Clinical Genomic Database

Benjamin D. Solomon[a,1], Anh-Dao Nguyen[b], Kelly A. Bear[a,c], and Tyra G. Wolfsberg[b]

[a]Medical Genetics Branch and [b]Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; and [c]Department of Pediatrics, Walter Reed National Military Medical Center, Bethesda, MD 20889

Edited by C. Thomas Caskey, Baylor College of Medicine, Houston, TX, and approved May 2, 2013 (received for review February 19, 2013)

Technological advances have greatly increased the availability of human genomic sequencing. However, the capacity to analyze genomic data in a clinically meaningful way lags behind the ability to generate such data. To help address this obstacle, we reviewed all conditions with genetic causes and constructed the Clinical Genomic Database (CGD) (http://research.nhgri.nih.gov/CGD/), a searchable, freely Web-accessible database of conditions based on the clinical utility of genetic diagnosis and the availability of specific medical interventions. The CGD currently includes a total of 2,616 genes organized clinically by affected organ systems and interventions (including preventive measures, disease surveillance, and medical or surgical interventions) that could be reasonably warranted by the identification of pathogenic mutations. To aid independent analysis and optimize new data incorporation, the CGD also includes all genetic conditions for which genetic knowledge may affect the selection of supportive care, informed medical decision-making, prognostic considerations, reproductive decisions, and allow avoidance of unnecessary testing, but for which specific interventions are not otherwise currently available. For each entry, the CGD includes the gene symbol, conditions, allelic conditions, clinical categorization (for both manifestations and interventions), mode of inheritance, affected age group, description of interventions/rationale, links to other complementary databases, including databases of variants and presumed pathogenic mutations, and links to PubMed references (>20,000). The CGD will be regularly maintained and updated to keep pace with scientific discovery. Further content-based expert opinions are actively solicited. Eventually, the CGD may assist the rapid curation of individual genomes as part of active medical care.

genome sequencing | genomic medicine | whole-genome sequencing

As a result of new technologies that allow efficient and affordable high-throughput sequencing, genomic sequencing is becoming increasingly prevalent in both research and clinical arenas. Currently, this type of sequencing commonly includes exome sequencing, sometimes referred to as "whole-exome sequencing." In exome sequencing, the protein-coding portions of known genes—comprised of ~1–2% of the 6 billion bases in the diploid genome, depending on the platform used—are sequenced (reviewed in ref. 1). Whole-genome sequencing, which additionally includes introns and gene regulatory regions, as well as the rest of the genome, is anticipated to become more widely used as methodologies evolve to allow decreased cost and to meet informatics challenges. Whole-genome sequencing may supplant exome sequencing in the relatively near future.

To date, the most impressive applications of human genomic sequencing (we refer here to both exome and genome sequencing as "genomic sequencing") have been the detection of the genetic causes of relatively rare conditions (1–3). However, genomic sequencing has myriad potential applications in more general clinical medicine, including in healthy individuals (3–8).

Despite the promise of the "age of genomic medicine," a key barrier to translating the power of genomic sequencing to the general clinical setting involves the time and resources required for clinically relevant analysis beyond searching for the cause of a single, usually relatively severe, disease. A number of freely or commercially available tools allow curation of individual genomes, including analysis of variant type, predicted pathogenicity of a particular variant, and associations of the gene or specific variant with known health conditions. After this level of curation, however, a key obstacle arises when trying to determine which detected variants may warrant further follow-up, including potential clinical interventions, or would otherwise alter patient-based care. Typically, clinically oriented analysis involves an approach in which detected potentially pathogenic variants affecting known disease-associated loci are largely individually queried to determine their clinical applicability (9–13).

To help address this problem, we manually investigated all conditions with known genetic causes. We constructed a database focusing on genetic data as relates to the availability of condition-specific interventions and how finding a pathogenic mutation would be anticipated to affect medical care. The current goal of this project is to disseminate the database to solicit content-oriented input, related to both clinical and molecular aspects of the database, from experts in individual genes and conditions. Eventually, this database may be used to aid in the efficient analysis of individual genomes for clinically significant health information. The Clinical Genomic Database (CGD) is freely available at: http://research.nhgri.nih.gov/CGD.

Results

At the time of writing, (April 2013), the CGD includes 2,616 genes in which mutations are known to cause human disease or have clinically significant pharmacogenomic implications. For 1,333 of these genes, medical interventions meeting the described criteria are available (*Materials and Methods*). The CGD includes an additional 1,283 genes for which these types of clinical interventions are not yet available based on current medical knowledge, but in which mutations may nonetheless be clinically relevant. Knowledge of mutations resulting in one of this latter group of conditions may thus be beneficial for a variety of reasons. These reasons include an enhanced ability to select optimal supportive care, make more fully informed medical choices, consider questions related to disease prognosis, make reproductive decisions, and avoid lengthy, expensive, and potentially risky "diagnostic odysseys."

The Web interface to the CGD allows searching by gene or condition, as well as browsing by clinical categories (for both manifestations and interventions). The CGD can be queried using single or multiple search terms, including large files of gene names or terms. For each entry, the database includes the gene symbol, conditions, allelic conditions, clinical categorization (by manifestation and intervention categories), mode of inheritance, age category (pediatric or adult) in which interventions are indicated based on descriptions in the medical literature, general descriptions of the interventions/rationale, and individually linked references (>20,000). See Table 1 for a summary of the categories included and the numbers of genes within each category; see Table

Author contributions: B.D.S. designed research; B.D.S., A.-D.N., and K.A.B. performed research; B.D.S., A.-D.N., K.A.B., and T.G.W. contributed new reagents/analytic tools; B.D.S., A.-D.N., K.A.B., and T.G.W. analyzed data; and B.D.S. and T.G.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. E-mail: solomonb@mail.nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1302575110/-/DCSupplemental.

Table 1. Organization of the 2,616 genes included in the by Manifestation categories and Intervention categories

Category	Number of genes in Manifestation categories*	Number of genes in Intervention categories†
Allergy/Immunology/Infectious	272	251
Audiologic/Otolaryngologic	273	150
Biochemical	452	175
Cardiovascular	486	267
Craniofacial	384	0
Dental	95	0
Dermatologic	391	33
Endocrine	300	153
Gastrointestinal	393	106
General	45	1,291
Genitourinary	183	27
Hematologic	326	247
Musculoskeletal	801	43
Neurologic	1183	46
Obstetric	38	35
Oncologic	230	229
Ophthalmologic	566	51
Pharmacogenomic	0‡	186
Pulmonary	96	60
Renal	355	133

Values shown may differ from updated versions available on the CGD website. To reflect the multisystemic nature of many genetic disorders and to allow comprehensive browsing, each entry may be listed under multiple categories.

*Manifestation categories include organ systems that are primarily affected by mutations in the corresponding gene. Recognition of these affected systems may aid in condition recognition, as well as supportive care. Genes not categorized by organ systems within the Manifestation categories are included in the General category here.

†Intervention categories include organ systems for which specific medical interventions are available. Genes not meeting the described criteria for these specific interventions (see *Materials and Methods*) are included in the General category here.

‡Pharmacogenomic-related genes are all categorized under the Intervention categories rather than Manifestation categories.

2 for an example of a single entry. The entire current contents (as of April 2013) are available as Dataset S1. The most current versions of selected data (e.g., individual or multiple entries or categories), as well as the entire contents, can also be freely downloaded directly through the CGD Web site.

Discussion

The clinical interpretation of genomic data involves multiple complex and controversial issues, and the body of literature on the subject is large and rapidly growing. One key challenge involves the ability to efficiently analyze vast amounts of data in a medically meaningful way. This issue will grow as large-scale sequencing becomes more frequently used in clinical situations. We expect that the creation and ongoing updating of the CGD to maintain currency will contribute to address this challenge.

We feel that the CGD fills a currently largely unfilled but critical niche in the field of clinical genomics and genomic medicine. Resources, such as Online Inheritance in Man (OMIM) (www.omim.org), provide vast repositories of rich clinical and genetic knowledge, but may be harder to query for efficient clinically oriented analysis. Variant-related databases, including the Human Gene Mutation Database (HGMD) (www.hgmd.cf.ac.uk and www.biobase-international.com/product/hgmd) are valuable when considering the potential pathogenicity of detected genetic variants, but do not focus directly on clinical implications in a particular healthcare situation. Other databases and tools, such as the Personal Genome Project (www.personalgenomes.org), contain robustly annotated variant sets and admirably enable powerful analysis of individual genomes (6). However, this latter type of application focuses on broader genomic exploration, especially involving genetic risk factors and association-based susceptibilities. This type of analysis tool may be more cumbersome when considering the optimal clinical approach to genetic findings in real time. Furthermore, the Personal Genome Project has deliberately selected an evaluation strategy involving a peer-produced model that will fill in a "genomic scaffold," intentionally resulting in the steady and ongoing construction of the analysis platform through the input of multiple contributors (6).

We do anticipate that the CGD will serve as a tool that can be used in conjunction with some of the above-mentioned platforms, as well as with other related resources. Admittedly, a key to advancing the field of genomic medicine involves merging the strengths of different resources, including both clinical and genetic/genomic datasets (14, 15). To expedite and encourage this process, the most recently updated contents of the CGD are freely downloadable through the Web site, either in its entirety or as selected portions (e.g., including entire clinical categories). The full content (current as of April 2013) is also available as a single file here (Dataset S1).

One way in which the CGD is different from other resources, which may in some respects be advantageous, is that a single (board-certified) clinical geneticist manually investigated and applied the same rationale to the availability of interventions for each gene/condition. Nevertheless, further large-scale validation and testing will be needed, and the CGD must be flexible to incorporate new data (16, 17). In general, databases and tools like the CGD must be designed to evolve with the pace of genetic/genomic and more general medical discovery. Reasons necessitating a dynamic database include, but are not limited to, new discoveries of genetic sources of disease, development of novel treatment methods, and new clinically relevant findings in previously described disorders. Without active maintenance, resources like the CGD will become almost immediately obsolete.

It is important to point out that the lack of available objective and uniform data related to each gene and condition raises challenges. One reason for this subjectivity is that many conditions have been described in only a small number of individuals, making drawing large-scale evidence-based conclusions difficult. Nevertheless, rare conditions are just as important to the individuals they affect as common conditions, and in the age of "personalized genomic medicine," the individual-scale is at the forefront.

In other words, the rarity and nature of many Mendelian conditions make classic "randomized, placebo-based, double-blind" studies very difficult or impossible. As relates to the creation of the CGD, despite attempts to apply uniform criteria in our analysis, determinations of which conditions have available specific interventions are clearly subjective. This subjectivity stems not only from medical judgment about the condition and its manifestations, but also from an assessment of the available interventions (16–19). One way to address this problem, as several investigations have done (6, 11, 16, 18), is to introduce a semi-quantitative rating system, but adding a numerical score may not always provide adequate justification.

There is also the perhaps larger problem related to the interpretation of the potential pathogenicity of specific variants, including in genes and conditions where the availability and benefit of early interventions is less in question (17, 18, 20). Even in the situation of relatively well-characterized conditions, such as phenylketonuria [resulting from mutations in *phenylalanine hydroxylase* (*PAH*)] (21–23) or high-penetrance cancer predisposition conditions (such as Lynch syndrome) (24–28), a genotype-first approach may be problematic because of challenges related to predicting the functional and clinical consequences of a detected variant. In other words, evaluating the potential benefits of

Table 2. Sample CGD entry for one gene

Gene	Entrez Gene ID	Condition	Inheritance	Age group	Allelic conditions	Manifestation categories	Intervention categories	Interventions/Rationale	References
CTNS	1497	Cystinosis	AR	Pediatric		Biochemical; Endocrine; Musculoskeletal; Ophthalmologic; Renal	Biochemical; Endocrine; Ophthalmologic; Renal	Cystine-depleting agents (cysteamine) instituted at an early age can be beneficial related to manifestations affecting multiple organ systems, perhaps including cognitive development, as well as renal manifestations (renal tubular Fanconi syndrome and glomerular damage), although renal transplantation may be necessary; Surveillance for renal manifestations and related sequelae can allow early detection and management of disease with (in addition to cystine-depleting agents), replacement of renal losses; Dietary management (e.g., ensuring sufficient caloric intake, and with vitamin D and phosphate supplementation) can be beneficial related to potential failure to thrive and hypophosphatemic rickets; Cysteamine eyedrops can be beneficial related to ophthalmologic sequelae; Surveillance for endocrine manifestations (e.g., hypothyroidism, or hypogonadism in males) can allow early detection and medical management, including potential use of growth hormone to optimize height in some individuals	6038997; 4914142; 5443335; 406375; 333912; 7112129; 3307383; 3335962; 3550461; 3674101; 3821824; 3335962; 3185663; 3292915; 381441; 2230837; 552398; 8455682; 8172256; 7593434; 9537412; 10556299; 10417278; 10444339; 10625078; 11001803; 10673275; 12110740; 12442267; 16603246; 17643777; 19863563; 20301574; 20803298; 20814825; 21305353; 21784456; 21868618; 21371554; 21900880; 22903658; 23001048; 23462307; 23538568

The CGD can be queried with a single gene symbol or condition, or with a list. Searches can also be limited by organ-system–based clinical categories (for both Manifestation and Intervention categories). To allow efficient access to complementary resources, including databases of variants and apparently pathogenic mutations, each gene/condition-specific links to the relevant entry (where available) in: 1000 Genomes (www.1000genomes.org); the Short Genetic Variations Database (The Database of Single Nucleotide Polymorphisms, dbSNP) (www.ncbi.nlm.nih.gov/snp/); GeneTests (www.ncbi.nlm.nih.gov/sites/GeneTests/); the HGMD (www.hgmd.cf.ac.uk/); the National Center for Biotechnology Information gene database (www.ncbi.nlm.nih.gov/gene); the National Heart, Lung, and Blood Institute Gene Ontology Exome Sequencing Project Exome Variant Server (http://evs.gs.washington.edu/EVS); OMIM (http://www.omim.org); Genetics Home Reference (http://ghr.nlm.nih.gov/); ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/). Each reference is directly linked to a PubMed abstract (>20,000 articles are individually referenced).

Information shown above may differ from the updated version available on the CGD website. For the sake of viewability, the "Comments" column (which appears on the CGD website) has been omitted from this example entry.

genetic/genomic diagnosis (i.e., before diagnosis on clinical grounds alone) of a disorder is challenging in itself, but is only one piece of the puzzle. An equally large piece involves the evaluation of specific variants, and strategies have been devised to address this issue. One such strategy involves "binning" variants (11, 18). According to this type of algorithm, a variant would be judged to be pathogenic for one of two reasons. The first reason is that the type of variant may by itself predict pathogenicity: for example, a truncating mutation or a large deletion in a gene in which loss-of-function is predicted to cause disease. The second reason, used for variants whose pathogenicity is more difficult to assess (e.g., novel missense variants), would involve the presence of strong evidence showing that the exact detected variant had been previously established as disease-causing (11, 12, 18, 20).

Along these lines, in any approach involving the clinical management of individual patients, genotypic information should not

operate in a vacuum (6, 29). Just as genotypic data, such as the type, location, and novelty of a particular amino acid substitution can aid in the interpretation of a variant, clinical information, including family and medical history, can help determine the consideration of a specific variant (5, 12, 29).

The first goal of the CGD is broad dissemination to solicit content-oriented feedback and input from experts studying relevant genes and conditions. This input can be used to continually revise and improve this resource, as the CGD will be regularly updated through a combination of automated and manual curation. The first long-term objective is the establishment of a user-friendly resource relevant to a wide group of clinicians that can be used as a reference resource in a variety of situations. Eventually, in addition to serving as a reference tool, the CGD may be used as a filter superimposed on automated binning algorithms to help allow efficient, clinically relevant annotation of human genomes.

Materials and Methods

To investigate conditions with identified genetic underpinnings, we individually read all entries in OMIM (www.omim.org) that included conditions with genetic causes, then cross-referenced all entries—and searched for additional entries—within the following publicly available databases: GeneTests (www.ncbi.nlm.nih.gov/sites/GeneTests), Pharmacogenomics Knowledge Base (www.pharmgkb.org), HGMD (www.hgmd.cf.ac.uk), and HGMD Professional (https://portal.biobase-international.com/hgmd) through a site-specific license. Pertaining to each gene and condition described in these databases, we directly analyzed the content of all cited primary references. Published literature that was not included in these databases was also queried through independent PubMed search (by gene and condition name). The most recent date of query was April 17, 2013.

The CGD has been constructed to reflect the multisystemic nature of many genetic conditions to allow more comprehensive browsing by clinical categories. In the CGD, genes were first categorized into Manifestation categories, or the organ systems primarily affected by mutations in the corresponding gene. For many of these organ systems, recognition of the condition's effects and related supportive care may be clinically beneficial. Conditions not grouped within a specific organ system under the Manifestation categories were included in the General category.

Next, genes were separately categorized under Intervention categories by the organ systems for which specific medical interventions were available. In determining the Intervention categories, the following points were considered. These points are based in part on arguments related to the selection of targets for routine newborn screening (30): (*i*) the condition must be clinically significant (i.e., at least some manifestations must result in morbidity or mortality); (*ii*) there must be a currently available, potentially beneficial intervention (this intervention may include preventive measures, surveillance, or medical or surgical treatments, although experimental/research-based interventions were not included); (*iii*) there should be advantage to early (genomic) diagnosis as opposed to discovery of the condition on purely clinical grounds (i.e., without genetic/genomic testing). Regarding this last point, precise diagnosis is challenging for many conditions, and correct recognition based on genetic/genomic diagnosis may allow interventions related to specific manifestations. The efficacy of these interventions would be diminished or lost with later diagnosis, such as might occur based primarily upon clinical presentation. For example, in certain types of Ehlers-Danlos syndrome, which may not always be recognized early enough to allow optimal medical care, genotype-based recognition may allow interventions related to certain cardiovascular manifestations, which may reduce associated morbidity and mortality (31).

For the Intervention categories, all genes not meeting the above criteria were included in the General category. As described above, for many such conditions although a more specific intervention may not be currently available, genetic knowledge may be beneficial related to a number of issues, including the selection of optimal supportive care, prognostic considerations related to medical-decision making, informing reproductive decisions, and avoidance of unnecessary testing as part of the diagnostic process. These entries contain similar information to those classified by organ system, with the exception that the interventions and rationale are not specifically described. Individual experts were contacted in many instances where the availability or efficacy of interventions was unclear.

The Web interface to the CGD allows searching by gene or condition or browsing by categories. For each entry, the database includes the gene symbol, conditions, allelic conditions (conditions resulting from mutations in the same gene, but which themselves may not have a specific intervention available; it must be noted that for many reportedly distinct conditions, there is clearly a phenotypic continuum, such that division into clinically separate conditions can be challenging), clinical categorization (as described above, by both manifestations as well as more specific interventions), inheritance, age [designated as either pediatric (less than 18 y of age) or adult] in which interventions are indicated based on descriptions in the medical literature, and general descriptions of the interventions/rationale. This latter category is not intended to serve in place of comprehensive treatment guidelines nor act as a clinical guide, but rather briefly describes the types of interventions that may be considered.

The CGD currently includes only single gene alterations; it does not include contiguous gene syndromes, although conditions with, for example, demonstrated digenic inheritance are included. Similarly, somatic alterations, such as commonly occur in cancerous processes, are not included, although if a germ-line change in the same gene has been shown to result in disease, those latter conditions are included. The current version does not include susceptibilities or genetic associations, such as those identified through an association-based study. As the database expands in the future, these types of additions would be considered.

ACKNOWLEDGMENTS. The authors thank Leslie G. Biesecker, Derek A. T. Cummings, James P. Evans, Donald W. Hadley, and Maximilian Muenke for support, mentorship, and critical input; Andreas D. Baxevanis for bioinformatic discussions and support; Mark Fredriksen for programming assistance; and all the experts who provided input related to individual genes and conditions. This research was supported by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health.

1. Bamshad MJ, et al. (2011) Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12(11):745–755.
2. Biesecker LG, Shianna KV, Mullikin JC (2011) Exome sequencing: The expert view. *Genome Biol* 12(9):128.
3. Gonzaga-Jauregui C, Lupski JR, Gibbs RA (2012) Human genome sequencing in health and disease. *Annu Rev Med* 63:35–61.
4. Bainbridge MN, et al. (2011) Whole-genome sequencing for optimized patient management. *Sci Transl Med* 3(87):87re3.
5. Solomon BD, et al.; NISC Comparative Sequencing Program (2011) Personalized genomic medicine: Lessons from the exome. *Mol Genet Metab* 104(1-2): 189–191.
6. Ball MP, et al. (2012) A public resource facilitating clinical use of genomes. *Proc Natl Acad Sci USA* 109(30):11920–11927.
7. Johnston JJ, et al. (2012) Secondary variants in individuals undergoing exome sequencing: Screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. *Am J Hum Genet* 91(1):97–108.
8. Solomon BD, Pineda-Alvarez DE, Bear KA, Mullikin JC, Evans JP; NISC Comparative Sequencing Program (2012) Applying genomic analysis to newborn screening. *Mol Syndromol* 3(2):59–67.
9. Lupski JR, et al. (2010) Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 362(13):1181–1191.
10. Tong P, et al. (2010) Sequencing and analysis of an Irish human genome. *Genome Biol* 11(9):R91.
11. Berg JS, Khoury MJ, Evans JP (2011) Deploying whole genome sequencing in clinical practice and public health: Meeting the challenge one bin at a time. *Genet Med* 13(6): 499–504.
12. Solomon BD, et al.; NISC comparative Sequencing Program (2012) Incidental medical information in whole-exome sequencing. *Pediatrics* 129(6):e1605–e1611.
13. Teer JK, Green ED, Mullikin JC, Biesecker LG (2012) VarSifter: Visualizing and analyzing exome-scale sequence variation data on a desktop computer. *Bioinformatics* 28(4):599–600.
14. Oetting WS, et al. (2013) Getting ready for the Human Phenome Project: The 2012 Forum of the Human Variome Project. *Hum Mutat* 34(4):661–666.
15. Hamosh A, et al. (2013) PhenoDB: A new web-based tool for the collection, storage, and analysis of phenotypic features. *Hum Mutat* 34(4):566–571.
16. Green RC, et al. (2012) Exploring concordance and discordance for return of incidental findings from clinical sequencing. *Genet Med* 14(4):405–410.
17. Manolio TA, et al. (2013) Implementing genomic medicine in the clinic: The future is here. *Genet Med* 15(4):258–267.
18. Berg JS, et al. (2013) An informatics approach to analyzing the incidentalome. *Genet Med* 15(1):36–44.
19. Evans JP, Berg JS (2011) Next-generation DNA sequencing, regulation, and the limits of paternalism: The next challenge. *JAMA* 306(21):2376–2377.
20. Xue Y, et al.; 1000 Genomes Project Consortium (2012) Deleterious- and disease-allele prevalence in healthy individuals: Insights from current predictions, mutation databases, and population-scale resequencing. *Am J Hum Genet* 91(6):1022–1032.

21. DiLella AG, Kwok SC, Ledley FD, Marvit J, Woo SL (1986) Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. *Biochemistry* 25(4): 743–749.
22. Holtzman NA, Kronmal RA, van Doorninck W, Azen C, Koch R (1986) Effect of age at loss of dietary control on intellectual performance and behavior of children with phenylketonuria. *N Engl J Med* 314(10):593–598.
23. Drogari E, Smith I, Beasley M, Lloyd JK (1987) Timing of strict diet in relation to fetal damage in maternal phenylketonuria. An international collaborative study by the MRC/DHSS Phenylketonuria Register. *Lancet* 2(8565):927–930.
24. Fishel R, et al. (1993) The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* 75(5):1027–1038.
25. Nicolaides NC, et al. (1994) Mutations of two PMS homologues in hereditary non-polyposis colon cancer. *Nature* 371(6492):75–80.
26. Papadopoulos N, et al. (1994) Mutation of a mutL homolog in hereditary colon cancer. *Science* 263(5153):1625–1629.
27. Miyaki M, et al. (1997) Germline mutation of MSH6 as the cause of hereditary non-polyposis colorectal cancer. *Nat Genet* 17(3):271–272.
28. Hampel H, et al. (2005) Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med* 352(18):1851–1860.
29. Spencer DH, et al. (2011) Direct-to-consumer genetic testing: Reliable or risky? *Clin Chem* 57(12):1641–1644.
30. American College of Medical Genetics' Newborn Screening Expert Group (2006) Newborn screening: Toward a uniform screening panel and system. *Genet Med* 8(Suppl 1):1S–252S.
31. Schwarze U, et al. (2004) Rare autosomal recessive cardiac valvular form of Ehlers-Danlos syndrome results from mutations in the COL1A2 gene that activate the nonsense-mediated RNA decay pathway. *Am J Hum Genet* 74(5):917–930.