Published in final edited form as:

Genet Med. 2013 September ; 15(9): 721–728. doi:10.1038/gim.2013.37.

Description and Pilot Results from a Novel Method for Evaluating Return of Incidental Findings from Next Generation Sequencing Technologies

Katrina A.B. Goddard, PhD[1], **Evelyn P. Whitlock, MD, MPH**[1], **Jonathan S. Berg, MD, PhD**[2], **Marc S. Williams, MD**[3], **Elizabeth M. Webber, MS**[1], **Jennifer A. Webster, MS**[1], **Jennifer S. Lin, MD, MCR**[1], **Kasmintan A. Schrader, MBBS**[4], **Doug Campos-Outcalt, MD, MPA**[5], **Kenneth Offit, MD, MPH**[4], **Heather Spencer Feigelson, PhD**[6], and **Celine Hollombe, MPH**[1]

[1]Center for Health Research, Kaiser Permanente Northwest, Portland, OR

[2]Department of Genetics, University of North Carolina, School of Medicine, Chapel Hill, NC

[3]Genomic Medicine Institute, Geisinger Health System, Danville, PA

[4]Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY

[5]Department of Family, Community, and Preventive Medicine, University of Arizona College of Medicine, Phoenix, AZ

[6]Institute for Health Research, Kaiser Permanente Colorado, Denver, CO

Abstract

Purpose—To develop, operationalize, and pilot test a transparent, reproducible, and evidence informed method to qualify when to report incidental findings from next generation sequencing technologies.

Methods—Using evidence-based principles, we propose a three stage process. Stage I ‘rules out’ incidental findings below a minimal threshold of evidence and is evaluated using inter-rater agreement and comparison with an expert-based approach. Stage II documents criteria for clinical actionability using a standardized approach to allow experts to consistently consider and recommend whether results should be routinely reported (Stage III). We used expert opinion to determine the face validity of Stages II and III using three case studies. We evaluated the time and effort for Stages I and II.

Corresponding author’s contact information: Center for Health Research, Kaiser Permanente Northwest, 3800 North Interstate Avenue, Portland, Oregon 97227, Telephone: (503) 335-6353, Fax: (503) 335-2400, Katrina.AB.Goddard@kpchr.org.

Potential Financial Conflicts of Interest: None disclosed

Supplementary information is available at the *Genetics in Medicine* website.

Author Contributions:
Conceived and designed the experiments: KABG JSB JSL KO KAS EMW JAW MSW EPW.
Development of methodology: KABG JSB DCO HSF JSL KAS EMW JAW MSW EPW.
Acquisition of data: KABG JSB DCO HSF KAS JAW EMW.
Analysis and interpretation of results: KABG JSB DCO HSF EMW JAW MSW EPW.
Wrote, reviewed and revised the manuscript: KABG JSB DCO HSF JSL KAS EMW JAW MSW EPW.
Provided administrative, technical, or material support: KABG KO EPW CH.

Results—For Stage I, we assessed 99 conditions and found high inter-rater agreement (89%), and strong agreement with a separate expert-based method. Case studies for familial adenomatous polyposis, hereditary hemochromatosis, and α1-Antitrypsin Deficiency were all recommended for routine reporting as incidental findings. The method requires less than three days per topic.

Conclusion—We establish an operational definition of clinically actionable incidental findings and provide documentation and pilot testing of a feasible method that is scalable to the whole genome.

Keywords

whole genome sequencing; clinical actionability; population screening; secondary findings; whole exome sequencing

INTRODUCTION

Use of next generation sequencing (NGS) technologies in clinical medicine and research is increasing, with tens of thousands of human genomes now completed.[1] The decreasing cost of generating whole genome or whole exome sequence (WGS, WES) data using NGS platforms means that it may soon be less expensive to routinely sequence the entire genome than to use many more targeted assays.[2] Thus, we are rapidly approaching a time when genome-scale sequencing tests may be analytically robust, clinically relevant and affordable enough to replace the single gene tests currently used in clinical practice.

These new technologies allow for a vast proportion of an individual's DNA to be queried, raising the issue of genomic incidental findings (IFs—unanticipated information discovered in the course of testing or medical care). [3] IFs will inevitably arise when NGS is not restricted to portions of the genome that are relevant to the indication for the test. Individual genomes that have been evaluated using WGS indicate that there are typically over 4 million variants observed per person compared with the reference sequence.[4–9] Each person carries an average of 50 to 100 heterozygous variants classified by the Human Gene Mutation Database as causing inherited disorders, although emerging information suggests that most of these are incorrect assertions of pathogenicity and/or errors of curation.[10,11] Without thoughtful interpretation, it would be a daunting task to return this information to patients in ways that would be helpful and accurate.[12]

While the number of clinically relevant IFs that could result from each sequence generated remains unclear, questions about the capacity of laboratories, providers, and health systems to manage the interpretation and presentation of findings to patients remain pertinent.[3,13] Because of the unprecedented ability of NGS to identify variants across the entire genome, most IFs will have unknown clinical significance. A process and criteria are needed to organize the return of results from IFs to enable end users to focus on the small number of clinically relevant variants.

Berg and colleagues proposed a framework for *a priori* categorization of genomic loci into three "bins" for management of IFs.[14] Any given individual will have few or no clinically actionable IFs (Bin 1).[11] These results should have well understood and validated

associations between genotype and phenotype, and are clinically actionable, meaning that interventions could delay or prevent manifestations, or health outcomes could be improved through early identification and treatment. IFs that are clinically valid but not actionable (Bin 2) are likely to be slightly more numerous.[11] These results have well understood and validated associations between genotype and phenotype, but are *not* immediately clinically actionable. Examples include variants relevant to pharmacogenomics (which only become actionable when a relevant drug is prescribed), most single gene disorders, and risk alleles such as *APOE* genotype for Alzheimer's disease risk (for which there is not yet effective prevention or treatment). Although some patients may be interested in receiving information about these IFs, Berg and colleagues argued that their return should not be automatic since the knowledge would not definitely improve health outcomes, and some results could result in psychosocial harms. Most genomic loci will fall into the third category of uninformative IFs (Bin 3), where there is an insufficient knowledge base on the association between genotype and phenotype. These IFs would typically not be considered for return of results in a clinical setting.

Several initiatives have begun to guide decisions about which IFs to return to patients or research participants from NGS platforms. At least three initiatives have been sponsored by the National Human Genome Research Institute including the electronic medical records and genomics (eMERGE) Network,[15] the Clinical Sequencing Exploratory Research (CSER) consortium, and the Return of Results (ROR) consortium. Professional organizations have established committees to assess NGS platforms (e.g., College of American Pathologists[16] and the American College of Medical Genetics and Genomics[2]). For the past 8 years, the CDC Office of Public Health Genomics has supported the independent Evaluation in Genomic Applications in Practice and Prevention (EGAPP) Working Group to develop evidence based recommendations on the use of genomic tests in clinical practice and disease prevention.[17] The EGAPP Working Group (EWG) sponsored and guided the Knowledge Synthesis Center (KSC) to develop a process to determine which gene/condition pairs belong in the category of clinically actionable IFs in adults. The KSC collaborated with several members of the Working Group to extend the framework proposed by Berg and colleagues, by clarifying and operationalizing criteria for determining clinically actionable IFs. The method includes a structured process to identify evidence, consider the evidence against *a priori* exclusion and inclusion criteria, and present evidence-based considerations to decision-makers.

METHODS

Method Development

We adapted the conceptual framework of population screening, which includes ethical, procedural, and statistical considerations. A higher standard is needed in the case of screening because the pre-test probability is low, which affects the likelihood that a detected mutation is a true health-impairing variant. There must be at least moderate certainty that screening-based identification of the gene/condition and subsequent management has been shown to have a net benefit—that health benefits will outweigh potential harms.[18]

Typically, screening tests should be held to a very high standard, with clinical[19,20] recommendations made only after a thorough, systematic review of the available evidence. A systematic review of each gene mutation would not be an effective or efficient approach, due to the sizeable, very active, but largely developmental clinical research base. Therefore, pragmatic choices were made to increase the feasibility and scalability of the method to the entire genome, such as limiting the sources that would be considered (i.e., a reproducible but not comprehensive search for evidence).[21] Our approach is influenced by the principles of systematic reviews, including: 1) the need for *a priori* objective criteria upon which to make decisions; 2) an emphasis on transparency and reproducibility of decision-making; 3) clear and transparent communication of methods and results; and 4) clear procedures to allow results to be revised as new evidence emerges. To develop these criteria, we considered published articles on population screening in general, newborn screening, or screening for specific genetic conditions.[22–32]

Our method is composed of three distinct stages. In stage I, we define rule-out criteria to rapidly eliminate genes/conditions that do not meet a baseline threshold. In stage II, genes/conditions that pass this initial threshold are subjected to a more resource intensive search of the literature with the relevant evidence synthesized into a short summary. In stage III, decision makers consider the summary for final selection of clinically actionable IFs.

Method Evaluation

Rater Agreement in Stage I—We evaluated the same 99 conditions selected by Green et al.[12] to assess the method for stage I. The 99 conditions were primarily selected from the GeneTests website (http://www.ncbi.nlm.nih.gov/sites/GeneTests/), based on the frequency of laboratory testing, among other considerations. Each condition was evaluated independently by two raters (K.G., J.W., E.M.W., K.S., H.F., D.C.O., or J.S.B.) and categorized as retained or ruled out. We calculated overall agreement and Cohen's kappa[33] as measures of inter-rater agreement. We adjudicated disagreements by discussion among raters to clarify reasons for disagreement, and erred towards 'retain' if there was ambiguity.

External Validity—We compared the categorization of the 99 conditions for stage I (retain/rule out) to the findings of an expert-based approach, which we refer to as the Green method.[12] In the Green method, 16 specialists (including J.S.B.), including clinical geneticists and molecular laboratory directors, evaluated 99 conditions as to whether they would routinely disclose or not disclose the information when discovered as an IF. Specialists were asked to assume the sequencing was perfectly accurate, family history was not available, the patient had no previously recognized clinical features consistent with the disease variant under consideration, the patient's sex was known, and the patient was an adult, but the exact age was unknown. Each specialist recorded their response independently (with no explicit guidelines regarding clinical actionability) The overall results were reported as the proportion of specialists who responded to 'disclose' the finding, which ranged from 50–100% across the 99 conditions. We infer that conditions with a high level of concordance among these experts represent those with the highest degree of clinical actionability and thus agreement that such information would be important to return as IFs.

Clinical Scenario

We assume an adult patient has received WGS or WES as part of their clinical care for an unspecified indication, and the sequencing has acceptable analytic validity. Sex is known, but not the patient's specific age, or any other personal or familial medical history. The patient is currently undiagnosed with the condition under consideration. This criterion does not imply that the patient is asymptomatic. Asymptomatic patients are either 'disease-free' or the condition is present, but the physical signs are 'undetected' by the patient or their clinician (e.g. tumors or precursor lesions may be present years before cancer symptoms appear). Symptomatic patients can also have symptoms that are non-specific, so the genetic condition is 'unrecognized'. For instance, a patient may be aware of and receive treatment for high cholesterol, but they may be unaware that they carry a mutation that causes familial hypercholesterolemia. All of these scenarios–disease-free, undetected disease, and unrecognized disease–are part of a spectrum of undiagnosed patients.

RESULTS

Stage I

The purpose of stage I is to eliminate genes/conditions from further consideration that do not meet a minimal threshold of clinical actionability. We expect a large majority of genes/conditions will be ruled out in this stage. Stage I has three pre-defined criteria addressed in five questions (Table 1). The three criteria are: 1) actionability, 2) moderate or high penetrance, and 3) association with a significant health condition. The criteria for stage I are documented on the Binning Dashboard (Supplemental Table 1). If any of the three criteria are not met, the gene/condition is excluded from further consideration. For each gene/condition it is most efficient to begin with an area suspected to not meet the criteria.

References used to assess the criteria include existing guidelines, systematic reviews, or expert-derived guidance [Table 2]. These resources are identified using a predefined method for searching existing databases for related materials. Only one reference is required to meet any given criteria. Thus, the search procedure is focused and not necessarily comprehensive.

This stage requires high sensitivity, so the gene/condition should be retained when it is uncertain if the minimum threshold is met. Two reviewers assess each gene/condition independently. Discrepant findings are adjudicated by consensus with additional input from a third party, as needed.

Criteria 1: Actionability—For disease-free patients, actionability implies an effective intervention to delay or prevent clinical manifestations or reduce disease impact. For patients with undetected disease, actionability includes screening for earlier diagnosis and to increase the likelihood of less burdensome disease. For symptomatic (but clinically unrecognized) patients, actionability includes alterations in patient management proven to provide benefit. Other actions include family management to improve outcomes for family members (referral to genetic services, or reproductive decision-making alone is not sufficient), or avoidance of circumstances for the patient (e.g., behavioral modifications, diet, exercise, smoking cessation).

Sufficient support of actionability is derived from a practice guideline, expert-derived guidance, or a systematic review. If no such guidance or review exists, our process defines the gene/condition pair as not actionable. If a guideline or systematic review is available, but only recommends care once symptoms have manifested, then the actionability criteria has not been met for an *undiagnosed* adult. If clinical signs and symptoms would always be recognized in childhood or early adolescence, the condition is not considered actionable *in adults*.

Criteria 2: Penetrance—At least one variant in the gene(s) under consideration should have high or moderate penetrance or risk in any population. We selected an arbitrary threshold for the penetrance of either 40% or a measure of relative risk of 2 or greater. While the absolute penetrance for a condition may be low, the relative risk compared with the general population may still be significant. For example, the penetrance of C282Y homozygosity in the *HFE* gene is 13.5% for the development of hereditary hemochromatosis;[34] however, the odds ratio for liver disease is 3.9 and up to 11 for hepatocellular carcinoma.[35] Penetrance data from studies in affected individuals may be used if data from an unselected population is not available. If the data regarding penetrance is unavailable, and all other criteria are met, the gene/condition should be retained to stage II.

Criteria 3: Significance/Burden of Condition—The condition should cause significant morbidity or mortality in adults. This does not include normal human variation such as eye color, hair color, skin color, or body size measurements such as height. We do not address infertility alone or reproductive decision making as part of this approach, recognizing that these could be important in some settings.

Agreement in Stage I

The inter-rater agreement was substantial with an overall agreement of 89% and a kappa of 0.70. Both reviewers retained 55 and both ruled-out 30 of the 99 conditions. For the 14 conditions where the reviewers disagreed, the reasons were: 4 disagreements about the presence or absence of a guideline, 1 disagreement about the overall actionability, 3 disagreements about the actionability in adults, 5 disagreements about whether the penetrance was high/moderate or low, and 1 disagreement about multiple areas. After resolving the disagreements, 62 gene/conditions were retained to stage II.

External Validity of Stage I

The results of the Stage I evaluation compared favorably with the expert-based Green method (Table 3). In most (88%) conditions that we retained after stage I, a high percentage (≥ 88%) of experts also agreed to recommend disclosure. Likewise, most conditions (92%) that we ruled-out after stage I similarly had lower agreement (<70%) among experts to disclose. When examining 37 conditions with only moderate agreement (75–81%) between experts in the Green method, we retained an intermediate proportion (46%), and 9 out of 14 disagreements between reviewers in our method were among these 37 conditions.

We ruled out six conditions that a high percentage (≥ 88%) of experts selected to disclose in the Green method. Four of these conditions (retinoblastoma, neurofibromatosis 1, MCAD deficiency, and isovaleric acidemia) have predominantly childhood onset, and are unlikely to present in an undiagnosed adult. One condition is not actionable (Tay Sachs disease), and one condition is not actionable in adults (Beckwith-Wiedemann Syndrome). Thus, none of these genes/conditions were deemed likely to ultimately fulfill the criteria of clinically actionable IFs had they been retained in stage II, and the apparent discrepancy between our results and the Green method does not indicate a serious problem with our method.

Stage II

The purpose of stage II is to document and summarize the readily available evidence related to key features of actionability to evaluate candidates for a clinically actionable IF. Although a relatively high bar for clinical actionability must be met for a gene/condition pair to qualify for routine reporting, we specify a relatively low threshold for the type of evidence that is permitted, including non-systematic or expert-based references. This approach is both pragmatic and efficient as this represents the most common available evidence for highly penetrant rare genetic conditions. There is transparency in the level of evidence by documenting the relative strength using a tiered system (Table 4) and quality rating the evidence reference using existing methods.[36,37]

In Stage II we conduct a reproducible search of existing synthesized literature using the predefined method (Table 2). Once a full search for potentially relevant references is completed, we examine these references to determine relevance. Any references deemed irrelevant are excluded.

To assess the relative quality of the identified references, we place each reference into one of four tiers (Table 4). The tiers are used to facilitate a hierarchical method for examining existing literature, starting with the most evidence-based. Different tiers of evidence can be present in the same document. For example, background information in a systematic review is not subjected to the same methods as data from the actual review questions.

We then produce a narrative summary with standardized information that is abstracted and documented for each gene/condition (Table 1; Supplemental Table 2). For each question, we use the reference(s) with the highest available tier of evidence for data abstraction. All sources are referenced if they are in agreement. If not, the reviewer determines the best reference based on considerations including quality, methods (e.g., search strategy, inclusion criteria, analytic methods, funding source), and date of publication (e.g., more recent publications may be more relevant). If there is not a best reference, we abstract data from all relevant references. The tier of evidence is recorded for each abstracted data element.

References are quality rated *only if* they are used as a reference for at least one question in the summary report, *and* there is a discrepancy between references. Thus, quality rating serves as a 'tie breaker' when we cannot otherwise decide on a best reference. For systematic reviews (Tier 1), we use the AMSTAR method.[37] For practice guidelines (Tier 2), we use the AGREE II method.[36] Quality rating for tiers 3 and 4 is not performed, because there is not an established method.

Stage II Case Studies

We conducted case studies for Familial Adenomatous Polyposis (FAP, Supplemental Table 3), Hereditary Hemochromatosis (HH, Supplemental Table 4), and α1-Antitrypsin Deficiency (AAT, Supplemental Table 5). These topics were selected to represent a spectrum in the expert opinion on clinical actionability. For each condition, we identified between 31 and 46 references, of which 13 to 30 were relevant to the topic (Table 5). About 40–50% of the relevant references were Tier 1, and 5 to 10 references were cited in each summary document. We did not complete the quality rating for any of these three topics, because no 'tie breakers' were needed.

Resource requirements for Stage I and Stage II

We estimated the resource requirements to complete these evaluations on a genome wide scale. We assume 2000 topics for review at Stage I; each topic is dual-reviewed and requires one hour to complete per person, and 25% of topics require adjudication of disagreements (15 minutes each for two people). Under these assumptions, stage I would require slightly more than two full time employees for one year. For Stage II, we assume that at most 25% of the 2000 topics in Stage I are retained, and each topic takes between 12 to 20 hours to complete. Stage II requires three to five full time employees for a year.

Stage III

The purpose of stage III is for a decision-making panel of experts to review the evidence in the summary document and make decisions about clinically actionable IFs. Different decision-making bodies may come to different decisions on which IFs should be routinely reported. Panels are convened by various stakeholders including professional organizations, payers, and governmental agencies. One such panel is the EWG, which has broad experience in methods development and making recommendations for genomic applications.[17,38,39] We presented the three case studies to the EWG. In each example, the group had consensus that the methodology produced sufficient documentation to decide that all three conditions are clinically actionable and IFs should be routinely reported (Supplemental Tables 3–5).

DISCUSSION

We have presented a transparent, reproducible, efficient, and evidence informed process for identifying clinically actionable IFs in adults. This work builds upon an established framework[11] by operationalizing the categorization of clinically actionable IFs.[14] We implemented the method in stage I for 99 conditions, and demonstrated both high inter-rater agreement (89%) and external validity compared with the Green et al. expert-based approach. While the Green method is not a gold standard, it is the only published method available for comparison. We pilot-tested the method using three case studies for stages II and III, which showed that the evidence summaries provided sufficient information for the EWG to recommend clinically actionable IFs that should be routinely reported. The process is scalable to the whole genome, and can be completed within a year for 2000 gene/condition pairs by about two people for Stage I and three to five people for Stage II.

As genome-scale sequencing tests are more routinely incorporated into clinical care, the ability to centralize the determination of clinically actionable IFs in a transparent and reproducible way is critical. Clinicians cannot be expected to individually assess the evidence across the entire genome, and then make individual determinations about returning IFs. Methods such as this one could facilitate centralized recommendations that can be confidently adopted by clinical laboratories and physicians, resulting in more consistent care across providers. Also, the evidence base is rapidly developing, so the need for transparency in decisions as to why gene/conditions are not clinically actionable supports updating recommendations as new information becomes available. Guidance on the appropriate reporting of IFs is relevant to many stakeholder groups.

Our method, while grounded in evidence-based principles, is also pragmatic in that it allows consideration of both expert-based and evidence-based guidelines. This approach recognizes the still-limited clinically relevant evidence on genetic mutations for most conditions, and the fact that very rare conditions may never accumulate the level of evidence required for traditional evidence-based practice guidelines. Our use of tiers to indicate the relative strength of the methodology used for reference documents allows flexibility for decision-makers. For example, if the usual presentation is sudden death, decision-makers might accept a low threshold of evidence in favor of prevention. The method provides balance through consideration of the potential risks and burdens, as well as benefits of interventions. Our approach uses consistent, but not comprehensive, search strategies to identify already summarized or synthesized evidence, allowing the method to be scalable to the whole genome. This systematic method would be made even more acceptable if combined with opportunities for public input, so that any non-reportable gene/condition pair can be revised with the availability of new (or overlooked) information, a practice that has been used by PLoS Currents/Genomic Tests (http://currents.plos.org/genomictests/).

There are several limitations of the proposed approach. Further experience, including clarification for some of the criteria, would strengthen the method. Semi-quantitative measures of the elements of actionability would avoid subjective interpretation of these criteria and facilitate thresholds to determine actionability. Evidence will change over time with advances in medical genetics, necessitating re-evaluation of genes/conditions. To account for these changes, the method needs an approach for updating and consideration of raised objections. The proposed method is conservative in assuming that no personal or family history information will be available to guide interpretation. Individual results will need to be contextualized by a clinician with expertise in genetics in order to personalize the interpretation.

This method is completely agnostic to the specific technology used to detect any given variant. Thus, we consider the assessment of analytic validity of molecular testing platforms as a necessary, but separate, issue. Although we do not intend to implement the method at the level of specific variants, in clinical practice variants within genes will need to be classified as deleterious or benign. This topic is being addressed by clinical laboratories and efforts such as ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) and the Human Variome Project (http://www.humanvariomeproject.org/).

This work is an important step toward identifying clinically actionable IFs on a genome-wide scale. Future work is needed to confirm our findings by applying all three stages to more cases and to address situations that were outside the scope of this work. Berg and colleagues explicitly recognize a category of variants that are clinically validated but not actionable (Bin 2), which can be further subdivided into categories based on the risk of psychosocial harm. The proposed method does not address how to classify IFs within this category NGS technology will likely be used in clinical settings that we did not address, including reproductive decision making, newborn screening, and pediatric cases. These scenarios will need significant revision of the method to address them. The method may achieve the widest adoption if the acceptability of the criteria and staged approach is assessed among diverse stakeholders, and potentially modified (e.g., selecting different thresholds) to include different perspectives. Nevertheless, we expect that application of this method will result in a robust framework for the subsequent analysis and management of IFs from individual genome-scale sequencing assays.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grant Support: Center for Disease Control and Prevention

This work was funded by a grant from the Centers for Disease Control and Prevention Office of Public Health Genomics, which established the Knowledge Synthesis Center (GD000076; PIs: Goddard, Whitlock). The authors would also like to thank the members of the EGAPP Working Group for their insightful comments on previous versions of this method, and for their participation in the case studies for stage III of the process. *EGAPP Working Group Members: Ned Calonge, MD, MPH (Chair) (The Colorado Trust); Roger D. Klein, MD, JD (Vice Chair) (Cleveland Clinic Foundation); Jonathan S. Berg, MD, PhD (Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill); Doug Campos-Outcalt, MD, MPA (Department of Family, Community and Preventive Medicine, University of Arizona College of Medicine, Phoenix); Benjamin Djulbegovic, MD, PhD (University of South Florida, and H. Lee Moffitt Cancer Center and Research Institute); Theodore G. Ganiats, MD (University of California San Diego School of Medicine, Department of Family and Preventive Medicine, and University of California San Diego Health Services Research Center); Cecile Janssens, PhD (Department of Epidemiology, Rollins School of Public Heath, Emory University); Kenneth Offit, MD, MPH (Clinical Genetics Service, Memorial Sloan-Kettering Cancer Center): Stephen G. Pauker, MD, MACP, FACC, ABMH (Division of Clinical Decision Making, Informatics and Telemedicine, Department of Medicine, Tufts Medical Center); Margaret Piper, PhD, MPH (Blue Cross/Blue Shield Association Technology Evaluation Center); Carolyn Sue Richards, PhD, FACMG (Knight Diagnostic Laboratories, Oregon Health & Science University); Ora L. Strickland, PhD, DSc (Hon.), RN, FAAN (College of Nursing & Health Sciences, Florida International University); David L. Veenstra, PharmD, PhD (Pharmaceutical Outcomes Research and Policy Program, and Institute for Public Health Genetics, University of Washington); Marc S. Williams, MD (Genomic Medicine Institute, Geisinger Health System); Doris T. Zallen, PhD (Department of Science and Technology in Society, Virginia Polytechnic Institute and State University).

Reference List

1. Katsnelson A. Human Genomes: Genomes by the thousands. Nature. 2010 Oct 8.467:1026–1027. [PubMed: 20981067]
2. American College of Medical Genetics and Genomics. [Accessed on: 10-4-2012] American College of Medical Genetics and Genomics (ACMG) Policy Statement Points to Consider in the Clinical Application of Genomic Sequencing. Available at: http://www.acmg.net/StaticContent/PPG/Clinical_Application_of_Genomic_Sequencing.pdf
3. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. JAMA. 2006 Jul 12; 296(2):212–215. [PubMed: 16835427]

4. Levy S, Sutton G, Ng PC, et al. The diploid genome sequence of an individual human. PLoS Biol. 2007 Sep 4.5(10):e254. [PubMed: 17803354]
5. Wheeler DA, Srinivasan M, Egholm M, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature. 2008 Apr 17; 452(7189):872–876. [PubMed: 18421352]
6. Wang J, Wang W, Li R, et al. The diploid genome sequence of an Asian individual. Nature. 2008 Nov 6; 456(7218):60–65. [PubMed: 18987735]
7. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008 Nov 6; 456(7218):53–59. [PubMed: 18987734]
8. Ashley EA, Butte AJ, Wheeler MT, et al. Clinical assessment incorporating a personal genome. Lancet. 2010 May 1; 375(9725):1525–1535. [PubMed: 20435227]
9. McKernan KJ, Peckham HE, Costa GL, et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Res. 2009 Sep; 19(9):1527–1541. [PubMed: 19546169]
10. Via M, Gignoux C, Burchard EG. The 1000 Genomes Project: new opportunities for research and social challenges. Genome Med. 2010; 2(1):3. [PubMed: 20193048]
11. Berg JS, Adams M, Nassar N, et al. An informatics approach to analyzing the incidentalome. Genet Med. 2013 Jan; 15(1):36–44. [PubMed: 22995991]
12. Green RC, Berg JS, Berry GT, et al. Exploring concordance and discordance for return of incidental findings from clinical sequencing. Genet Med. 2012 Apr; 14(4):405–410. [PubMed: 22422049]
13. Biesecker LG, Burke W, Kohane I, Plon SE, Zimmern R. Next-generation sequencing in the clinic: are we ready? Nat Rev Genet. 2012 Nov; 13(11):818–824. [PubMed: 23076269]
14. Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. Genet Med. 2011 Jun; 13(6):499–504. [PubMed: 21558861]
15. Fullerton SM, Wolf WA, Brothers KB, et al. Return of individual research results from genome-wide association studies: experience of the Electronic Medical Records and Genomics (eMERGE) Network. Genet Med. 2012 Apr; 14(4):424–431. [PubMed: 22361898]
16. Schrijver I, Aziz N, Farkas DH, et al. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the association for molecular pathology. J Mol Diagn. 2012 Nov; 14(6):525–540. [PubMed: 22918138]
17. 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. 2009 Jan; 11(1):3–14. [PubMed: 18813139]
18. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. Ann Intern Med. 2007 Dec 18; 147(12):871–875. [PubMed: 18087058]
19. Institute of Medicine. [Accessed on: 11-21-2012] Clinical Practice Guidelines We Can Trust. Available at: http://www.iom.edu/Reports/2011/Clinical-Practice-Guidelines-We-Can-Trust.aspx
20. U.S. Preventive Services Task Force. [Accessed on: 1-14-2013] USPSTF Procedure Manual Section 5: Methods for Arriving at a Recommendation. Available at: http://www.uspreventiveservicestaskforce.org/uspstf08/methods/procmanual5.htm
21. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev. 2012; 1(1):10. [PubMed: 22587960]
22. Andermann A, Blancquaert I, Beauchamp S, Dery V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bull World Health Organ. 2008 Apr; 86(4):317–319. [PubMed: 18438522]
23. Wilson, JMG.; Jungner, G. WHO Chronicle. Vol. 22. Geneva: World Health Organization; 1968. Principles and practice of screening for disease; p. 473Public Health Papers, #34
24. UK National Screening Committee. [Accessed on: 7-23-2012] UK National Screening Committee Criteria. Available at: http://www.screening.nhs.uk/criteria
25. Cogswell ME, McDonnell SM, Khoury MJ, Franks AL, Burke W, Brittenham G. Iron overload, public health, and genetics: evaluating the evidence for hemochromatosis screening. Ann Intern Med. 1998 Dec 1; 129(11):971–979. [PubMed: 9867750]

26. Calonge N, Green NS, Rinaldo P, et al. Committee report: Method for evaluating conditions nominated for population-based screening of newborns and children. Genet Med. 2010 Mar; 12(3): 153–159. [PubMed: 20154628]
27. Watson MS. Current status of newborn screening: decision-making about the conditions to include in screening programs. Ment Retard Dev Disabil Res Rev. 2006; 12(4):230–235. [PubMed: 17183572]
28. Padilla CD. Towards universal newborn screening in developing countries: obstacles and the way forward. Ann Acad Med Singapore. 2008 Dec; 37(12 Suppl):6–4. [PubMed: 19904464]
29. Therrell BL Jr. Considerations in choosing screening conditions: one (US) approach. Ann Acad Med Singapore. 2008 Dec; 37(12 Suppl):22–25. [PubMed: 19904477]
30. Khoury MJ, McCabe LL, McCabe ER. Population screening in the age of genomic medicine. N Engl J Med. 2003 Jan 2; 348(1):50–58. [PubMed: 12510043]
31. Harris R, Sawaya GF, Moyer VA, Calonge N. Reconsidering the criteria for evaluating proposed screening programs: reflections from 4 current and former members of the U.S. Preventive services task force. Epidemiol Rev. 2011 Jul; 33(1):20–35. [PubMed: 21666224]
32. Burke W, Tarini B, Press NA, Evans JP. Genetic screening. Epidemiol Rev. 2011 Jul; 33(1):148–164. [PubMed: 21709145]
33. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960; 20:37–46.
34. European Association For The Study Of The Liver. EASL clinical practice guidelines for HFE hemochromatosis. J Hepatol. 2010 Jul; 53(1):3–22. [PubMed: 20471131]
35. Ellervik C, Birgens H, Tybjaerg-Hansen A, Nordestgaard BG. Hemochromatosis genotypes and risk of 31 disease endpoints: meta-analyses including 66,000 cases and 226,000 controls. Hepatology. 2007 Oct; 46(4):1071–1080. [PubMed: 17828789]
36. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010 Dec 14; 182(18):E839–E842. [PubMed: 20603348]
37. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009 Oct; 62(10):1013–1020. [PubMed: 19230606]
38. Veenstra DL, Piper M, Haddow JE, et al. Improving the efficiency and relevance of evidence-based recommendations in the era of whole-genome sequencing: an EGAPP methods update. Genet Med. 2012 Sep 6.
39. Botkin JR, Teutsch SM, Kaye CI, et al. Outcomes of interest in evidence-based evaluations of genetic tests. Genet Med. 2010 Apr; 12(4):228–235. [PubMed: 20118789]
40. [Accessed on: 11-21-2012] Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Available at: http://www.cochrane-handbook.org/

Table 1

Criteria for Stages I & II

Stage I		
Criteria	**Questions**	
Actionability	Is there a practice guideline or systematic review for the genetic condition?	
	Does the practice guideline or systematic review indicate that the result is actionable in one or more of the following ways: a) patient management, b) surveillance or screening, c) family management, or d) circumstances to avoid?	
	Is the result actionable in an undiagnosed adult with the genetic condition?	
Penetrance	Is there at least one known pathogenic variant with at least moderate penetrance (≥40%) or moderate relative risk (≥2) for important health implications in any population?	
Significance/Burden of disease	Is this condition an important health problem?	
Stage II		
Criteria	**Questions**	
Significance/Burden of disease	What is the nature of the threat to health for an individual carrying a deleterious allele?	
	Disease Incidence	Disease Prevalence
	Clinical Features (Signs/Symptoms)	Natural History
	Significance/Burden of Condition	
Actionability	How effective are available interventions for preventing the harm?	
	Patient Management	Surveillance
	Family Management	Circumstances to Avoid
Penetrance	What is the chance that this threat will materialize?	
	Prevalence	Penetrance
	Expressivity	Relative Risk
Acceptability of Intervention	How acceptable are available interventions in terms of the burdens or risks placed on the individual?	
Risks	Would the underlying risk or condition escape detection prior to harm in the setting of recommended care?	
	Chance to escape clinical detection in adults	

Table 2

Search method to identify references of evidence based practice guidelines, systematic reviews, or expert consensus based practice guidelines. For stage I, these steps are completed only until *at least one* relevant reference is identified. For stage II, all steps are completed to identify a comprehensive list of references based on this search method.

Step	Method
1	Identify synonyms for the condition, name of involved gene(s), and the OMIM identification number from the following sites: OMIM: http://www.ncbi.nlm.nih.gov/omim GeneTest Review: http://www.ncbi.nlm.nih.gov/sites/GeneTests/review
2	If a GeneTest review exists, search for a link to a treatment guideline within the text (e.g., a “Guideline” flag) or reference list.
3	Search OMIM for a link to a guideline.
4	Search guidelines.gov for condition name and any synonyms identified in step 1.
5	In Pubmed: Select MeSH from the search pull down menu. Search for the condition. In the list of results select the one that is relevant. On the right side select “Add to search builder” and then “search pubmed”. On the side under article types select “practice guidelines” and “systematic reviews”.
6	If no relevant MeSH term is identified using the condition name see if MeSH terms exist for the synonyms, individual genes, or a larger category that the disease might fall under (e.g., inborn errors of metabolism). If no terms are identified or the term is too broad, search Pubmed using the condition or gene name as text words and limit article types to “practice guidelines” and “systematic reviews”.
7	Consider the following sources if no systematic reviews or practice guidelines are identified in the previous steps: OrphaNet: http://www.orpha.net/consor/cgi-bin/index.php Clinical Utility Gene Cards: http://www.nature.com/ejhg/archive/categ_genecard_012011.html

Table 3

External validity compared with expert opinion-based method by Green et al.[12]

Green Method	Proposed Method*		
% Responded yes, disclose	Retain	Rule–out	Total
100%	20 (95%)	1	21
94%	11 (78%)	3	14
88%	13 (87%)	2	15
81%	6	8 (43%)	14
75%	11	12 (45%)	23
69%	0	6 (100%)	6
63%	1	2 (66%)	3
56%	0	2 (100%)	2
50%	0	1 (100%)	1
Total	62	37	99

* Results after adjudication

Table 4

Evidence Tiers for Stage II.

Tier	Reference Type
1	Evidence from a systematic review, or a meta-analysis, or a clinical practice guideline clearly based on a systematic review[1].
2	Evidence from clinical practice guidelines or broad-based expert consensus with some level of evidence review, but using unclear methods or using sources that were not systematically identified[1].
3	Evidence from another reference with non-systematic review of evidence (e.g., GeneTest Reviews, OrphaNet, and Clinical Utility Gene Cards, opinion of a single or few (<5) experts) with additional primary literature cited.
4	Evidence from another reference with non-systematic review of evidence (e.g., GeneTest Reviews, OrphaNet, and Clinical Utility Gene Cards, opinion of a single or few (<5) experts) with no citations to primary data sources.

[1] systematic review of the evidence means that traditional systematic review methods are followed including: a) a clearly stated set of objectives, b) an explicit, reproducible methodology, c) systematic search that attempts to identify all studies that would meet the eligibility criteria, d) inclusion and exclusion criteria for studies are pre-defined, and e) an assessment of the validity of findings in the included studies, and f) a systematic presentation and synthesis of the characteristics and findings of the included studies.[40]

Table 5

Identified References for Three Case Studies in Stage II.

	FAP	HH	AAT Deficiency
Total Number of References Identified	46	34	31
Number of Related References	28	13	16
Number of References in Each Tier			
Tier 1	14	5	8
Tier 2	11	2	4
Tier 3 or 4	3	6	4
Number of References Cited	5	10	8