GENETIC TESTING AND MOLECULAR BIOMARKERS
Volume 21, Number 3, 2017

Pp. 184–194
DOI: 10.1089/gtmb.2016.0397

Which Results to Return: Subjective Judgments in Selecting Medically Actionable Genes

Gabriel Lázaro-Muñoz,[1,2] John M. Conley,[1,3] Arlene M. Davis,[1,4,5] Anya E.R. Prince,[1] and R. Jean Cadigan[1,4]

Background: Advances in genomics have led to calls for returning information about medically actionable genes (MAGs) to patients, research subjects, biobank participants, and through screening programs, the general adult population. Which MAGs are returned affects the harms and benefits of every genetic testing endeavor. Despite published recommendations of selection criteria for MAGs to return, scant data exist regarding how decision makers actually apply such criteria.
Methods: The process and criteria used by researchers when selecting MAGs for a preventive genomic sequencing program targeting the general adult population were examined. The authors observed and audio-recorded the gene selection meetings, and analyzed meeting transcripts, gene scoring sheets, and meeting handouts.
Results: To select MAGs, the committee imported, from a preexisting project, ''a semiquantitative metric'' that scores genes on five criteria. Numerous subjective judgments and conceptual challenges in defining and applying the five criteria complicated the selection process. Criteria-related challenges also included the limited evidence available about facts fundamental to the scoring decisions and the emergence and application of criteria that were not part of the original metric.
Conclusions: When identifying MAGs appropriate for screening and return, decision makers must expect and prepare to address such issues as the inevitability of subjective judgments, limited evidence about fundamental decision-making elements, the conceptual complexity of defining criteria, and the emergence of unplanned criteria during the gene selection process.

Keywords: medically actionable genes, genetic screening, return of individual results

Introduction

As genome-scale sequencing becomes cheaper and more commonplace, and as more genomic information is perceived to have medical utility, two related questions regarding the return of genetic results become critical: which genes should be analyzed for pathogenic variants, and how should those genes be chosen? The answers are already relevant in clinical practice and are gaining relevance in other settings: in public health, if population-based preventive genomic sequencing (PGS) becomes a reality, and in biobanking, as some biobanks return findings to their identified participants (Henderson *et al.*, 2013; Cadigan *et al.*, this volume). To inform emerging practice in these various settings, this case study addresses these two important questions in the context of PGS through examination of the gene selection process in a novel ongoing study called GeneScreen, which we describe in more detail below.

A key concept in answering these questions is medically actionable genes (MAGs): those genes that may contain pathogenic variants associated with a poor health outcome that can be mitigated by an available intervention. Approximately 0.5–1% of people are estimated to carry deleterious variants in MAGs (Evans *et al.*, 2013). Despite this small percentage, proponents of public health genomics argue that if screening for MAGs occurred opportunistically (e.g., whenever whole genome or exome sequencing [WGS/WES] is conducted for clinical or research purposes) or at the population level through PGS programs, millions of people who unknowingly carry medically actionable pathogenic

[1]Center for Genomics and Society, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
[2]Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, Texas.
[3]School of Law, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
[4]Department of Social Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.
[5]Center for Bioethics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

variants could be identified (Evans *et al.*, 2013; Green *et al.*, 2013; Berg *et al.*, 2016).

However, genetic experts disagree about which genes should be considered MAGs and returned to individual patients, research subjects, or biobank participants (Fullerton *et al.*, 2012; Berg *et al.*, 2013; Goddard *et al.*, 2013; McGuire *et al.*, 2013; PCSBI, 2013; van El *et al.*, 2013; Wolf *et al.*, 2013; Jarvik *et al.*, 2014; Klitzman *et al.*, 2014). Due to uncertainty in identifying pathogenic variants (Richards *et al.*, 2015; Van Driest *et al.*, 2016) and limited data about their penetrance in the general population (Green *et al.*, 2013; Prince *et al.*, 2014; Adams *et al.*, 2016), groups that return findings from MAGs risk unnecessary medicalization, especially for people without symptoms or family histories of genetic conditions, potentially leading to overtreatment, unwarranted anxiety, and other harms (Burke *et al.*, 2001, 2013; Khoury *et al.*, 2008; Klitzman *et al.*, 2013; Prince *et al.*, 2014; Lázaro-Muñoz *et al.*, 2015; Levine and Steinberg, 2015; Adams *et al.*, 2016). The prospect of harm makes it vitally important to base the choice of which genetic results to return on well-researched criteria and sound selection processes.

The first attempt to address which specific MAGs should be returned at the policy level occurred in 2013, when the American College of Medical Genetics and Genomics (ACMG) published its recommendations regarding MAGs to analyze and report to patients whenever WGS/WES is performed for clinical purposes (Green *et al.*, 2013). The recommendations came from an ACMG-appointed Working Group that utilized a deliberative expert consensus process to develop a list of 57 (later reduced to 56) genes that should be analyzed for actionable pathogenic variants, and the results returned. In November 2016, the list was increased to 59 genes (Kalia *et al.*, 2016).

The 2013 ACMG recommendations generated significant critiques (Burke *et al.*, 2013). These included complaints that the process for developing the guidelines was not sufficiently ''deliberative and inclusive'' (Ross *et al.*, 2013:525) and that ''[t]he broad criteria used [to select genes] would actually justify a much longer list of genes'' that can affect the replicability of the selection process (Wolf *et al.*, 2013:1050). Citing the 2011 National Academy of Medicine's (NAM) recommendation of eight standards to use to develop clinical practice guidelines (IOM, 2011), Ross *et al.* (2013) argued that the ACMG Working Group failed to meet most of the standards, including involving diverse stakeholders in decision making and eliminating bias due to conflicts of interest. Despite such critiques, the ACMG recommendations are used not only in clinical care but also in research settings. For example, Geisinger Health System aims to recruit over 100,000 of its members to participate in the ''MyCode Community Health Initiative.'' This precision medicine project and biobank will return results for 76 MAGs (associated with 27 health conditions), the bulk of which are the ACMG 56 (Geisinger, 2016).

The Geisinger MyCode project is by no means the only biobanking effort to return MAGs findings to individuals. One large-scale example is the Precision Medicine Initiative, which aspires to return individual and cohort-level results to its million or more participants (Collins and Varmus, 2015). Many scholars have written on the responsibilities of biobanks to return individual results to participants who indicate they want them, and the criteria to assess which results should be returned (e.g., Wolf *et al.*, 2012; Holm *et al.*, 2014). The criteria almost always hinge on medical actionability.

As genomic research advances, the number of genetic variants defined as medically actionable will increase. With this in mind, Berg *et al.* (2016) developed a ''semiquantitative metric,'' a scoring mechanism to use for assessing the medical actionability of gene-disease pairs. The metric responds to some of the critiques of the ACMG process in an effort to create a replicable tool that selection committees could employ to examine actionability in different contexts. The ''semi'' in Berg *et al.* (2016:472) semiquantitative metric was intended to suggest the ''nuances in application'' of the metric, specifically the ''subjectivity inherent in scoring.'' The metric was thus designed to promote rigor and transparency in gene selection, while at the same time allowing for some degree of flexibility in its application.

In this article, we critically examine a gene selection process to better understand the subjectivity that Berg *et al.* (2016) note. We perform this assessment through an ethnographic study of the deliberations of a committee of researchers (the Gene Selection Committee [GSC]) as they used Berg *et al.* (2016) semiquantitative metric to facilitate decision making in their analysis of candidate genes for a research project called GeneScreen. With support from the National Human Genome Research Institute of the National Institutes of Health (NIH), GeneScreen is developing one of the first PGS pilot programs targeting the general adult population. It aims to screen 1000 adults, return results of MAGs to participants, and analyze the ensuing harms and benefits of screening (Prince *et al.*, 2014; GeneScreen Project, 2016). GeneScreen recruits individuals for screening through two mechanisms: patients in general medicine clinics in North Carolina, and participants in a Kaiser Permanente biobank. GeneScreen researchers from the University of North Carolina (UNC) and Kaiser Permanente hope to obtain further funding to expand GeneScreen to include 20,000 additional Kaiser Permanente biobank participants.

Our real-time ethnographic study was designed into GeneScreen to enhance transparency and provide useful data to inform future gene selection efforts. The ethnographic study was intended to be inductive and open to issues as they emerged. The management of the subjectivity that Berg *et al.* (2016) foresaw—that is, the inevitably messy interaction between quantification and human judgment—immediately emerged as a critical issue, and it is the focus of this article. To our knowledge, this is the first published report of an observational study of an actual MAGs selection process.

Overview of GeneScreen's gene selection process

GeneScreen's GSC was composed of researchers whose fields include medical genetics, genetic counseling, bioethics, law, anthropology, philosophy, sociology, psychology, and nursing. A medical geneticist chaired the committee. Berg *et al.* (2016) originally developed the semiquantitative metric for use in a separate study called NCGENES (North Carolina Clinical Genomic Evaluation by NextGen Exome Sequencing) (Berg *et al.*, 2016 description of the metric reflects some slight revisions from the version used by GeneScreen). Several GSC members are also researchers in NCGENES.

NCGENES is part of the NIH-funded Clinical Sequencing Exploratory Research consortium (CSER, 2016). It examines

Table 1. Semiquantitative Metric: Official Criteria

Criterion	*Guiding question*	*Scores*
Severity	What is the nature of the threat to health for an individual carrying a deleterious allele in this gene?	3 = Death with no chance to intervene 2 = Diagnosis with death as distinctly possible outcome 1 = Significant morbidity 0 = Modest or no morbidity
Likelihood	What is the chance that a serious threat will materialize?	3 = >50% 2 = 25–49% 1 = 5–24% 0 = <5%
Effectiveness	How effective are interventions for preventing the harm?	3 = Highly effective 2 = Modestly effective 1 = Minimally effective 0 = Not effective
Acceptability	How acceptable are the interventions in terms of the burdens or risks placed on the individual?	3 = Highly acceptable 2 = Modestly acceptable 1 = Minimally acceptable 0 = Not acceptable
Knowledge	What kind of evidence is available to score severity, likelihood, effectiveness, and acceptability?	At the beginning of the gene selection process: 2 = A great deal is known about the condition, the gene and the above parameters 1 = Adequate body of knowledge exists to make clinical decision and recommendations when necessary 0 = Insufficient knowledge of above parameters to make decisions about intervention During the gene selection process it was modified to: 3 = Substantial evidence 2 = Moderate evidence 1 = Minimal evidence 0 = Controversial or poor evidence

The information was obtained from one of the handouts provided to GSC members during GeneScreen's gene selection process. Given GeneScreen's connection with NCGENES, that handout has some similarities to table 1 in Berg *et al.* (2016).

the utility of WES for identifying underlying genetic causes of undiagnosed conditions in several patient populations (e.g., cancer and cardiac diseases) (Foreman *et al.*, 2012; NCGENES, 2016). Using the metric, NCGENES identified 168 MAGs to be opportunistically screened as secondary targets and returned to patient-participants, in addition to primary target genes associated with their symptoms.

GeneScreen's GSC evaluated candidate genes for genomic screening in the general adult population. The GSC started with the metric and scores developed in NCGENES. The metric scores genes based on five criteria: (1) severity of disease outcome, (2) likelihood of severe outcome, (3) effectiveness of intervention, (4) acceptability of the intervention, and (5) knowledge base about the first four criteria (Berg *et al.*, 2016) (Table 1). Each criterion was scored from 0 to 3 (as we will discuss, knowledge base was initially scored from 0 to 2, but later changed) and scores were added to obtain a total score, with each candidate gene having a maximum possible score of 15. In theory, the higher the gene's total score the more suitable it would be for general population screening and return. GeneScreen's nine-member community advisory board (CAB) was also consulted about some GSC decisions.

Methods

This case study of GeneScreen's GSC meetings took place at UNC-Chapel Hill. UNC's Institutional Review Board evaluated the protocol and determined it to be exempt from review. All authors are GeneScreen researchers and three were GSC members. At least one of the anthropologist authors attended each of the seven 1-h long GSC meetings as a participant-observer. The meetings were recorded and transcribed. The results presented here are based on the authors' analyses of the GSC meeting transcripts, gene scoring sheets and other meeting handouts, and the authors' notes of the meetings. The authors reviewed the meeting transcripts and other documents, discussed the structure and dynamics of the decision-making process, identified emerging topics such as occurrences of subjective judgments, developed a coding scheme related to the criteria used to select genes for the screening program, and coded the transcripts using ATLAS.ti. The authors then reviewed the ATLAS.ti-generated transcript excerpts. These excerpts were divided among the authors who analyzed how the criteria were applied and drafted reports for each code. The authors reached a consensus regarding each code report, revised the drafts, and approved the final version of each report.

Results

Results below are divided into three sections. First, we define subjective judgments and illustrate how these judgments affected important aspects of the gene selection process. Next, we detail the complexity of applying each of the metric's ''official'' criteria and some of the subjective judgments that influenced gene scoring. Finally, we describe ''unofficial'' criteria that emerged during the selection process that proved important in selecting genes to screen in this context.

Subjective judgments

Subjective judgments are determinations where reasonable individuals could reach differing conclusions upon evaluation of the available information, or ''judgments that may be contested'' (Miller *et al.*, 2014). The 2011 NAM report for developing trustworthy clinical practice guidelines warns that subjective judgments often pass unnoticed but ''importantly influence conclusions'' (IOM, 2011:122). As the creators of the metric have acknowledged (Berg *et al.*, 2016), the metric's application to a specific context inevitably involves subjective judgments that cannot be viewed as directly compelled by evidence. Subjective judgments are more common when there is a lack of evidence or conflicting evidence on the issue in question (IOM, 2011). Given the lack of research and evidence about many aspects of population-based PGS programs and some other genomic testing endeavors (Prince *et al.*, 2014; Manrai *et al.*, 2016), it is not surprising that subjective judgments were common during the GSC gene selection process.

An illustration of the nature and importance of subjective judgments was observed early in the GSC's deliberations. It involved the fact that GeneScreen would be inviting people from the general population to be screened for genetic conditions. Some GSC members made the judgment that such a program should have stricter gene inclusion criteria than one that opportunistically screens for variants in genes for which the sequencing data have already been generated for research or clinical purposes. As a GSC member commented:

> In NCGENES we're saying ''We've done [WES] sequencing anyway. We…got the results. Now what do we do with it? And here are the ones we feel we should return.'' Very different than [GeneScreen where we are] saying ''We're gonna go out into the population, and we're gonna look to see if you have mutations in these genes.''

The argument against this distinction is that—like individuals from the general population—those who are opportunistically screened for pathogenic variants in secondary target MAGs generally lack symptoms or a family history of the secondary conditions examined. It would thus be reasonable to conclude that both groups are exposed to similar harms and benefits related to unsuspected pathogenic variants in MAGs, and the fact that raw sequencing data have been generated in one context but not the other does not justify a different threshold. Furthermore, there is no research available to suggest that the harms and benefits of screening for unsuspected genetic conditions in these two groups differ in meaningful ways. Therefore, the higher inclusion threshold for general population screening was a subjective judgment that had a significant impact on the number of genes that were even considered for GeneScreen. GeneScreen only considered 41 of the 168 MAGs from NCGENES because the GSC established a preset score threshold of ≥11.

Despite the semiquantitative metric's apparent simplicity, its application proved complex, and the complexities often involved subjective judgments. Moreover, it also became evident that subjective judgments related to factors outside the metric criteria played a significant role in the decision-making process (see Unofficial Criteria section). Recognizing this reality, one member cautioned the GSC: ''The metric isn't everything. The metric guides us but is not definitive. We can't just blindly turn the crank and say we should screen for these.''

Official criteria

In this section, we present data on how the metric's five ''official'' criteria were applied, detailing the challenges that emerged during the GSC's deliberations and highlighting some of the subjective judgments that played important roles in shaping gene scoring and selection (Tables 1 and 2 for GeneScreen's final list of genes and their scores).

Severity. Severity was a critical criterion, and one fraught with subjective judgments. The GSC decided that the more severe the potential health outcome associated with a gene, the greater the justification for inclusion. In fact, one member suggested that any gene associated with conditions that scored on the lower end (0 or 1) of the severity scale should be automatically excluded. While no ''auto-fails'' were integrated into the scoring system, none of the candidate genes had a severity score of 0 and very few scored a 1. Death was regarded as the worst possible outcome, but the only genes that received the highest score of 3 were those associated with conditions for which, as a GSC member described, ''you could drop dead all of a sudden'' (e.g., aortic dissection and cardiac arrhythmia). Genes associated with conditions for which death is a longer-term possibility (e.g., colorectal cancer and familial hypercholesterolemia) received a 2.

Genes associated with multiple, distinct severe health outcomes were particularly complicated to score for severity. When examining these genes, the GSC considered three alternatives: score (1) the most severe outcome, (2) the most penetrant outcome, or (3) multiple outcomes. The usual practice was to score ''the worst'' outcome, as a member commented. For example, for *FBN1*—associated with Marfan syndrome—aortic dissection was scored instead of any of the ocular or skeletal phenotypes also associated with the condition (Dietz, 2014). However, another member countered, ''We're not consistent…sometimes it's the worst, and sometimes it's the more likely. Because for *BRCA* we did breast cancer. I would argue ovarian cancer is worse.'' Similarly, for Lynch syndrome-associated genes, colorectal cancer—the most likely health outcome—was selected over ovarian or gastric cancer (Kohlmann and Gruber, 2014), which some might argue are more severe and difficult to manage. Choosing which outcome to score was a subjective judgment with critical implications for the decision-making process because it largely determined the scores for the other four metric criteria. Furthermore, there was no clearly stated reason as to why the GSC usually scored the most severe condition associated with a gene but in some cases scored the most penetrant, despite internal discussion of this discrepancy.

Table 2. Scoring Sheet for the 17 Genes Selected for GeneScreen's Adult Preventive Genomic Sequencing Pilot Program

		Disease outcome			*Clinical outcome*				
Gene	*Disease*	*Outcome scored*	*Severity*	*Likelihood*	*Intervention scored*	*Effectiveness*	*Acceptability*	*Knowledge*	*Total score*
APC	FAP	Colorectal cancer	2	3	Colonoscopy	3	2	3	13
LDLR	Familial hypercholesterolemia	Hypercholesterolemia	2	3	Statins	2	3	3	12, 13
BRCA1, *BRCA2*	Hereditary breast and ovarian cancer	Breast cancer	2	3	Prophylactic mastectomy	3	1	3	12
FBN1	Marfan syndrome	Aortic dissection	3	2	Annual echocardiogram	2	3	3	13
HFE	Hereditary hemochromatosis	Multiple system iron overload	2	1	Yearly ferritin check	3	3	3	12
KCNQ1, KCNH2, SCN5A	Romano-Ward long QT syndrome	Sudden death due to arrhythmia	3	2	EKG screening	2	3	3	13
MLH1, *MSH2*, *MSH6*, *PMS2*	Lynch syndrome	Colorectal cancer	2	3, 3, 2, 2	Colonoscopy	3	2	3	13, 13, 12, 12
MUTYH	Attenuated FAP, MUTYH-associated polyposis	Colorectal cancer	2	3	Colonoscopy	3	2	3	13
RET	MEN, FMTC	Medullary thyroid cancer	2	3	Thyroidectomy	3	2	3	13
RYR1	Malignant hyperthermia susceptibility	Anesthesia-induced malignant hyperthermia	2	1	Avoid certain anesthetics	3	3	3	12
SERPINA1	A1AT deficiency	Pulmonary disease	2	3	Avoid smoking	1	3	2	11

This table is a modified version of the scoring sheets provided to GSC members during GeneScreen's gene selection meetings. The outcomes and interventions shown were the ones used to calculate the final scores for GeneScreen.

A1AT, Alpha-1-antitrypsin; EKG, electrocardiogram; FAP, familial adenomatous polyposis; FMTC, familial medullary thyroid cancer; ICD, implantable cardioverter defibrillator; MEN, multiple endocrine neoplasia.

Likelihood. The likelihood of developing a severe outcome influenced the selection process, but did not generate as many subjective judgments as the other metric criteria. This may be due to the relative clarity of the definition—synonymous with penetrance—and the fact that, unlike the other criteria, likelihood is objectively quantifiable. As one member said, ''penetrance…is what it is.'' The GSC did acknowledge that current calculations likely overestimate penetrance in the general population, yet decided to use these estimates nonetheless. Referencing the penetrance of Romano-Ward Long QT syndrome-associated variants, a member commented:

> It's probably not as high as we think because we've always ascertained these people by looking at all the people who have had lethal dysrhythmias… We haven't been looking at the people who have mutations who didn't come to the doctor cause they didn't have lethal dysrhythmias.

Initially, the general consensus was that likelihood ''should not loom super large'' in the selection process. However, later meetings highlighted concerns that it is important to consider penetrance to minimize the risk of overdiagnosis and overtreatment in the general population. Ultimately, the vast majority of genes selected were associated with highly penetrant conditions (scores of 2 or 3) and only two of the genes selected, *HFE* and *RYR1*, scored a 1.

Effectiveness. Effectiveness of intervention was a complex assessment in which subjective judgments—in the sense of judgments that seemed to be based on personal preferences that could be contested—were influential. On the one hand, a gene that might score high in other categories could score poorly in effectiveness and be disregarded. A member warned:

> I do not want to go out looking for stuff in people. Say [to them], ''You've got a disease that could be severe'' and say, ''By the way, we really don't have very good ways of dealing with it.''

On the other hand, a gene that scored poorly in other categories could still make the list if an effective and acceptable intervention was available. This occurred in the case of *HFE*, a gene associated with hemochromatosis:

> The penetrance isn't high…. it should get dinged as it does on our metric with that, but it is more than compensated for by the fact that you've got this really acceptable, absolutely effective intervention…. which is phlebotomy.

These examples suggest that—although the metric did not differentially weight criteria—in practice, a criterion like effectiveness, at times was given more weight by decision makers than criteria such as severity and likelihood.

Scoring conditions with multiple possible effective interventions was particularly difficult. Some GSC clinicians argued for scoring the intervention they personally would initially recommend, which often involved a subjective judgment that balanced the effectiveness and burden of the possible interventions. For example, double mastectomy was the intervention scored for *BRCA1/2* mutations instead of increased surveillance, which the GSC considered less burdensome. A clinician stated, ''What I tell [patients] is, 'if you choose surveillance, you're doing nothing to prevent. The only way to prevent breast cancer is bilateral mastectomy.'''

For other genes, a less effective but more acceptable intervention was chosen. For example, when discussing interventions for Romano-Ward Long QT syndrome, a member commented:

> [Beta-blockers are] moderately effective…certainly not a guaranteed success…. The ICD [implantable cardioverter defibrillator] actually is highly effective…. If you score ICD, you'd get a higher effectiveness but a lower acceptability.

Ultimately, the GSC scored electrocardiogram (EKG) screening as the recommended intervention. Like beta-blockers, EKG screening was considered moderately effective, but highly acceptable.

Some clinician members also expressed concern about the efficacy of behavioral interventions. This revealed a bias among clinicians toward endorsing genes associated with conditions for which there are ''medical'' interventions, and generated sharp disagreements along disciplinary lines. A medical geneticist noted:

> [B]ehavioral interventions…. They're weak…. They're hard to implement. They nudge risk. Right. And I think that combined with the weakness of most genetic predictors is what is responsible for the failure of public health genomics so far.

A social psychologist countered that behavioral interventions might only seem weak because their effectiveness is often assessed before their impact can be measured. This dispute was particularly sharp in the case of *SERPINA1*, whose related condition (alpha-1 antitrypsin deficiency) is managed primarily by smoking avoidance; the gene was ultimately included.

Acceptability. Acceptability of the recommended intervention was an important but challenging consideration. A GSC member commented:

> We could go searching for *CDH1* mutations that cause a high risk of gastric cancer…. We have [a] very good intervention…Here's the problem…the only intervention is removing your stomach which is not only a big surgery,…it's a tough way to live…. To ignore acceptability would mean that it would score very highly, and I actually would feel real troubled by us going out into the general population and screening for something where the only intervention is that.

Acceptability was acknowledged by the GSC to be ''the most subjective'' of all criteria, and its scoring was described as ''squishy.'' This was in part due to the difficulty of defining acceptability and the lack of relevant research in the context of population-based genomic screening. Scoring acceptability involved numerous subjective judgments. For example, to facilitate decision making, acceptability was determined by comparing interventions associated with different genetic conditions under consideration. As a GSC member commented, ''It's not just 'Where would you put colonoscopy?' it's 'How does colonoscopy rate relative to having your stomach out relative to an annual blood test?'''

Furthermore, when evaluating the acceptability of an intervention, the GSC decided to downplay the impact of contextual features such as participant characteristics, cost of interventions, severity, and likelihood of a severe outcome. Some GSC members questioned this decision, and suggested that a burdensome intervention may become more acceptable if the likelihood of developing a severe outcome is high. However, the approach followed by the GSC was typified by this comment:

> The slippery slope is [that] if it means I'm gonna die, everything's acceptable…. To me the way to look at acceptability is you pluck somebody out of the population, you don't tell them the context, and…you do a colonoscopy…. How burdensome is that versus…you draw their blood?''

Some GSC members were concerned about whose perspective should be used to score acceptability: that of clinicians, researchers and academics, or the general public? A clinician member suggested that the CAB could help score acceptability because ''[W]e have tried very hard to ignore the fact that we're clinicians when we come up with the acceptability scores…. [But] we have no idea if we're right compared to the lay person.'' As it turned out, at a subsequent CAB meeting, CAB members had little consensus regarding acceptability scoring. Unlike the approach ultimately followed by the researchers, CAB members were reluctant to score acceptability without considering multiple contextual features such as those mentioned above. The CAB's nuanced view of acceptability resisted the orderly metric scoring and further highlights the subjective nature of this criterion.

Knowledge base. The metric's knowledge criterion is intended to ensure that claims about the other criteria are scientifically well-founded. However, knowledge base proved to be a moving target. Early in GeneScreen's selection process the GSC interrogated the knowledge criterion imported from NCGENES. An initial problem was that NCGENES scored knowledge on a three-point scale (0, 1, 2) that some GSC members believed did not provide sufficient gradation or guidance (Table 1). One GSC member characterized it as ''knowledge good or knowledge bad.'' The same member highlighted a key underlying problem with defining knowledge base:

> [T]here is a difference between relying on what people in my field say and my clinical experience. ''This is how confident I am at doing this.''…Which is how some people define knowledge base. And other people define it as, ''Can I find a journal article in a peer reviewed journal that looked at people who did different things and got different outcomes?''

GSC members believed it was particularly important to be rigorous when assessing knowledge base in the context of general population screening. As one member advised, ''The last thing…we should do is be running out and looking for problems in people when we really don't even understand the disease, the interventions, the process.'' Another member proposed:

> I'm thinking of a three-point [in fact, four-point—0, 1, 2, 3] knowledge base scale.… [T]hree would be things that have primary literature about the condition itself. Two would be things where there's a biological similarity to a different disease that has good literature…like *PTEN* and breast cancer. [O]ne is…where we're really hand waving and saying 'Well, expert opinion says, but nobody has studied.' Then zero is [where there is controversial or poor evidence].

The GSC agreed to revise the criterion as suggested, adopting an expanded scale with each score defined more rigorously. This likely helped minimize subjectivity, but at some points clinician members used their own clinical experience to resolve ambiguities in applying the revised scale.

Unofficial criteria

Other decision-making elements beyond the official metric criteria influenced the gene selection process. We labeled these informal but powerful factors the ''unofficial criteria'' (Table 3). These unofficial criteria—population prevalence, cost of sequencing the gene, and age of onset of the associated

Table 3. Unofficial Criteria That Influenced Gene Selection

Criterion	*Rationale*	*Illustrative quotes*
Population prevalence of pathogenic variants in the gene	Because GeneScreen will only enroll 1000 participants, screening genes with low prevalence of pathogenic variants will not advance the research aim of understanding the implications of returning a positive result.	''Something that has a mutation prevalence of one in a million there's not much point in looking for it in our sample of a thousand people.''
		''There are two syndromes on this list that have a high enough population prevalence where we're going to find them: Lynch Syndrome and *BRCA1/2*. Cumulatively they have about a one in two hundred population prevalence…. That's where the money is for us.''
Cost of sequencing the gene	GeneScreen has a set budget and sequencing large genes increases costs. Also, the prevalence of pathogenic variants in some genes is low enough that they would increase costs, but are not likely to yield positive results in the study's sample.	''We are constrained by costs…. we will probably have to factor in how big is the gene.''
		''[*APC* is] a big gene, and it's going to cost a lot of money [to sequence].''
Age of onset	GeneScreen targets the general adult population. For conditions that express before adulthood, participants would likely already be impacted by the condition, which reduces the utility of genomic screening for that condition in adults.	''The idea of doing little kids and looking at their *APC* gene that causes familial polyposis coli with essentially a 100% cancer rate [and is] highly intervenable…that would make a lot of sense to me. But we're doing adults, and if the average age of our participants if forty, the chance that they're going to have a previously undetected *APC* mutation is really low.''

condition—were largely motivated by GeneScreen's specific context, research questions, and budgetary limits, but their emergence also highlighted the pervasive influence of subjective judgments in the gene selection process.

Population prevalence. The prevalence of pathogenic variants in MAGs was an influential consideration. Prevalence of pathogenic variants is, of course, quantifiable, and is thus objective in one sense. But the decision to consider prevalence even though it was not part of the official criteria and, ultimately, to give it great weight were subjective judgments. A GSC member argued, ''Something that has a mutation prevalence of one in a million there's not much point in looking for it in our sample of a thousand people,'' and later added, ''we should have a tight, high-yield list.'' The same member also commented: ''There are two syndromes on this list that have a high enough population prevalence where we're gonna find them. It's Lynch Syndrome and *BRCA1/2*. Cumulatively they have about a one in two hundred population prevalence…. That's where the money is for us.''

Prevalence became such a central consideration that it was added to the scoring sheet alongside the official criteria scores, a status not accorded the other unofficial criteria. Discussions about prevalence prompted one member to ask, ''So is frequency a new category?'' Another replied, ''No, but it's a consideration.'' Concerns about low prevalence were key for the exclusion of some genes that scored high on the metric (i.e., ≥11), including *VHL*, *PTEN*, *TGFBR 1*, and *TGFBR 2*. In contrast, high prevalence favored inclusion of genes ''dinged'' on other grounds, such as *SERPINA1* with its ''weak'' behavioral intervention and *HFE* with its low penetrance. Prevalence thus led the GSC to reject some higher scoring genes from the final list and to include some with low scores on certain criteria.

Cost of sequencing the gene. Early in the selection process a GSC member argued that the list should be short because ''something meeting our rigorous criteria will be somewhat unusual, and secondly we are constrained by costs…. we will probably have to factor in how big is the gene.'' When discussing the possible inclusion of *APC*, the same member commented, ''It's a big gene, and it's gonna cost a lot of money [to sequence].'' *APC* was eventually kept on the list, but other genes such as *TTN* were excluded because of their size. A member commented, ''Well, why not screen for *TTN*…? Well, it's gigantic,'' later adding, ''it's cost.''

The cost of sequencing was also often cited as a reason for excluding high-scoring genes with low prevalence of pathogenic variants. A member argued that genes associated with pathogenic variants that occur in one in a million individuals should not be included ''even if [they] meet our other criteria, [because] it costs money to do this.'' Including these genes would increase costs but would not advance GeneScreen's research goal of studying the impact of disclosing positive findings to participants from the general population. Cost thus exerted influence but specific costs were never estimated or discussed in relation to the budget.

Age of onset. GeneScreen was designed to target the general adult population. Therefore, some GSC members were concerned that including genes associated with conditions that present before adulthood would undermine the utility of the screening. One member argued, ''The whole point of screening is to find something early so you can prevent a problem. If people are gonna have the problems [by the time they participate in the program], then why would you waste time screening?'' One member argued against considering age-based risk because of concerns about the lack of data on age of onset for many genetic conditions. Others felt uncomfortable about excluding conditions that generally have an early age of onset because some individuals could nonetheless begin experiencing symptoms later in life. Age of onset issues were considered when evaluating the utility of screening for pathogenic variants in genes such as *APC*, *RET*, *FBN1*, *BRCA 1*, and *BRCA 2*. For example, when evaluating *APC* a clinician commented, ''we're [screening] adults…if the average age of our participants is forty, the chance that they're gonna have a previously undetected APC mutation is really low.'' The discussion of this factor was episodic and the GSC set no general standards.

Discussion

In this study of one of the first attempts at selecting genes for an adult PGS program (a program that recruits participants through clinic populations and through a healthcare organization's biobank), we have focused on the subjective judgments that arise when a group attempts to select MAGs to examine and results to return. This analysis is useful not only for those interested in implementing PGS programs, but also for biobankers and other researchers who are under increasing pressure to return medically actionable individual results to participants (Wolf *et al.*, 2012; Jarvik *et al.*, 2014). Even for biobankers and researchers focused on populations with specific diseases, this analysis of scores relevant to screening asymptomatic adults remains relevant given that any secondary findings that are returned may be unrelated to the disease and not indicated by any symptoms or family history. Berg *et al.* (2016:473) acknowledge ''the necessarily subjective nature of any assessment of actionability,'' and recommend that ''the scores and evidence base generated by the application of the semiquantitative metric are best considered an initial starting point for more nuanced discussions about individual conditions or particular clinical applications.'' In this article, we explored the subjectivity in the application process and the discussions and decision making that occurred when the metric was used to select genes that would be examined and returned in a genomic testing program. Subjective judgments in the application of the metric are problematic because they add arbitrariness to the process and threaten its reproducibility. A rigorous and well-informed gene selection process is an essential first step for any genomic testing endeavor because it can help properly manage subjective judgments, and maximize the benefits and minimize harms of testing.

In future gene selection processes, some of the specific challenges that introduced subjectivity into the GSC process could be more effectively managed through advance planning. Below we discuss some key lessons revealed by our analysis that may be particularly helpful when using a metric to identify MAGs to return to individuals.

Managing subjectivity with diverse stakeholders. The 2011 NAM report recommends that groups consist of individuals from clinical and nonclinical disciplines with representation from more than one clinical specialty and from the general public, including patients and caregivers (IOM, 2011:83–86). NAM bolsters its recommendation for clinician diversity by citing studies showing that clinicians have a ''lower threshold'' for recommending procedures performed within their own specialty (IOM, 2011:85).

The GSC did seek the views of the general public by asking the CAB for assistance with scoring the acceptability of the interventions, and the CAB's response illustrated the potential for varying perspectives about how to evaluate this criterion. CAB members felt they could not score acceptability without considering contextual features, an idea that the GSC had already rejected. Ultimately, it was not clear how, if at all, the CAB's perspectives were integrated into the GSC's decision-making process. One particular problem was that CAB members were not invited to participate in the GSC's deliberative meetings; instead, the CAB met on its own, with some GSC members observing and reporting the CAB's views back to the full GSC.

Furthermore, although a wide variety of nonclinical fields were represented in the GSC, an analysis of speaking turns showed that nonclinician members of the GSC were much less likely to speak during the GSC meetings compared to clinicians. Most nonclinicians seemed to lack sufficient knowledge about the specific genes to contribute substantially—or even to attempt to contribute—to conversations about most of the criteria on the metric. In addition, as stated above, there were no CAB members in the gene selection deliberative meetings. Together, this meant that genetic clinicians had a disproportionate influence on the MAG selection process.

The inability to incorporate the CAB's views and the disproportionate influence of clinicians raises an important general question: what role should nonclinicians play in the selection process? The category of nonclinicians can include community members, academics, policy makers, and others. The participation of diverse stakeholders in the selection process can promote a balanced representation of interests affected by genomic testing programs and encourage questioning from different perspectives, which can help identify and address subjective judgments and shared assumptions.

Although the GSC membership was highly interdisciplinary, it lacked clinical diversity (all of the GSC clinicians specialized in genetics), excluded community members from deliberative meetings, and nonclinician academics were much less likely to speak than clinicians. Furthermore, simply adding input from diverse stakeholders into a gene selection process does not mean their perspectives will be effectively integrated. Stakeholders have different perspectives and levels of understanding about genomics, medical care, and the ethical, legal, and social implications of genomic testing programs, which can complicate communication and make it difficult to integrate diverse perspectives. Future initiatives should not only add diverse stakeholders, but carefully consider how they can contribute to decision making, and what training may be necessary to help all stakeholders contribute effectively.

Evaluating genes associated with multiple health outcomes. Candidate genes are often associated with more than one severe health outcome. While individuals identified with a pathogenic variant in these genes will likely be informed that they are at an increased risk of developing all of these outcomes (Kocarnik and Fullerton, 2014), the variant will have different penetrance levels for each potential health outcome and some of these outcomes will not be medically actionable. Therefore, a gene may meet metric criteria based on its association with one medically actionable outcome, but it may not meet the criteria for inclusion if other outcomes are considered.

Although some GSC members argued for scoring multiple outcomes, GeneScreen's GSC ultimately decided to score just one outcome per gene regardless of pleiotropy (e.g., *BRCA 1*, *BRCA 2*; Table 2). This subjective judgment likely was influenced by an interest in maintaining a user-friendly metric and the constraints of time, given the need to move the larger research project forward. However, in-depth examination of how to manage MAGs associated with multiple health outcomes, and the possibility of integrating multiple outcomes into calculations, should be considered by future initiatives to more accurately assess the risks and benefits of analyzing and reporting this type of MAGs. Furthermore, while the GSC generally scored the most severe condition or outcome, in some cases it scored the most penetrant, and never developed an explicit policy about which of these to score. Nor was it clear what drove this choice in individual cases. Sometimes it seemed to be an ethical judgment based on a hypothetical participant's best interest, whereas in other instances the choice seemed to reflect an unstated clinical preference. This type of inconsistency should be acknowledged and addressed to minimize arbitrariness. Future initiatives should carefully consider how they will manage such decisions before they begin their gene selection process because of the potential impact on the ultimate list of genes and the ensuing harms and benefits to those screened.

Availability of multiple interventions. A similar problem was revealed when the GSC evaluated genes associated with genetic conditions that have multiple possible interventions. For outcomes that are medically actionable, interventions may have different levels of effectiveness and burden. GSC members made a subjective judgment when they decided that for gene selection purposes they would score the intervention that clinician members believed struck the best balance between effectiveness and burden. However, others could decide to score the intervention that patients most commonly select, the most effective one, the one generally recommended by professional organizations, or even multiple interventions. Different choices could lead to different assessments about the medical actionability of a genetic condition, so future groups need to carefully examine the respective harms and benefits of each option.

Considering acceptability as a criterion. Some of the dilemmas discussed in the preceding sections involve evaluation of the acceptability of an intervention, which underscores the problematic nature of defining and applying an acceptability criterion. Furthermore, acceptability rests on a subjective judgment motivated by the medical ethics principle of nonmaleficence. That is, the goal is to avoid harm by not screening for and returning genomic health risks for which the only available intervention would be viewed, at least by some, as too onerous. However, one could argue that

this is an unduly paternalistic approach and that individuals carrying actionable genomic risks should have the opportunity to decide for themselves whether the intervention is too onerous. If future initiatives still decide to use acceptability as a criterion, it will be important to acknowledge its inherent subjective nature and manage it by carefully defining what is acceptable and why, and whose perspectives should count in scoring it. In addition, future initiatives could also try to manage the subjective nature of criteria such as acceptability and severity by examining the perspective of members of the general population and professional clinical organizations about the interventions and conditions under consideration.

Managing unofficial criteria. In addition to the official criteria, the population prevalence of pathogenic variants in each MAG, the cost of sequencing specific MAGs, and the age of onset of symptoms of associated genetic conditions also influenced the GSC's decisions. Consideration of these unofficial criteria does not necessarily suggest that the formal gene selection process was flawed. Unofficial criteria provided a way to address important concerns specific to the context of GeneScreen that were voiced only after the process was underway. However, to minimize the impact of subjective judgments, such as the arbitrary use of decisional elements outside the official metric criteria, future selection processes should consider at the outset whether the specified decision-making tools contain all of the criteria that are likely to be relevant for their context. If unofficial criteria emerge during the process and are not promptly identified and interrogated by decision makers, the process may be at risk of arbitrariness. Decision makers must be poised to identify such criteria, determine their utility, and then define and transparently implement them as appropriate.

Limitations

Given the nature of a case study, our analysis is based on the process and criteria employed by a single center. Furthermore, the analysis is based on GeneScreen's GSC meetings and does not include discussions that may have occurred outside of these meetings, nor deliberations about the metric that occurred in the predecessor NCGENES project. Many of GeneScreen's investigators were also part of the NCGENES process, and they may have brought to the GeneScreen selection process shared assumptions or important conclusions that are not captured in the data we considered. Thus, GeneScreen's GSC meetings and the related data provided a revealing but inevitably incomplete look into several of the complicated issues that arise in selecting MAGs for examination and return.

Conclusions

Our findings suggest that MAGs selection criteria such as those utilized by GeneScreen are important guides in the process, but those who use them ''can't just blindly turn the crank.'' Instead, decision makers should be reflective in developing and applying selection criteria, to ensure that those criteria address all the relevant aspects of informed decision making in the context at hand. At the same time, they should strive to make sure that inevitable subjective judgments and unplanned but influential criteria are acknowledged transparently and assessed and managed with reference to available evidence. Advances in genomic sequencing will increase calls for both preventive genomic screening programs for the general population, and the development of return of results policies for biobankers and other researchers. The selection of genes to examine and return as a part of such programs and policies requires close scrutiny to ensure that these are advisable in the first place and that, if implemented, they maximize their social and medical utility.

Acknowledgments

The authors would like to thank Eric T. Juengst, Debra Skinner, Gail E. Henderson, Jonathan S. Berg, and Michael Adams for helpful comments and discussions about topics discussed in this article, and members of NCGENES (NIH Grant 1U01HG006487-01 [Evans, PI]) who were instrumental in developing the semiquantitative metric that is integral to the analyses presented here. Research for this article was funded by the National Human Genome Research Institute of the National Institutes of Health (NIH) Grant 2P50HG004488 (Henderson, PI), ''Center for Genomics and Society,'' and K99HG008689 (Lázaro-Muñoz, PI). The views expressed are those of the authors alone, and do not necessarily reflect views of NIH or all CGS investigators.

Author Disclosure Statement

The authors (G.L.M., J.M.C., A.M.D., A.E.R.P., and R.J.C.) do not have competing financial interests or other conflicts of interest to report.

References

Adams MC, Evans JP, Henderson GE, *et al.* (2016) The promise and peril of genomic screening in the general population. Genet Med 18:593–599.

Berg JS, Amendola LM, Eng C, *et al.* (2013) Processes and preliminary outputs for identification of actionable genes as incidental findings in genomic sequence data in the Clinical Sequencing Exploratory Research Consortium. Genet Med 15:860–867.

Berg JS, Foreman AKM, O'Daniel JM, *et al.* (2016) A semiquantitative metric for evaluating clinical actionability of incidental or secondary findings from genome-scale sequencing. Genet Med 18:467–475.

Burke W, Antommaria AH, Bennet R, *et al.* (2013) Recommendations for returning genomic incidental findings? We need to talk! Genet Med 15:855–859.

Burke W, Coughlin SS, Lee NC, *et al.* (2001) Application of population screening principles to genetic screening for adult-onset conditions. Genet Test 5:201–211.

Cadigan RJ, Edwards TP, Lassiter D, *et al.* ''Forward thinking'' in U.S. biobanking. Genet Test Mol Bioma (In press).

Clinical Sequencing Exploratory Research (CSER) (2016) Moving the genome into the clinic. Available at https://cser-consortium.org, accessed September 1, 2016.

Collins FS, Varmus H (2015) A new initiative on precision medicine. NEJM 372:793–795.

Dietz HC (2014) Marfan syndrome. In: Pagon RA, Adam MP, Ardinger HH, *et al.* (eds) GeneReviews® [Internet]. University of Washington, Seattle, WA. Available at http://www.ncbi.nlm.nih.gov/books/NBK1335, accessed September 1, 2016.

Evans JP, Berg JS, Olshan AF, *et al.* (2013) We screen newborns, don't we? Realizing the promise of public health genomics. Genet Med 15:332–334.

Foreman AKM, Lee K, Evans JP (2012) The NCGENES project: exploring the new world of genome sequencing. NC Med J 74:500–504.

Fullerton SM, Wolf WA, Brothers KB, *et al.* (2012) Return of individual research results from genome-wide association studies: experience of the Electronic Medical Records and Genomics (eMERGE) network. Genet Med 14:424–431.

Geisinger (2016) MyCode® Community Health Initiative. Available at www.geisinger.org/for-researchers/partnering-with-patients/pages/mycode-health-initiative.html, accessed September 1, 2016.

GeneScreen Project (2016) Center Research Project. Center for Genomics and Society at the University of North Carolina at Chapel Hill, 2P50HG004488 (Henderson, PI). Available at http://genomics.unc.edu/genomicsandsociety/Research.html, accessed September 1, 2016.

Goddard KAB, Whitlock EP, Berg JS, *et al.* (2013) Description and pilot results from a novel method for evaluating return of incidental findings from next-generation sequencing technologies. Genet Med 15:721–728.

Green RC, Berg JS, Grody WW, *et al.* (2013) ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med 15:565–574.

Henderson GE, Edwards TP, Cadigan RJ, *et al.* (2013) Stewardship practices of US biobanks. Sci Transl Med 5:215cm7.

Holm IA, Savage SK, Green RC, *et al.* (2014) Guidelines for return of research results from pediatric genomic studies: deliberations of the Boston Children's Hospital Gene Partnership Informed Cohort Oversight Board. Genet Med 16: 547–552.

Institute of Medicine (IOM) (2011) Clinical Practice Guidelines We Can Trust. National Academies Press, Washington, DC.

Jarvik GP, Amendola LM, Berg JS, *et al.* (2014) Return of genomic results to research participants: the floor, the ceiling, and the choices in between. Am J Hum Genet 94:818–826.

Kalia SS, Adelman K, Bale SJ, *et al.* (2016) Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2. 0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med. Available at www.nature.com/gim/journal/vaop/ncurrent/abs/gim2016190a.html, accessed November 28, 2016.

Khoury MJ, Berg A, Coates R, *et al.* (2008) The evidence dilemma in genomic medicine. Health Aff 27:1600–1611.

Klitzman R, Appelbaum PS, Chung W (2013) Return of secondary genomic findings vs patient autonomy: implications for medical care. JAMA 301:369–370.

Klitzman R, Buquez B, Appelbaum PS, *et al.* (2014) Processes and factors involved in decisions regarding return of incidental genomic findings in research. Genet Med 16:311–317.

Kocarnik JM, Fullerton SM (2014) Returning pleiotropic results from genetic testing to patients and research participants. JAMA 311:795–796.

Kohlmann W, Gruber SB (2014) Lynch syndrome. In: Pagon RA, Adam MP, Ardinger HH, *et al.* (eds) GeneReviews® [Internet]. University of Washington, Seattle, WA. Available at www.ncbi.nlm.nih.gov/books/NBK1211, accessed September 1, 2016.

Lázaro-Muñoz G, Conley JM, Davis AM, *et al.* (2015) Looking for trouble: preventive genomic sequencing in the general population and the role of patient choice. AJOB 15:3–14.

Levine B, Steinberg K (2015) Proposed shift in screening for breast cancer. JAMA 313:525.

Manrai AK, Ioannidis JPA, Kohane IS (2016) Clinical genomics: from pathogenicity claims to quantitative risk estimates. JAMA 315:1233–1234.

McGuire AL, Joffe S, Koenig BA, *et al.* (2013) Point-counterpoint. Ethics and genomic incidental findings. Science 340:1047–1048.

Miller FG, Joffe S, Kesselheim AS (2014) Evidence, errors, and ethics. Perspect Biol Med 57:299–307.

NCGENES (2016) Welcome–North Carolina Clinical Evaluation by NextGen Exome Sequencing, 1U01HG006487-01 (Evans, PI). Available at www.med.unc.edu/ncgenes, accessed September 1, 2016.

Presidential Commission for the Study of Bioethical Issues (PCSBI) (2013) Anticipate and communicate: ethical management of incidental and secondary findings in the clinical, research, and direct-to-consumer contexts. Available at http://bioethics.gov/sites/default/files/FINALAnticipateCommunicate_PCSBI_0.pdf, accessed September 1, 2016.

Prince AER, Berg JS, Evans JP, *et al.* (2014) Genomic screening of the general adult population: key concepts for assessing net benefit with systematic evidence reviews. Genet Med 17:441–443.

Richards S, Aziz N, Bale S, *et al.* (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–423.

Ross LF, Rothstein MA, Clayton EW (2013) Premature guidance about whole-genome sequencing. Per Med 10:523–526.

Van Driest SL, Wells QS, Stallings S, *et al.* (2016) Association of arrhythmia-related genetic variants with phenotypes documented in electronic medical records. JAMA 315:47–57.

van El CG, Cornel MC, Borry P, *et al.* (2013) Whole-genome sequencing in health care: recommendations of the European Society of Human Genetics. Eur J Hum Genet 21:580–584.

Wolf SM, Annas GJ, Elias S (2013) Point-counterpoint. Patient autonomy and incidental findings in clinical genomics. Science 340:1049–1050.

Wolf SM, Crock BN, Van Ness B, *et al.* (2012) Managing incidental findings and research results in genomic research involving biobanks and archived data sets. Genet Med 14: 361–384.

Address correspondence to:
Gabriel Lázaro-Muñoz, PhD, JD, MBE
Center for Medical Ethics and Health Policy
Baylor College of Medicine
One Baylor Plaza, Suite 310D
Houston, TX 77030

E-mail: Gabriel.Lazaro-Munoz@bcm.edu