to as RNAseq or whole-transcriptome shotgun sequencing. The sequencing of cDNA generated from total RNA. Transcriptome sequencing can provide data on gene expression, alternatively spliced transcripts, non-coding RNA and gene fusions or rearrangements.

Whole-exome sequencing
Also referred to as targeted exome capture. The selective application of next-generation sequencing to the coding regions of the genome using complementary oligonucleotide probes that selectively hybridize and capture the desired genomic regions of interest. Whole-exome sequencing represents approximately 20,000 genes or a little more than 1% of the whole genome, and is therefore a cheaper strategy than whole-genome sequencing. Targeted gene sequencing can be completed for a shorter defined list of genes: for example, for 200 to 1,000 or more cancer-related genes.

Whole-genome sequencing
Complete sequencing of an organism's entire DNA sequence, including exons and non-coding genome regions.

relationships among mutations, drugs, tissue types and response, whereas randomization is not essential as long as an antitumour effect can be clearly identified.

An additional goal of such exploratory Phase II trials is to dissect the mechanisms of primary and secondary resistance. In the case of patients with advanced metastatic cancer who receive molecularly targeted therapies, few will experience curative treatment. Potentially curative treatment is likely to require rational combination therapy that must be developed based on an understanding of acquired drug resistance.

Two innovative study designs that use genomic enrichment strategies are currently being utilized to develop targeted therapies (FIG. 3). In both designs, patients are selected according to molecular eligibility and therefore these studies include patients with cancers of any tissue of origin. In the exploratory design (FIG. 3a), tumour biopsy samples are collected before treatment and at the time of disease progression. These biopsy samples enable genomic analyses and provide an opportunity to identify mechanisms of drug resistance and, in turn, to inform future trial design with candidate biomarkers that are to be explored in a later expansion stage. The expansion stage involves a larger number of patients, with randomization across many centres, and measures traditional end points including progression-free and overall survival.

The second trial design (FIG. 3b), known as the sequential study design, is an extension of the exploratory design. It is in use by some pharmaceutical companies with several agents in their pipelines that are ready for combination therapies, as it allows the addition of a second targeted drug based on candidate resistance mechanisms identified after profiling patients who respond to the first drug and then progress. This design requires some existing insight into the predicted mechanisms of acquired resistance so that second-line targeted therapies may be rationally included in the trial as options for incremental therapy. Thus, this design strategy may arise based on the earlier exploratory design that identified initial therapeutic promise and mechanisms of acquired resistance.

Both study designs require research biopsy samples to be obtained before treatment and at time of progression to enable sequencing and analysis. Within each actionable mutation, the grouping of patients who receive the same targeted drug allows the comparison between primary non-responders and responders (and thereby determines primary resistance); at a later stage, it allows the comparison of patients pretreatment and post-progression (secondary resistance). Candidate subsets of patients with proven molecular biomarker targets can be promoted to an expansion stage in a separate clinical protocol involving focused candidate biomarker enrichment, with the goal of achieving clinical validation via randomization to therapy versus control. This can be expanded to several clinical centres to achieve accrual goals. Alternatively, the expansion stage can entail a combination strategy with a drug that targets an emerging resistance mechanism. Difficulties here include the challenges posed by accrual for trials involving rare genomic alterations, combining investigational drugs from different companies and the costs associated with clinical trials and drug development.

Ultimately, the goal of these clinical trial designs is to identify genomic biomarkers that predict a clinical response (clinical validity) for a matching targeted therapy and to improve overall outcomes (clinical utility). Furthermore, the incorporation of NGS in clinical trials has the potential to accelerate the development of companion diagnostics for approved drugs[50]. Companion diagnostics provide data on the potential safety and efficacy of a corresponding therapeutic agent. As these studies focus on the development and regulatory approval of both biomarkers and drugs, early discussions with regulatory agencies are crucial for planning such clinical trials.

Future directions

NGS and other omics-based strategies have brought oncology to exciting crossroads where there is tremendous opportunity for biomarker-defined trials. This will require the co-evolution of innovative genomics-based trial designs and sequencing technologies, and will provide the data for linking tumour genomics to therapeutic effectiveness. Meanwhile, important caveats and limitations to this approach remain, including the challenges posed by disease heterogeneity[25] and the complexity and influence of the epigenome. As genome projects identify the role of epigenetic changes in determining patient response to therapy, the identification of these epigenetic changes will be incorporated into clinical NGS strategies. Furthermore, technologies and methodologies that allow the clinical-grade evaluation of single cells will facilitate the study of cancer stem cells, tumour subclones and components of the adjacent tumour microenvironment. Thus, current NGS strategies

are a starting point that will evolve over time to incorporate additional technologies and address these caveats. In the near future, we anticipate a transition towards whole-genome sequencing, increased depth of sequencing and methylation-specific sequencing. Developing these technologies into analytically validated diagnostics that can be used for patient care will be a major challenge. The availability of drugs that are effective against the molecular deregulation caused by the identified mutations also present a substantial rate-limiting step, and many of the genes that are most frequently mutated in cancer may not be druggable at present.

New funding and philanthropic mechanisms may be needed to bridge these roadblocks to progress[41]. Although the challenges ahead are substantial and overcoming them will require scientific innovation as well as national focus and leadership, the opportunities to advance cancer therapeutics have never been greater.

Richard Simon is at the Biometric Research Branch, US National Cancer Institute, Bethesda, Maryland 20892-7434, USA.

Sameek Roychowdhury is at the Department of Internal Medicine, Division of Medical Oncology, Ohio State University, Biomedical Research Tower Room 508, 460 West 12th Avenue, Columbus, Ohio 43210, USA.

Sameek Roychowdhury is also at the Comprehensive Cancer Center and at the Department of Pharmacology, Ohio State University, Columbus, Ohio 43210, USA.

Correspondence to S.R.
e-mail: Sameek.roychowdhury@osumc.edu

doi:10.1038/nrd3979

1. Demeure, M. J. *et al.* Cancer of the ampulla of Vater: analysis of the whole genome sequence exposes a potential therapeutic vulnerability. *Genome Med.* **4**, 56 (2012).
2. Mardis, E. R. A decade's perspective on DNA sequencing technology. *Nature* **470**, 198–203 (2011).
3. Hudson, T. J. *et al.* International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
4. Ledford, H. Big science: the cancer genome challenge. *Nature* **464**, 972–974 (2010).
5. Committee on a Framework for Development a New Taxonomy of Disease & the National Research Council. *Toward Precision Medicine: Building a Knowlege Network for Biomedical Research and a New Taxonomy of Disease* (National Academies Press, 2011).
6. MacConaill, L. E. *et al.* Profiling critical cancer gene mutations in clinical tumor samples. *PLoS ONE* **4**, e7887 (2009).
7. Dias-Santagata, D. *et al.* Rapid targeted mutational analysis of human tumours: a clinical platform to guide personalized cancer medicine. *EMBO Mol. Med.* **2**, 146–158 (2010).
8. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotech.* **27**, 182–189 (2009).
9. Lipson, D. *et al.* Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nature Med.* **18**, 382–384 (2012).
10. Wagle, N. *et al.* High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov.* **2**, 82–93 (2012).
11. Gargis, A. S. *et al.* Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nature Biotech.* **30**, 1033–1036 (2012).
12. Beadling, C. *et al.* Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping. *J. Mol. Diagn.* **15**, 171–176 (2013).
13. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials, Board on Health Care Services, Board on Health Sciences Policy & The Institute of Medicine. *Evolution of Translational Omics: Lessons Learned and the Path Forward* (eds Christine, M., Micheel, S. J. N. & Omenn, G. S.) (National Academies Press, 2012).
14. Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nature Rev. Genet.* **11**, 685–696 (2010).
15. Jia, W. *et al.* SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol.* **14**, R12 (2013).
16. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
17. Sathirapongsasuti, J. F. *et al.* Exome sequencing-based copy-number variation and loss of heterozygosity detection: exome CNV. *Bioinformatics* **27**, 2648–2654 (2011).
18. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotech.* **31**, 213–219 (2013).
19. Prahallad, A. *et al.* Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **483**, 100–103 (2012).
20. Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Rev. Genet.* **12**, 628–640 (2011).
21. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
22. Waldron, L., Simpson, P., Parmigiani, G. & Huttenhower, C. Report on emerging technologies for translational bioinformatics: a symposium on gene expression profiling for archival tissues. *BMC Cancer* **12**, 124 (2012).
23. Kerick, M. *et al.* Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *BMC Med. Genomics* **4**, 68 (2011).
24. Olson, E. M., Lin, N. U., Krop, I. E. & Winer, E. P. The ethical use of mandatory research biopsies. *Nature Rev. Clin. Oncol.* **8**, 620–625 (2011).
25. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
26. Swanton, C. Intratumor heterogeneity: evolution through space and time. *Cancer Res.* **72**, 4875–4882 (2012).
27. Gorre, M. E. *et al.* Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **293**, 876–880 (2001).
28. Carver, B. S. *et al.* Reciprocal feedback regulation of PI3K and androgen receptor signaling in PTEN-deficient prostate Cancer. *Cancer Cell* **19**, 575–586 (2011).
29. Mao, M. *et al.* Resistance to BRAF inhibition in BRAF-mutant colon cancer can be overcome with PI3K inhibition or demethylating agents. *Clin. Cancer Res.* **19**, 657–667 (2013).
30. Greger, J. G. *et al.* Combinations of BRAF, MEK, and PI3K/mTOR inhibitors overcome acquired resistance to the BRAF inhibitor GSK2118436 dabrafenib, mediated by NRAS or MEK mutations. *Mol. Cancer Ther.* **11**, 909–920 (2012).
31. Flaherty, K. T. *et al.* Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. *N. Engl. J. Med.* **367**, 1694–1703 (2012).
32. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
33. Welch, J. S. *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
34. Ahmed, J. *et al.* CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic Acids Res.* **39**, D960–D967 (2011).
35. Roychowdhury, S. *et al.* Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci. Transl. Med.* **3**, 111ra121 (2011).
36. Turner, N. & Grose, R. Fibroblast growth factor signalling: from development to cancer. *Nature Rev. Cancer* **10**, 116–129 (2010).
37. Kwak, E. L. *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* **363**, 1693–1703 (2010).
38. Le Tourneau, C. *et al.* Designs and challenges for personalized medicine studies in oncology: focus on the SHIVA trial. *Target. Oncol.* **7**, 253–265 (2012).
39. Herlyn, M. New dream team for melanoma therapy. *Pigment Cell Melanoma Res.* **25**, 279–280 (2012).
40. Gautschi, O. *et al.* A patient with BRAF V600E lung adenocarcinoma responding to vemurafenib. *J. Thorac. Oncol.* **7**, e23–e24 (2012).
41. Printz, C. "Stand up to cancer" focuses on accelerating promising research: campaign has raised approximately $180 million. *Cancer* **117**, 1557–1559 (2011).
42. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Rev. Genet.* **10**, 57–63 (2009).
43. Dietrich, S. *et al.* BRAF inhibition in refractory hairy-cell leukemia. *N. Engl. J. Med.* **366**, 2038–2040 (2012).
44. Simon, R. Clinical trials for predictive medicine: new challenges and paradigms. *Clin. Trials* **7**, 516–524 (2010).
45. Simon, R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Per. Med.* **7**, 33–47 (2010).
46. Slamon, D. J. *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344**, 783–792 (2001).
47. Horstmann, E. *et al.* Risks and benefits of phase 1 oncology trials, 1991 through 2002. *N. Engl. J. Med.* **352**, 895–904 (2005).
48. Sharma, M. R. & Schilsky, R. L. Role of randomized Phase III trials in an era of effective targeted therapies. *Nature Reviews Clin. Oncol.* **9**, 208–214 (2011).
49. Dagher, R. *et al.* Approval summary: imatinib mesylate in the treatment of metastatic and/or unresectable malignant gastrointestinal stromal tumors. *Clin. Cancer Res.* **8**, 3034–3038 (2002).
50. Parkinson, D. R., Johnson, B. E. & Sledge, G. W. Making personalized cancer medicine a reality: challenges and opportunities in the development of biomarkers and companion diagnostics. *Clin. Cancer Res.* **18**, 619–624 (2012).
51. Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).
52. Chalhoub, N. & Baker, S. J. PTEN and the PI3-kinase pathway in cancer. *Annu. Rev. Pathol.* **4**, 127–150 (2009).
53. Cheung, L. W. *et al.* High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. *Cancer Discov.* **1**, 170–185 (2011).
54. Carpten, J. D. *et al.* A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. *Nature* **448**, 439–444 (2007).
55. Banerji, S. *et al.* Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**, 405–409 (2012).
56. Hollander, M. C., Blumenthal, G. M. & Dennis, P. A. PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nature Rev. Cancer* **11**, 289–301 (2011).
57. Sato, T., Nakashima, A., Guo, L., Coffman, K. & Tamanoi, F. Single amino-acid changes that confer constitutive activation of mTOR are discovered in human cancer. *Oncogene* **29**, 2746–2752 (2010).
58. Chan, J. A. *et al.* Pathogenesis of tuberous sclerosis subependymal giant cell astrocytomas: biallelic inactivation of TSC1 or TSC2 leads to mTOR activation. *J. Neuropathol. Exp. Neurol.* **63**, 1236–1242 (2004).
59. Hardt, M., Chantaravisoot, N. & Tamanoi, F. Activating mutations of TOR (target of rapamycin). *Genes Cells* **16**, 141–151 (2011).
60. Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949–954 (2002).

61. Slamon, D. J. & Clark, G. M. Amplification of c-erbB-2 and aggressive human breast tumors? *Science* **240**, 1795–1798 (1988).
62. Johnson, R. L. *et al.* Human homolog of *patched*, a candidate gene for the basal cell nevus syndrome. *Science* **272**, 1668–1671 (1996).
63. Xie, J. *et al.* Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* **391**, 90–92 (1998).
64. Xie, J. *et al.* Mutations of the PATCHED gene in several types of sporadic extracutaneous tumors. *Cancer Res.* **57**, 2369–2372 (1997).
65. Gherardi, E., Birchmeier, W., Birchmeier, C. & Vande Woude, G. Targeting MET in cancer: rationale and progress. *Nature Rev. Cancer* **12**, 89–103 (2012).
66. Quintas-Cardama, A., Kantarjian, H., Cortes, J. & Verstovsek, S. Janus kinase inhibitors for the treatment of myeloproliferative neoplasias and beyond. *Nature Rev. Drug Discov.* **10**, 127–140 (2011).
67. Lapenna, S. & Giordano, A. Cell cycle kinases as therapeutic targets for cancer. *Nature Rev. Drug Discov.* **8**, 547–566 (2009).
68. Lens, S. M., Voest, E. E. & Medema, R. H. Shared and separate functions of polo-like kinases and aurora kinases in cancer. *Nature Rev. Cancer* **10**, 825–841 (2010).
69. Ryan, C. J. & Tindall, D. J. Androgen receptor rediscovered: the new biology and targeting the androgen receptor therapeutically. *J. Clin. Oncol.* **29**, 3651–3658 (2011).
70. Stirewalt, D. L. & Radich, J. P. The role of FLT3 in haematopoietic malignancies. *Nature Rev. Cancer* **3**, 650–665 (2003).
71. Shangary, S. & Wang, S. Targeting the MDM2–p53 interaction for cancer therapy. *Clin. Cancer Res.* **14**, 5318–5324 (2008).
72. Corless, C. L., Barnett, C. M. & Heinrich, M. C. Gastrointestinal stromal tumours: origin and molecular oncology. *Nature Rev. Cancer* **11**, 865–878 (2011).
73. Soda, M. *et al.* Identification of the transforming *EML4–ALK* fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
74. Chen, Y. *et al.* Oncogenic mutations of ALK kinase in neuroblastoma. *Nature* **455**, 971–974 (2008).
75. Bergethon, K. *et al.* ROS1 rearrangements define a unique molecular class of lung cancers. *J. Clin. Oncol.* **30**, 863–870 (2012).
76. Matulonis, U. A. *et al.* High throughput interrogation of somatic mutations in high grade serous cancer of the ovary. *PLoS ONE* **6**, e24433 (2011).
77. Weiss, G. J. *et al.* Paired tumor and normal whole genome sequencing of metastatic olfactory neuroblastoma. *PLoS ONE* **7**, e37029 (2012).

Acknowledgements
S.R. is supported by Pelotonia, the American Cancer Society (grant MRSG-12-194-01), the Landon-Foundation AACR Innovator Award for Personalized Cancer Medicine, and a Young Investigator Award from the Prostate Cancer Foundation. Special thanks to J. Bush for reading the manuscript and A. Marsell for administrative support.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
Catalogue of somatic mutations in cancer (COSMIC): http://cancer.sanger.ac.uk/cancergenome/projects/cosmic
DrugBank: http://www.drugbank.ca
My Cancer Genome: http://www.mycancergenome.org
Pharmacogenomics KnowledgeBase: http://www.pharmgkb.org
Therapeutic Targets Database: http://bidd.nus.edu.sg/group/cjttd

FURTHER INFORMATION
Sameek Roychowdhury's homepage: http://cancer.osu.edu/research/cancerresearch/facilities/labs/roychowdhury/pages/index.aspx
Cancer Diagnosis Program — US National Cancer Institute: http://www.cancerdiagnosis.nci.nih.gov
Cancer Genome Atlas: http://cancergenome.nih.gov
"CAP moves forward with next generation sequencing initiatives" (29 September press release; CAP website): http://www.cap.org/apps/cap.portal?_nfpb=true&cntvwrPtlt_actionOverride=%2Fportlets%2FcontentViewer%2Fshow&_windowLabel=cntvwrPtlt&cntvwrPtlt{actionForm.contentReference}=media_resources%2Fnewsrel_cap_moves_forward.html&_state=maximized&_pageLabel=cntvwr
CollabRx: http://www.collabrx.com
Division of Cancer Treatment and Diagnosis — US National Cancer Institute: http://dctd.cancer.gov
GeneInsight: http://geneinsight.com
International Cancer Genome Consortium: http://icgc.org
Lung Cancer Mutation Consortium: http://www.golcmc.com/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF