Comparative Effectiveness Research, Genomics-Enabled Personalized Medicine, and Rapid Learning Health Care: A Common Bond

Geoffrey S. Ginsburg and Nicole M. Kuderer

Geoffrey S. Ginsburg, Duke University Medical Center, Duke Center for Personalized Medicine, Institute for Genome Sciences and Policy; and Nicole M. Kuderer, Duke Cancer Institute, Duke University Medical Center, Duke University, Durham, NC.

Submitted February 23, 2012; accepted June 18, 2012; published online ahead of print at www.jco.org on October 15, 2012.

Supported by Grants No. 5UC2CA148041-02 (G.S.G.), No. 5UC2CA148041-02 (N.M.K.), and No. 1KM1CA156687-01 (N.M.K.) from the National Cancer Institute, and by an American Society of Clinical Oncology Young Investigator Award.

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

Corresponding author: Geoffrey S. Ginsburg, MD, PhD, Duke Institute for Genome Sciences & Policy, Duke University, 101 Science Dr, DUMC Box 3382, Office 2111, Durham, NC 27708; e-mail: Geoffrey.Ginsburg@duke.edu.

© 2012 by American Society of Clinical Oncology

0732-183X/12/3034-4233/\$20.00

DOI: 10.1200/JCO.2012.42.6114

ABSTRACT

Despite stunning advances in our understanding of the genetics and the molecular basis for cancer, many patients with cancer are not yet receiving therapy tailored specifically to their tumor biology. The translation of these advances into clinical practice has been hindered, in part, by the lack of evidence for biomarkers supporting the personalized medicine approach. Most stakeholders agree that the translation of biomarkers into clinical care requires evidence of clinical utility. The highest level of evidence comes from randomized controlled clinical trials (RCTs). However, in many instances, there may be no RCTs that are feasible for assessing the clinical utility of potentially valuable genomic biomarkers. In the absence of RCTs, evidence generation will require well-designed cohort studies for comparative effectiveness research (CER) that link detailed clinical information to tumor biology and genomic data. CER also uses systematic reviews, evidence-quality appraisal, and health outcomes research to provide a methodologic framework for assessing biologic patient subgroups. Rapid learning health care (RLHC) is a model in which diverse data are made available, ideally in a robust and real-time fashion, potentially facilitating CER and personalized medicine. Nonetheless, to realize the full potential of personalized care using RLHC requires advances in CER and biostatistics methodology and the development of interoperable informatics systems, which has been recognized by the National Cancer Institute's program for CER and personalized medicine. The integration of CER methodology and genomics linked to RLHC should enhance, expedite, and expand the evidence generation required for fully realizing personalized cancer care.

J Clin Oncol 30:4233-4242. © 2012 by American Society of Clinical Oncology

INTRODUCTION

This year a total of 1,638,910 new cases of cancer and 577,190 deaths from cancer are projected to occur in the United States.¹ Although overall cancer incidence rates have remained relatively stable from 2004 through 2008, cancer death rates have decreased by more than 1.5% per year for both men and women. Most patients with cancer receive treatment based on results from studies performed on hundreds, if not thousands, of similar patients, but not based on the genetics or biology of individuals or their disease.¹ However, cancer is clearly a heterogeneous disease whose presentation and response are likely determined by the patient's underlying genetics or biology. The past decade has witnessed unprecedented discovery and development of prognostic and predictive biomarkers that now offer an opportunity to stratify patients for risk of progression, to direct treatments to those most likely to respond, and to reduce unnecessary treatment toxicities caused by ineffective treatments that are not

going to be associated with improved outcome.² Thus, truly personalized oncology medicine is now considered a realistic approach for helping patients with cancer.^{3,4} This strategy has been enhanced by investment in projects such as The Cancer Genome Atlas (TCGA),⁵ which take advantage of high-throughput sequencing of specific genes in human tumors, in genome-wide surveys of chromosomal abnormalities, and in the development of gene expression data.⁵ Genome-wide data from these types of studies have contributed to the discovery of prognostic and potentially predictive biomarkers and the delineation of pathways that appear to drive the oncogenic phenotype (Table 1). Moreover, specific somatic mutations in genes encoding signaling molecules and their respective pathways have led to the development of small molecule antagonists and several targeted therapies aimed at these mutations and pathways (Table 2). The molecular heterogeneity of cancer resulting from the acquisition of multiple genetic alterations that contribute to the development of the tumor underlies the heterogeneity of

Table 1. Major Pathways Impacted by Somatic Genetic Changes

Major Pathway
Signal transduction
P13-kinase pathway
AKT pathway
Phospholipases
Protein kinase C (PKC) family
Mad family
Smad family
c-Met
Receptor tyrosine kinases (PDGFR, EGFR, KIT, etc.)
G-protein coupled receptor families
Jnk family
RHO family
Erk family
Mek family
Jak pathway
TGF pathways
Ras pathways
Src pathway
Wnt/frizzled pathway
Transcription factors
Myc family
Jun
Fos
NFκB family
Nuclear receptors (ER, AR, PPAR, RAR, etc.)
Myb
E2F family
STAT family
Cell cycle
Cdks
Cyclins
Cip/Kip family
Rb family
Ink family
CDK phosphatases
CDK kinases
Apoptosis regulators
Caspases
ICE family
FAS/FAS ligand
TRAIL
Bcl/Bad/Bax family
DNA damage/repair
p53
AT
PARP
Gadd45
MDM family
GST family
Rad52
DNA-dependent protein kinase
BRCA1/BARD1
Invasion/metastasis/angiogenesis
Integrins
MMP family
TIMP family
Laminin family
Cathepsin family
CD44
Endothelin

(continued in next column)

Table 1. Major Pathways Impacted by Somatic Genetic Changes (continued)

Major Pathway
Cadherin family
Catenin family
Thrombospondins
Metalloproteinases
Receptor tyrosine kinases
VEGF (and other growth factors)

Abbreviations: ADP, adenosine diphosphate; AR, adrenergic receptor; EGFR, epidermal growth factor receptor; ER, estrogen receptor; KIT, kinase tyrosine; NFκB, nuclear factor kappa B; PARP, poly ADP-ribose polymerase; PDGFR, platelet-derived growth factor receptor; PPAR, peroxisome proliferator-activated receptor; RAR, retinoic acid receptor; TRAIL, tumor necrosis factor-related apoptosis-inducing ligand; VEGF, vascular endothelial growth factor.

patient outcomes and response to therapy. Thus, it is clear that cancer is not a single disease but rather a collection of diseases with unique characteristics.^{6,7}

Three major challenges impede the implementation of these stunning advances in cancer genomics and hinder fundamental changes in clinical practice. First, there is a need to develop validated predictive biomarkers by matching existing therapies with data on individual patient outcomes. Second, there is a critical need to develop strategies linking genomic data to clinical outcomes data such that evidence for test utility can be evaluated systematically. Some of this evidence will be developed in the context of prospective clinical trials

Table 2. Major Targeted Cancer Therapies

Target	Disease	Drug	Clinical Development Status
PML-RAR-α	PML	Tretinoin	Approved
Bcr/abl	CML	Imatinib	Approved
		Dasatinib	Approved
c-Kit	GIST PDGF-α	Imatinib	Approved
Flt-3	AML	Semaxanib, midostaurin, tandutinib	Early-stage clinical trials
EGFR	NSCLC	Gefitinib	Approved/withdrawn
	NSCLC and pancreatic cancer	Erlotinib	Approved
Antiangiogenesis	MDS	Lenolidamide	Approved
Other major antiangiogenesis		Bevacizumab	Approved
		Thalidomide	Approved
		Sunitinib	Approved
		Sorafenib	Approved
Proteasome	MM	Bortezomib	Approved
HER2	Breast cancer	Trastuzumab	Approved
		Lapatinib	Approved
BRAF (V600E mutation)		Vemurafenib	Approved
EML4-ALK translocation		Crizotinib	Approved

Abbreviations: AML, acute myeloid leukemia; CML, chronic myeloid leukemia; EGFR, epidermal growth factor receptor; GIST, GI stromal tumor; HER2, human epidermal growth factor receptor 2; MDS, myelodysplastic syndrome; MM, multiple myeloma; NSCLC, non-small-cell lung cancer; PDGF-α, platelet-derived growth factor alpha; PML, promyelocytic leukemia; RAR-α, retinoic acid receptor alpha.

that test the hypothesis that biomarker-informed care is superior to the current standard-of-care methods—the highest level of evidence derived from comparative effectiveness research (CER). For ethical reasons or reasons of feasibility, in many situations, this evidence will need to be developed in the context of large observational cohort studies that are rich in genomic data and clinical and patient-reported outcomes measures. In either case, this would mean routinely obtaining tumor tissue by invasive biopsies—a major hurdle for combining genomic data with clinical outcomes data. Finally, there is the challenge of rigorously evaluating evidence consistently across organizations and making recommendations that drive the appropriate adoption of these novel biomarkers into clinical practice guidelines and clinical use.

There is growing optimism about the successful application of genomic understanding and modern technology so that treatment choices are based on the individual and the biology of their disease and not based on a population. However, without the ability to test hypotheses for clinical and molecular subsets of patients in both clinical trial and real-world settings, oncology will remain a population-based approach and not an individualized one. The current emphasis in CER represents, perhaps paradoxically, an opportunity to catalyze the development of the data that will enable our ability to stratify oncology populations more robustly than before and drive the current cancer practice toward personalized medicine.

CER: WHAT IS IT AND WHY IS IT IMPORTANT?

Compared with the efficacy of an intervention (the extent to which the intervention does more good than harm under ideal circumstances in highly selected patients) assessed in a randomized clinical trial (RCT), CER evaluates the effectiveness of an intervention in real-world practice settings.^{8,9} CER involves a comparison of the effectiveness of two (or more) interventions to determine which works best for a given health care problem. CER usually evaluates treatment effects across large study populations and therefore reports the average treatment effects. In general, therefore, CER does not investigate whether patient subgroups exhibit differences in response to an intervention; thus, it appears to be at odds with personalized medicine. A recent study¹⁰ showed that only 13% of CER studies focused on subpopulations other than white middle-age adults. In fact, CER studies have rarely accommodated the collection of genomic data. However, recent institutional and governmental investments in resources that support collecting CER data combined with genomic data should provide the opportunity for analysis at the subpopulation level and allow examination of the effectiveness of an intervention in patient subgroups. The resulting evidence from these data will likely better enable personalized medicine.

PERSONALIZED MEDICINE: WHAT IS IT AND WHY IS IT IMPORTANT?

Personalized medicine is a concept for health care, but it is not a process like CER. It is a broad and rapidly advancing field that is informed by each person's unique clinical, molecular, genomic, and environmental information. Health care that embraces personalized medicine ideally is individualizing patient care across the continuum

(from health to disease).¹¹ The overarching goal of personalized medicine is to optimize outcomes for each individual through an unprecedented customization of patient care. Our ability to fully execute this vision of care for oncology patients is hampered, in part, by our inability to link molecular data from the individual or their tumor to clinical, treatment, and outcomes data. The linkage of clinical data with genomic information could potentially result in the development of predictive clinico-molecular profiles that will further enhance personalized oncology care.

THE CER PROCESS

Lyman⁸ provides a summary framework of the CER process and its importance in guidelines development (Fig 1), beginning with systematic reviews of the literature and meta-analyses, evidence synthesis, and effectiveness decision modeling. The CER framework also makes use of existing methods of epidemiologic, population, and health outcomes research and evidence-based medicine by using data from prospective and retrospective cohort studies and registries, but with a focus on nonselected patients from general practice who provide key information on the real-world patient experience.¹² Cohort or population studies can also further inform the design of definitive CER RCTs with a special focus on treatment effectiveness in the real-world setting (rather than clinical trials of efficacy, which limit trial participants to highly selected otherwise healthy patient populations) to answer specific clinical and biomarker questions.⁸ To make recommendations for using genomic biomarkers in caring for patients with cancer, there needs to be a process that evaluates the substantial and rapidly increasing body of evidence for directed strategies from laboratory, translational, population, behavioral, and modeling research that may inform clinical choices and guidelines, regulatory and health policy decisions, as well as the design and analysis of confirmatory CER trials.^{13,14} In this sense, CER is an important component of evidence-based medicine (EBM) that provides research results for the systematic review, evidence synthesis, and evaluation of EBM. CER expands EBM's reach beyond RCTs to the general patient population and to special subpopulations (eg, the elderly, pregnant women, minorities, or patient subsets with comorbid conditions) that

Fig 1. Schema of evidence sources for comparative effectiveness research (CER) in oncology that provide opportunities for evidence quality appraisal and synthesis, clinical decision simulation and evaluation, and the development of clinical practice guidelines.⁸

are not routinely included in efficacy clinical trials. To better define the gaps in our knowledge for genomic approaches to oncology and future directions for CER, we need to do the following: (1) develop and apply timely systematic reviews and analytic tools, which include optimal searches, selection, abstraction, quality appraisal, and analysis that provide for rigorous evidence-based evaluation of genomics approaches to personalized cancer care; (2) establish and use disease-focused multidisciplinary research teams of translational clinical investigators, genomic experts, biostatisticians, and health outcomes research methodologists to conduct and evaluate the data analysis and systematic reviews; (3) integrate the evidence synthesis with the evaluation of emerging data from the longitudinal registries, clinical trials, or pragmatic trials to guide the selection of genomic biomarkers for optimally designed phase III confirmatory studies; (4) develop and evaluate clinical simulation models of critical decision strategies; and (5) provide clinical and policy recommendations through the formulation of evidence-based clinical practice guidelines.^{8,14}

SYNERGIES AND OPPORTUNITIES BETWEEN CER AND PERSONALIZED MEDICINE

Personalized medicine interventions—diagnostics, genomics, and so on—should be evidence based. Increasingly, this means showing that an intervention has some demonstrably favorable impact on health outcomes in real-world practice settings. For genetic and genomic testing, this requires demonstrating not only that a test can accurately detect a particular gene or biomarker (analytic validity; Table 3), but that the test result reliably identifies or predicts a corresponding disease or clinical phenotype (clinical validity; Table 3). Ultimately, genomic information, like that from other diagnostic tests, needs to improve clinical decision making and have a positive impact on relevant patient health outcomes (clinical utility; Table 3). Most genomics-based tests in use today have demonstrated analytic validity and clinical validity but have yet to demonstrate clinical utility.¹⁵ The lack of clinical utility data is a result of the paucity of outcomes data associated with the test in either clinical trials or in clinical practice. Given that it is highly unlikely that RCTs will be conducted to determine clinical utility for the myriad potentially valuable genomic diagnostics, the greatest hope for evidence generation for these novel testing platforms will be either from testing hypotheses in the context of RCTs from which appropriate biologic specimens and clinical data have been obtained or in the context of cohort studies or biorepositories, in

which there is linkage of biologic specimens and molecular and reliable clinical outcomes data from routine clinical care.

CYP2D6 Genotyping

There have been several poignant examples in oncology and in other disease areas that illustrate the need for rigorous evidence for biomarkers and the systems to provide it. The debate over testing for genetic variants in *CYP2D6* that are associated with tamoxifen metabolism in the context of decisions about therapy for breast cancer is one of those cases that highlights the need for evidence of clinical utility to support testing and treatment decisions. The pharmacology of tamoxifen suggests that the drug depends on being converted by a drug-metabolizing enzyme, cytochrome P450 2D6 (*CYP2D6*), from a prodrug to potent antiestrogen metabolites, especially endoxifen.¹⁶ Approximately 7% to 10% of white patients have reduced *CYP2D6* activity resulting from a non- or underfunctioning polymorphism of the *CYP2D6* gene.¹⁷ Initially, studies of long-term outcomes had been mostly retrospective, and some suggested that women with *CYP2D6* polymorphisms have a higher risk of recurrence than women with wild-type *CYP2D6*,^{17,18} although within these studies, there were conflicting results.^{19,20} More recently, prospective-retrospective analyses of DNA samples from the large prospective RCTs of tamoxifen have failed to show an association between *CYP2D6* genotype and clinical outcomes in patients taking tamoxifen, despite attempting to adjust for major clinical confounders.^{21–23} Thus, currently *CYP2D6* genotyping is not indicated for women with estrogen receptor–positive breast cancer who are candidates for tamoxifen therapy.^{19–21,24,25} However, it is possible that even in these prospective RCTs, data on key confounders for tamoxifen metabolism were not sufficiently available for a definitive analysis because of the retrospective nature of the analysis. Prospective cohort studies designed to evaluate the association between specific tamoxifen metabolites and clinical outcomes, and that also assess and adjust for key confounders, should be able to further elucidate whether this particular pharmacogenomic research question is worth any additional inquiry.¹⁹

KRAS Mutation Testing

Conversely, for *KRAS* mutation detection testing in colorectal cancer, the evidence resulting from several prospective-retrospective analyses of samples collected from multiple drug registration trials demonstrated that individuals with wild-type *KRAS* had significant improvements in tumor response with epidermal growth factor receptor (EGFR) inhibitors cetuximab or panitumumab and chemotherapy compared with patients with mutated *KRAS*.²⁶ Although these biomarker data were exclusively assessed retrospectively by using prospective trial data, the consistency of numerous registration RCTs for existing EGFR inhibitors in conjunction with the high rates of tumor specimen collection in the more recent trials was deemed compelling enough evidence that in July 2009, the US Food and Drug Administration announced revisions to the prescribing information for EGFR inhibitors and colorectal cancer and limited their drug label to the approximately 60% of individuals whose tumors harbor the wild-type *KRAS* gene.²⁷

Other Examples of Genomic Tests

In *KRAS* testing for EGFR inhibitors, prospective RCTs were available in which appropriate outcomes data were able to be linked to genomic data to develop the strongest available evidence. In other

Table 3. Evaluation of Genomic Tests and Other Biomarkers

Assessment Level	Description
Analytic validity	A test's ability to accurately and reliably measure the molecular target of interest (genotype)
Clinical validity	A test's ability to detect or predict the associated disorder (phenotype), including clinical sensitivity (or the clinical detection rate), clinical specificity, and positive and negative predictive values
Clinical utility	A test's ability to affect clinical decisions and to improve patient outcomes in clinical practice

Adapted from the Centers for Disease Control's Office of Public Health Genomics.^{14a}

settings of genomic markers with limited data from RCTs, or with limited data on clinical utility in a particular subpopulation, a prospective RCT has been recommended by the National Cancer Institute (NCI) and other agencies for the purpose of definitive biomarker validation. Examples include the ongoing MINDACT [Microarray in Node-Negative and 1-3 Node-Positive Disease May Avoid Chemotherapy] trial for the MammaPrint 70-gene RNA profile in patients with node-negative and node-positive breast cancer,²⁸ the TAILORx [Trial Assigning Individualized Options for Treatment (Rx)] trial for Oncotype DX 21-gene RNA profile testing in patients with intermediate-risk, node-negative breast cancer,²⁹ and the RxPONDER [Rx for Positive Node, Endocrine Responsive Breast Cancer] trial in patients with one to three positive lymph nodes.³⁰ However, in many other instances, there may be no RCTs that are feasible or ethically justifiable. This is especially the case in the biomarker development phase. Once a specific treatment is the established standard of care, that makes it unethical to randomly assign patients to a control arm of no therapy until there are sufficient data on the biomarker's clinical validity (Table 3). Thus, to assess the clinical validity of a new predictive or prognostic test without existing RCT data will require rigorously designed cohort studies, ideally prospective CER studies in real-world settings. Prospective CER cohort studies—with the biomarker as the primary reason for the study instead of the drug as the focus of the primary study design—may permit more comprehensive assessments of the biomarker under study because they allow for more extensive adjustment of clinical confounding factors. Furthermore, these types of studies may investigate and incorporate other outcomes of interest that are not routinely captured in RCTs, such as patient-reported outcomes or cost information.

WHAT KIND OF EVIDENCE IS NEEDED AND HOW IS EVIDENCE GENERATED?

The stakeholder community for personalized medicine is diverse and includes regulators, policymakers, health care providers, payers, academia, industry, government researchers, and patients. In a perfect world, the evidentiary threshold for acceptance (adoption, reimbursement, and regulation) would be the same for all groups. It is clear, however, that more dialogue and coordination among stakeholders is needed to facilitate the development of the necessary evidence base. It is equally apparent that test development and reimbursement need to focus on the clinical utility of the test and the net benefit to patients.¹⁴ Finally, the analysis of the evidence must be adapted to the clinical setting and to the evidence needed for a particular application.

No organization has owned the evaluation of genetic testing; it has been a distributed process with widespread heterogeneity among stakeholders regarding the evidentiary threshold for test acceptance or adoption. Hayes et al³¹ developed a specific framework for evaluating the clinical utility of tumor markers by using a utility grading system and assessing the level of evidence. Altman and Lyman³² have specified a framework for designing breast cancer prognostic marker studies with a focus on methodologic challenges. More recent studies^{14,33} have further refined biomarker development frameworks and the level of evidence for clinical utility. In 2005, the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) working group was established as an independent evidentiary review body that emerged

from a Centers for Disease Control and Prevention (CDC) –funded project to develop evidence recommendations on the appropriate use of genetic tests.^{34,35} Challenges to establishing analytic validity are that necessary information is often missing or proprietary, clinical validity may be biased by patient selection, and clinical utility suffers from the lack of necessary RCTs or biased observational studies.³⁶

In oncology, the evidence for the clinical utility of biomarkers or genetic tests traditionally has been generated in three major ways: (1) performing prospective clinical trials in which the marker is the primary objective, of which there are only a few; (2) using archived specimens, ideally from RCTs or prospective observational studies; and (3) by performing retrospective analysis of biorepositories or tumor registries. Table 4 is a summary of major experimental methods that can be used to generate evidence that might serve as the basis for test recommendations and possible future clinical adoption. Significant investment will be required to develop and refine methods that are better suited to address questions at the subgroup level and to achieve adequate statistical power to make valid inferences for such subgroups.

NCI PROGRAM IN CER AND PERSONALIZED MEDICINE

Against the backdrop of major advances in the field of cancer biomarker research and other advances in cancer genomics, the NCI initiated an American Recovery and Reinvestment Act (ARRA)–funded program aimed at systematic research to compare the clinical effectiveness and cost-effectiveness of cancer care and prevention based on genomic tools and markers. The *NCI Guidelines for ARRA Research and Research Infrastructure Grand Opportunities: Comparative Effectiveness Research in Genomic and Personalized Medicine* opportunity offered in 2009 supported 2-year efforts to advance methods for analysis, synthesis, modeling, and evaluation of the clinical validity and utility of existing and emerging genomic and personalized medicine applications in cancer control and prevention, to accelerate the development of genomic and personalized medicine by planning CER initiatives, and to enhance clinical and population data infrastructure to support current and future CER initiatives in genomic and personalized medicine.

Seven research proposals were funded under the approximately \$28M 2-year program (Appendix Table A1, online only) that covers a spectrum of interventions from family history utility for cancer screening and prevention to genetic risk testing and pharmacogenomics. What is extremely attractive and unique about these proposals is their innovation in CER to help bring about genomic and personalized care paradigms in oncology. For example, several are taking advantage of the electronic medical record to provide prospective outcomes data in response to predefined interventions (family history, genetic risk testing, or pharmacogenetics) using case cluster designs, and others are using the electronic medical record to define cohorts of clinical and genomic subpopulations and to carefully examine their outcomes to various therapeutics. Some proposals are developing integrated data warehouses that contain linked sample data, patient data, and molecular data, all in the context of usual care. These structures may provide the opportunity to evaluate the effectiveness of standard and targeted therapies not only from clinical and

Table 4. Various Methods for Developing Evidence for CER and Personalized Medicine

Evidence Type	Description	Advantage or Limitations for CER	Advantage or Limitations for Personalized Medicine	Example
Randomized controlled trials	<p>Gold standard for establishing an intervention's causal effect on patient health outcomes in predefined, carefully controlled conditions</p> <p>Can establish efficacy and safety for drugs and some medical devices</p> <p>Can establish utility for diagnostic tests</p> <p>Comparator is typically placebo or no treatment</p> <p>Typically patient-level randomization and analyses</p>	<p>Usually does not generate evidence for real-world effectiveness in diverse patient populations</p> <p>If cluster randomization is used with an active comparator (usually standard of care), this can generate high-quality CER data</p>	<p>Subgroup analyses and analyses focusing on variability rather than those that average main effects (ie, identify groups of individuals that do well or poorly) can identify hypothesis generating analyses for potential high-impact groups (which can be based on genomic data, if collected) to be used in future randomized controlled trials to establish causality, efficacy, safety, and/or clinical utility</p>	<p>MINDACT trial uses the MammaPrint 70-gene profile³⁷; TAILORx trial uses the Oncotype Dx 21-gene assay³⁸</p> <p>To provide prospective evidence about the clinical utility of these tests, to determine the need for adjuvant chemotherapy, and to identify women who are unlikely to benefit from chemotherapy</p>
Pragmatic clinical trials	<p>Standard for establishing an intervention's effect on patient health outcomes in real-world settings and populations (diverse/heterogeneous)</p> <p>Outcomes include a broad range of measures that are relevant to decision makers, clinicians, and patients</p> <p>Comparator is typically standard of care</p> <p>Clinician-, clinic-, or system-level randomization and analyses (cluster)</p>	<p>Results are generalizable and useful for real-world effectiveness</p> <p>Effect sizes may be small (but still clinically meaningful) when comparing the impact of active treatments; thus, to generate clinically meaningful and statistically significant results, very large sample sizes may be required (with their associated high costs)</p>	<p>Including genomic data in study measures will begin to generate effectiveness outcomes for personalized medicine</p> <p>Currently, this type of data is not commonly captured. To do so will require a health information technology data infrastructure</p>	<p>Family History Project uses a hybrid type 2 effectiveness and implementation model³⁹</p> <p>To assess the impact and operating characteristics of a family history decision support intervention for cancer screening on patient and provider behaviors</p>
Adaptive clinical trials	<p>Strives for the most effective use of an intervention (optimized) to maximize its effect on patient health outcomes in real-world settings and populations (diverse/heterogeneous)</p> <p>Outcomes include a broad range of measures that are relevant to decision makers, clinicians, and patients</p> <p>Measures include data for assessing the population distribution probability of an outcome at baseline (prior distribution) and after the intervention (posterior distribution)</p> <p>Data are analyzed frequently (at set intervals) and used to inform the study direction</p> <p>Comparator is baseline (prior to intervention) and standard of care</p>	<p>Because the intervention's form or use is modified during the study to maximize its impact, adaptive clinical trials do not usually generate adequate effectiveness data although they do capture some</p> <p>They can be used to inform future effectiveness studies, once optimization has occurred</p>	<p>Since a goal of adaptive clinical trials is to modify study design to optimize an intervention's effect, they not only gather data on personalized medicine characteristics (including genomic data, if collected) but also shift resources to focus enrollment on these groups</p> <p>Results can inform the design of future definitive personalized medicine randomized controlled trials or effectiveness studies</p>	<p>I-SPY-2 breast cancer trial systematically assigns phase II agents while rapidly learning about the impact of these agents on patients based on specific molecular characteristics of their tumors. Drugs that show beneficial changes will be preferentially assigned and will move through the trial more rapidly. Agents that do not show the likelihood of improved response rate associated with any predefined biomarker will be dropped from the trial⁴⁰</p>
Observational cohort studies and registries	<p>Generates hypotheses about linkages between an intervention and patient health outcomes in real-world settings and populations (diverse/heterogeneous)</p> <p>There is no randomization; all interventions are provided in the course of normal clinical care</p> <p>Outcomes typically include those available in administrative databases, such as clinical variables and health care use</p> <p>Measures are usually longitudinal and available at unstructured time points</p>	<p>Very useful for generating CER hypotheses but does not establish causality</p> <p>How useful it is depends on the quality and extent of data available for collection</p>	<p>Very useful for generating personalized medicine hypotheses and for performing biomarker development and initial validation studies but does not provide final proof for the link between personalized medicine characteristics/subgroups and health outcomes</p> <p>How useful it is depends on the quality and extent of data available for collection (including whether genomic data are available)</p> <p>Results can inform the design of future definitive personalized medicine randomized controlled trials or effectiveness studies</p>	<p>TCGA seeks to identify somatic mutations that are specifically associated with the type of cancer being sequenced and other types of larger-scale genomic changes, such as copy number changes and/or chromosomal translocations, and gene expression that contribute to cancer development and/or progression⁴¹</p>

Abbreviations: CER, comparative effectiveness research; MINDACT, Microarray in Node-Negative and 1-3 Node-Positive Disease May Avoid Chemotherapy trial; TCGA, The Cancer Genome Atlas.

Fig 2. (A) The rapid-learning health system model. (B) The rapid learning health system applied to comparative effectiveness research and personalized medicine. CER, comparative effectiveness research.

genomic levels but also from an economic and patient-reported perspective. As a result of these efforts, novel multidisciplinary approaches and structures have been formed that may provide a unique and completely different strategy to CER in genomics⁴² that may become a model for others to emulate. Collectively, these groups have expanded methods for capturing genomics data from the literature and other sources, which form the basis for evidentiary review and the generation of novel hypotheses for the impact of genomic medicine on outcomes.^{43,44} Although it is still early in the lifecycle of research to evaluate the value of this investment, it is likely that these CER projects will advance the field of personalized genomic medicine.

BUILDING A BRIDGE BETWEEN THE LABORATORY AND THE CLINIC

The Rapid Learning Health Care Model Linked to CER and Personalized Medicine

Translational cancer research and the personalization of health care are made possible by matching increasingly detailed patient data (eg, demographics, clinical and treatment parameters, behavioral indicators, tumor and blood biomarkers, genetic and genomic data) with health outcomes data, which simultaneously allows the comparison of outcomes of clinical practice to best available evidence. Rapid learning health care (RLHC) is a model in which such diverse and interoperable data are made available, ideally in a robust and real-time fashion, to potentially support clinical practice while simultaneously providing possibilities to fuel

CER and innovation (Fig 2A). Ideally, in RLHC, clinical information is entered automatically into a comprehensive data system (eg, a data warehouse), is aggregated continuously at the point of care, and is analyzed both for clinical application and for generation of research evidence. This strategy aims to make discovery a natural outgrowth of patient care and also has the potential to stimulate innovation, quality, safety, and value and to help bridge the gap between clinical care and research.⁴⁵⁻⁴⁷ The quality and availability of these comprehensive data and the resulting need for data quality assessment and processing and data cleaning and modeling are central concerns for the development of a realistic and realized model of rapid learning for CER that enables personalized and genomics-guided medical care. RLHC holds great promise for developing large and diverse databases in a time- and cost-effective manner, which would be otherwise unthinkable. From an analyst's point of view, RLHC would likely enable the linkage of clinical and biomarker information to other key patient subgroup outcomes of interest that are often elusive, including health care use, quality of life, and patient-reported outcomes data. The data attributes of the RLHC model fundamentally depend on the establishment of a sound, reliable, and flexible health informatics system, substantial biostatistical support, and involvement of health outcomes methodologists. The full potential of personalized care using RLHC requires advances in CER and biostatistics methodology and the development of interoperable informatics systems, both of which are areas of focus of the NCI's program for CER and personalized medicine.

Linking Clinical, Health Outcomes, and Genomics Data and CER and Rapid Learning in Oncology

NCI ARRA funding has enabled the development of a personalized CER model at Duke University whose objectives are to evaluate the association between biomarkers and cancer outcomes (Fig 2B). Modeled as much as possible after an existing prospective cancer cohort study^{48,49} and other variables from randomized trials, the data infrastructure plans to link mostly electronically captured demographic, clinical, treatment, and laboratory data with tumor and blood biospecimens, genomic data, tumor response and health outcomes data, as well as electronically captured patient-reported outcomes. These data, merged with data on health care resource use residing in the Duke University Health System data warehouse, facilitate the evaluation of biomarkers along with clinical outcomes research and economic analyses.

Systematic Review and Quality Appraisal of Clinical Multigene Prediction Assays of Chemotherapy Response

Despite the enthusiasm concerning the potential for genomic predictors to provide a more personalized approach to selecting chemotherapeutic regimens, there is still considerable uncertainty about and controversy over optimal study methodology and validity for this new technology.^{50,51} In accordance with the second major research component of the NCI-funded Duke program on *Genomic Comparative Effectiveness Research*, we developed multiple comprehensive systematic literature searches and appraised study quality specifically for multigene microarray signatures predictive of response to systemic chemotherapy.^{43,52,53} Data were extracted in a dual-blinded fashion. In addition, a formal study of quality appraisal was developed and performed by using a quality scoring system modified from EGAPP, REMARK [Reporting Recommendations for Tumor Marker Prognostic Studies], the STREGA [Strengthening the Reporting of Genetic Association Studies] statements, and others.^{34,54-56} Our early data show that reported validation studies of multigene microarray prediction signatures of chemotherapy benefit vary enormously in terms of study design, methodology, and quality; in addition, most studies fail to report genomic model specifications and standard clinical prognostic or predictive factors.^{43,52}

CHALLENGES TO DEVELOPING THE COMMON BOND BETWEEN CER AND PERSONALIZED GENOMIC MEDICINE

The Need for a Cancer Informatics Infrastructure

The technical capability to generate the vast quantities of molecular data required to enable personalized medicine is increasingly not the rate-limiting step in clinical translational research. The larger challenge, as we have already mentioned, is in linking molecular data with clinical data. The vast majority of clinical data—from clinical trials, clinical research, and clinical care—are often collected in proprietary data formats that seriously restrict the ability to use them for downstream integrated analysis. As multisite trials become commonplace, collecting data in a uniform standard across different systems at diverse locations becomes critical. Developing data standards for molecular data is equally critical.

The Need for Novel Biostatistics Methodology, Biostatisticians, and Health Outcomes Methodologists

When the clinical and molecular data derived from tumors and blood specimens are aggregated into data warehouses with genes, proteins, metabolites, and images, and when methods are available to convert that information into analyzable data sets, the resulting analytic challenges are formidable. When the multivariable (multiple interacting inputs) and multivariate (multiple interacting outcomes) characteristics of biologic systems of cancer are appreciated and are to be incorporated into clinical practice, the use of statistical modeling will become indispensable, and continuing advances are needed in the field of biostatistics. The problems of dealing with the uncertainty caused by thousands of observations per individual patient are now made more severe by the potential presence of tens of thousands of patients, each with thousands of measurements. The issues of dealing with nonindependent observations, variable clustering, and multivariate outcomes require considerable attention. Missing data, particularly in longitudinal analysis, remain problematic. Even more problematic is the issue of developing synthetic models from disparate data sources with differing data quality and reliability. The science of evaluating predictions and characterizing probabilistic statements in the human context will require significant ongoing research before real-time data integration into clinical practice. Finally, explaining advanced statistics to the medical community will take a concerted effort, because the commonly familiar statistical tools will not suffice in the context of these highly complex new data sets and analyses.

The Need for Linked Biorepositories

The NCI came to an important realization regarding the cost of high-quality biospecimens through its TCGA project.⁵ The underlying premise was that TCGA data would be directly integrated because investigators across all participating centers would conduct analyses on the same biomolecules extracted from the same biospecimen derived from the same tumor. When a search was conducted for samples that met the technical requirements of the project (ie, 500 samples of each type of cancer along with matched normal tissue or blood from the same patient for a reference genome), the NCI discovered that 75% to 95% of retrospective samples across the country failed to meet high-quality standards for molecular analysis across multiple platforms.⁵⁷ Academic institutions have, in general, considered that there is no scientific merit to the coordination of endeavors to bank biologic specimens; thus, specimen banking has been undertaken with only minimal funding, without adequate management or informatics and data support, and under the institutional radar.⁵⁸

The ethics and societal rules governing the use of these inevitably well-organized and valuable tissue repositories must be unified, clarified, and continuously updated to deal with our increasing capacity for using biologic information for human benefit or detriment. This mix of technology and oversight is obviously beyond the capabilities of most individual investigators and requires the exploitation of economies of scale at higher institutional levels. It is also important that these initiatives be planned to not only span specific diseases but also to span institutions and promote cross collaboration and sharing of the precious resources and samples archived in these biorepositories. This will allow rapid comparison and validation of new findings and predictions. The NCI, through the cancer Biomedical Informatics Grid (caBIG) community, has started to prototype a solution to this challenge by adopting the existing and widely recognized data standards

and by working closely with various organizations to develop vocabularies and define common data elements and data models where there are currently no such standards.

SUMMARY

The rapid pace of innovation and the often limited evidence for clinical utility of existing biomarkers present significant challenges to realizing the full potential of personalized medicine in oncology. Development of an integrated CER strategy in oncology that embraces the rapid learning system model combined with a strong foundation in statistical methodology and health outcomes research and linked to cancer genomics holds promise in enhancing the quality, synthesis, and generation of evidence and improving the translation of important genomic biomarkers into clinical guidelines and clinical practice.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Although all authors completed the disclosure declaration, the following author(s) and/or an author's immediate family member(s) indicated a

financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.

Employment or Leadership Position: None **Consultant or Advisory Role:** Geoffrey S. Ginsburg, Universal Oncology (U); Nicole M. Kuderer, sanofi-aventis (U) **Stock Ownership:** Geoffrey S. Ginsburg, Universal Oncology **Honoraria:** None **Research Funding:** Geoffrey S. Ginsburg, Pfizer; Nicole M. Kuderer, Amgen **Expert Testimony:** None **Other Remuneration:** None

AUTHOR CONTRIBUTIONS

Conception and design: All authors

Financial support: All authors

Administrative support: All authors

Provision of study materials or patients: All authors

Collection and assembly of data: All authors

Data analysis and interpretation: All authors

Manuscript writing: All authors

Final approval of manuscript: All authors

REFERENCES

1. Siegel R, Naishadham D, Jemal A: Cancer statistics, 2012. *CA Cancer J Clin* 62:10-29, 2012
2. Dancey JE, Bedard PL, Onetto N, et al: The genetic basis for cancer treatment decisions. *Cell* 148:409-420, 2012
3. Schilsky RL: Tumor heterogeneity as the foundation of personalized cancer treatment. *Semin Oncol* 38:171-172, 2011
4. Freedman AN, Sansbury LB, Figg WD, et al: Cancer pharmacogenomics and pharmacoepidemiology: Setting a research agenda to accelerate translation. *J Natl Cancer Inst* 102:1698-1705, 2010
5. National Cancer Institute: The National Genome Atlas. <http://cancergenome.nih.gov/>
6. Sorlie T, Tibshirani R, Parker J, et al: Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 100:8418-8423, 2003
7. Parker JS, Mullins M, Cheang MC, et al: Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 27:1160-1167, 2009
8. Lyman GH: Comparative effectiveness research in oncology: The need for clarity, transparency and vision. *Cancer Invest* 27:593-597, 2009
9. Lyman GH, Hirsch B: Comparative effectiveness research and genomic personalized medicine. *Personalized Med* 7:223-227, 2010
10. Jacobson GA: Comparative clinical effectiveness and cost-effectiveness research: Background, history, and overview. Washington, DC, Congressional Research Service, 2007
11. Ginsburg GS, Willard HF: Genomic and personalized medicine: Foundations and applications. *Transl Res* 154:277-287, 2009
12. Sox HC, Goodman SN: The methods of comparative effectiveness research. *Annu Rev Public Health* 33:425-445, 2012
13. Somerfield MR, Einhaus K, Hagerty KL, et al: American Society of Clinical Oncology clinical practice guidelines: Opportunities and challenges. *J Clin Oncol* 26:4022-4026, 2008
14. Febbo PG, Ladanyi M, Aldape KD, et al: NCCN Task Force report: Evaluating the clinical utility of tumor markers in oncology. *J Natl Compr Canc Netw* 5:S1-S32, 2011 (suppl 9)
- 14a. Centers for Disease Control: Office of Public Health Genomics. <http://www.cdc.gov/genomics/gtesting/ACCE/index.htm>
15. Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS): U.S. System of Oversight of Genetic Testing: A Response to the Charge of the Secretary of Health and Human Services. http://oba.od.nih.gov/oba/sacghs/reports/sacghs_oversight_report.pdf
16. Jordan VC, Collins MM, Rowsby L, et al: A monohydroxylated metabolite of tamoxifen with potent antioestrogenic activity. *J Endocrinol* 75:305-316, 1977
17. Goetz MP, Rae JM, Suman VJ, et al: Pharmacogenetics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. *J Clin Oncol* 23:9312-9318, 2005
18. Schroth W, Antoniadou L, Fritz P, et al: Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. *J Clin Oncol* 25:5187-5193, 2007
19. Kuderer NM, Peppercorn J: CYP2D6 testing in breast cancer: Ready for prime time? *Oncology (Williston Park)* 23:1223-1232, 2009
20. Higgins MJ, Rae JM, Flockhart DA, et al: Pharmacogenetics of tamoxifen: Who should undergo CYP2D6 genetic testing? *J Natl Compr Canc Netw* 7:203-213, 2009
21. Kelly CM, Pritchard KI: CYP2D6 genotype as a marker for benefit of adjuvant tamoxifen in postmenopausal women: Lessons learned. *J Natl Cancer Inst* 104:427-428, 2012
22. Rae JM, Drury S, Hayes DF, et al: CYP2D6 and UGT2B7 genotype and risk of recurrence in tamoxifen-treated breast cancer patients. *J Natl Cancer Inst* 104:452-460, 2012
23. Regan MM, Leyland-Jones B, Bouzyk M, et al: CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: The Breast International Group 1-98 Trial. *J Natl Cancer Inst* 104:441-451, 2012
24. Leyland-Jones B, Regan MM, Bouzyk M, et al: Outcome according to CYP2D6 genotype among postmenopausal women with endocrine-responsive early invasive breast cancer randomized in the BIG 1-98 Trial. 33rd Annual San Antonio Breast Cancer Symposium, San Antonio, TX, December 8-12, 2010 (abstr S1-8)
25. Rae JM, Drury S, Hayes DF, et al: Lack of correlation between gene variants in tamoxifen metabolizing enzymes with primary endpoints in the ATAC Trial. 33rd Annual San Antonio Breast Cancer Symposium, San Antonio, TX, December 8-12, 2010 (abstr S-1-7)
26. Allegra CJ, Jessup JM, Somerfield MR, et al: American Society of Clinical Oncology provisional clinical opinion: Testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J Clin Oncol* 27:2091-2096, 2009
27. US Food and Drug Administration, Office of Medical Products and Tobacco: Cetuximab (Erbix) and panitumumab (Vectibix): Class labeling changes to anti-EGFR monoclonal antibodies, cetuximab (Erbix) and panitumumab (Vectibix)—KRAS Mutations. <http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm172905.htm>
28. Rutgers E, Piccart-Gebhart MJ, Bogaerts J, et al: The EORTC 10041/BIG 03-04 MINDACT trial is feasible: Results of the pilot phase. *Eur J Cancer* 47:2742-2749, 2011
29. Eastern Cooperative Oncology Group: Patient Education Materials Available for TAILORx. <http://www.ecog.org/general/tailorx.html>
30. SWOG: SWOG RxPONDER trial (S1007): Patient information. <http://www.swog.org/Visitors/S1007/patients.asp>
31. Hayes DF, Bast RC, Desch CE, et al: Tumor marker utility grading system: A framework to evaluate clinical utility of tumor markers. *J Natl Cancer Inst* 88:1456-1466, 1996

32. Altman DG, Lyman GH: Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Res Treat* 52:289-303, 1998
33. Simon RM, Paik S, Hayes DF: Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst* 101:1446-1452, 2009
34. Teutsch SM, Bradley LA, Palomaki GE, et al: The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: Methods of the EGAPP Working Group. *Genet Med* 11:3-14, 2009
35. Evaluation of Genomic Applications in Practice and Prevention (EGAPP): EGAPP Working Group recommendations. <http://www.egappreviews.org/recommendations/index.htm>
36. Wizemann T, Berger AC: Generating Evidence for Genomic Diagnostic Test Development: Roundtable on Translating Genomic-Based Research for Health Board on Health Sciences Policy. Washington, DC, Institute of Medicine, The National Academies Press, 2011
37. European Organisation for Research and Treatment of Cancer (EORTC): EORTC trial 10041 (BIG 3-04): MINDACT (Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy). http://www.eortc.be/services/unit/mindact/MINDACT_websiteii.asp
38. National Cancer Institute: The TAILORx breast cancer trial (Trial Assigning Individualized Options for Treatment [Rx]). <http://www.cancer.gov/clinicaltrials/noteworthy-trials/tailorx>
39. The Genomedical Connection. <http://www.genomedical.com/>
40. I-SPY 2 trial. <http://ispy2.org/>
41. National Cancer Institute: The Cancer Genome Atlas (TCGA). <http://cancergenome.nih.gov/>
42. Ramsey SD, Veenstra D, Tunis SR, et al: How comparative effectiveness research can help advance 'personalized medicine' in cancer treatment. *Health Aff (Millwood)* 30:2259-2268, 2011
43. Kuderer NM, Culakova E, Huang M, et al: Quality appraisal of clinical validation studies for multigene prediction assays of chemotherapy response in early-stage breast cancer. *J Clin Oncol* 29:214s, 2011 (suppl 15; abstr 3082)
44. Goddard KAB, Knaus WA, Whitlock E, et al: Decision-making in cancer genomic medicine using comparative effectiveness research. *Genet Med* 14:633-642, 2012
45. Etheredge LM: Creating a high-performance system for comparative effectiveness research. *Health Aff (Millwood)* 29:1761-1767, 2010
46. Abernethy AP, Etheredge LM, Ganz PA, et al: Rapid-learning system for cancer care. *J Clin Oncol* 28:4268-4274, 2010
47. Ginsburg GS, Staples J, Abernethy AP: Academic medical centers: Ripe for rapid-learning personalized health care. *Sci Transl Med* 3:101cm27, 2011
48. Lyman GH, Kuderer NM, Crawford J, et al: Predicting individual risk of neutropenic complications in patients receiving cancer chemotherapy. *Cancer* 117:1917-1927, 2011
49. Khorana AA, Kuderer NM, Culakova E, et al: Development and validation of a predictive model for chemotherapy-associated thrombosis. *Blood* 111:4902-4907, 2008
50. Subramanian J, Simon R: Gene expression-based prognostic signatures in lung cancer: Ready for clinical use? *J Natl Cancer Inst* 102:464-474, 2010
51. Subramanian J, Simon R: What should physicians look for in evaluating prognostic gene-expression signatures? *Nat Rev Clin Oncol* 7:327-334, 2010
52. Culakova E, Poniewierski MS, Huang M, et al: Assessment of genomic prognostic signatures as predictors of response to neoadjuvant chemotherapy in patients with early stage breast cancer. 34th San Antonio Breast Conference, San Antonio, TX, December 6-10, 2011 (abstr P3-14-04)
53. Lyman GH, Culakova E, Poniewierski MS, et al: Multigene signature assays in patients with early-stage breast cancer (ESBC) receiving neoadjuvant chemotherapy: An NCI-funded systematic review and evidence summary of predictive performance. 34th Annual San Antonio Breast Conference, San Antonio, TX, December 6-10, 2011 (abstr P5-13-17)
54. McShane LM, Altman DG, Sauerbrei W, et al: REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 93:387-391, 2005
55. Little J, Higgins JP, Ioannidis JP, et al: STrengthening the REporting of Genetic Association studies (STREGA): An extension of the STROBE Statement. *Ann Intern Med* 150:206-215, 2009
56. Dupuy A, Simon RM: Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J Natl Cancer Inst* 99:147-157, 2007
57. Compton C, Kelly A: Biospecimen banking in the post genome era, in Willard H, Ginsburg GS (eds): *Genomic and Personalized Medicine* (ed 2). Amsterdam, the Netherlands, Elsevier, 2012
58. Ginsburg GS, Burke TW, Febbo P: Centralized biorepositories for genetic and genomic research. *JAMA* 299:1359-1361, 2008