

## Horizon scanning for new genomic tests

Marta Gwinn, MD, MPH<sup>1</sup>, Daurice A. Grossniklaus, PhD, MEd<sup>2</sup>, Wei Yu, PhD<sup>2</sup>,  
Stephanie Melillo, MPH<sup>2</sup>, Anja Wulf<sup>3</sup>, Jennifer Flome, MPH<sup>1</sup>, W. David Dotson, PhD<sup>2</sup>,  
and Muin J. Khoury, MD, PhD<sup>2</sup>

**Purpose:** The development of health-related genomic tests is decentralized and dynamic, involving government, academic, and commercial entities. Consequently, it is not easy to determine which tests are in development, currently available, or discontinued. We developed and assessed the usefulness of a systematic approach to identifying new genomic tests on the Internet. **Methods:** We devised targeted queries of Web pages, newspaper articles, and blogs (Google Alerts) to identify new genomic tests. We finalized search and review procedures during a pilot phase that ended in March 2010. Queries continue to run daily and are compiled weekly; selected data are indexed in an online database, the Genomic Applications in Practice and Prevention Finder. **Results:** After the pilot phase, our scan detected approximately two to three new genomic tests per week. Nearly two thirds of all tests (122/188, 65%) were related to cancer; only 6% were related to hereditary disorders. Although 88 (47%) of the tests, including 2 marketed directly to consumers, were commercially available, only 12 (6%) claimed United States Food and Drug Administration licensure. **Conclusion:** Systematic surveillance of the Internet provides information about genomic tests that can be used in combination with other resources to evaluate genomic tests. The Genomic Applications in Practice and Prevention Finder makes this information accessible to a wide group of stakeholders. *Genet Med* 2011;13(2):161–165.

**Key Words:** genetic test, genomic test, horizon scanning, surveillance, database

Many sectors—scientific, medical, commercial, and regulatory—have an interest in the development and evaluation of new, health-related genomic tests. As a result of public and private investment, many new genomic tests are being offered to health care providers and, in some cases, directly to consumers. This activity is decentralized and dynamic, involving a complex network of relationships among government, academic, and commercial entities. Consequently, it is not easy to determine which genomic tests are in development, currently available, or discontinued. Information about the tests is mostly scattered and often incomplete.

Efforts are underway to consolidate relevant information and make it accessible to stakeholders and the public. Since 1993,

information about genetic tests for mostly rare, single-gene and chromosomal disorders has been available online from GeneTests.<sup>1,2</sup> Several endeavors—such as the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) project<sup>3</sup> and the Secretary's Advisory Committee on Heritable Disorders in Newborns and Children<sup>4</sup>—concentrate on systematic reviews of evidence for the validity and utility of existing genetic tests.

Horizon scans complement these retrospective review processes by identifying new, health care-related genetic tests before they become widely used. For example, the Agency for Healthcare Research and Quality has published annual horizon scans since 2006, with a focus alternating between cancer and other conditions. The 2009 report reviewed selected pharmacogenetic tests and the 2010 report updates information on genetic tests for noncancer conditions.<sup>5,6</sup> An updated list of companies offering genetic tests directly to consumers is available from the Genetics and Public Policy Center.<sup>7</sup> Most recently, the National Institutes of Health has announced plans to develop a new, voluntary Genetic Testing Registry at National Center for Biotechnology.<sup>8,9</sup>

Since 2001, the Office of Public Health Genomics (OPHG) at the Centers for Disease Control and Prevention has actively monitored online news and scientific resources for information on genomic research and its potential implications for public health. OPHG posts an online weekly update, which contains links to news stories and scientific articles (<http://www.cdc.gov/genomics/update/current.htm>). In addition, OPHG staff curates an online database of scientific publications on genetics and disease in human populations, which is populated by a weekly sweep of PubMed.<sup>10</sup> Since 2005, the OPHG-sponsored EGAPP project staff has also conducted periodic online searches for genomic tests that are in transition from research to clinical practice.

In mid-2009, we initiated a review of these online search procedures to develop a more systematic and comprehensive horizon scanning process to identify new, health-related genetic and genomic tests. We describe the development of this process and evaluate the initial results, collected from mid-October 2009 through June 2010. Finally, we consider how horizon scanning can contribute to future evaluation of genomic testing.

### METHODS

We focused our review on tests that analyzed (1) genetic variants (e.g., single nucleotide polymorphisms or copy number variants), singly or in combination; (2) gene expression; or (3) protein biomarkers (excluding single antibodies). We also sought to identify relevant tests based on other methods (such as fluorescent in situ hybridization, immunohistochemistry, or mass spectrometry) and assays of other molecular analytes (such as epigenetic markers and micro-RNAs), either alone or in combination. Although we were primarily interested in tests related to screening, diagnosis, management, and prediction of adult-onset and chronic diseases, we did not exclude newly described tests for inherited disorders. We did not attempt a

From the <sup>1</sup>McKing Consulting Corporation; <sup>2</sup>Office of Public Health Genomics, OSELS, Centers for Disease Control and Prevention; and <sup>3</sup>Cadence Group, Atlanta, Georgia.

Marta Gwinn, MD, MPH, OPHG/OSELS/CDC, 1600 Clifton Road, mail stop E-61, Atlanta, GA 30333. E-mail: mgwinn@cdc.gov.

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

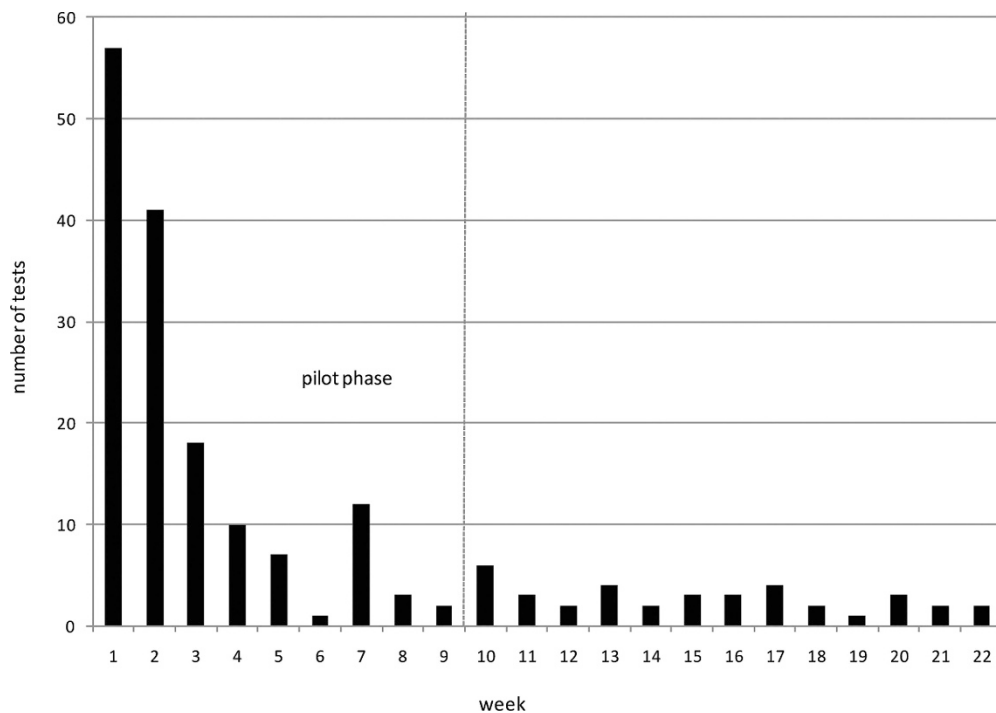
Disclosure: The authors declare no conflict of interest.

Submitted for publication August 10, 2010.

Accepted for publication October 9, 2010.

Published online ahead of print January 12, 2011.

DOI: 10.1097/GIM.0b013e3182011661



**Fig. 1.** Number of newly identified genomic tests by week, GAPP Finder, October 12, 2009, to June 30, 2010.

retrospective survey of previously developed tests, including some that are already in widespread use.

We used Google Alerts (<http://www.google.com/alerts>) to search for Web pages, newspaper articles, and blogs that might identify a new, health-related genomic test. We first composed a set of Google Alert queries by combining the words test or testing with a word from the following list: gene, genetic, DNA, molecular, expression, or proteomic. We did not use the term genomic because, although it describes these tests as a group, it is rarely used to describe specific tests. We then created a set of selected three-word combinations by adding a term from a second list: cancer, cardiovascular disease, and diabetes. Results from the two- and three-word queries were returned daily by e-mail and compiled weekly. Results were reviewed in two stages: first, for general relevance and to remove duplicates, and next, to identify relevant tests that had not been found previously.

From July 27 through October 5, 2009, we monitored the relative yields of our queries and eliminated those that returned no useful or duplicate results. Although disease-specific queries for cancer, cardiovascular disease, and diabetes yielded few unique items, we retained them because of their potential public health importance. The final set of queries included the following search terms: cancer gene test, cancer molecular test, cancer test, cancer test human, cardiovascular disease genetic test, diabetes genetic test, gene expression, genetic test, genetic testing, and proteomic test.

During a pilot phase (October 12, 2009, through March 31, 2010), the final set of queries was run daily and compiled weekly to create a list of online mentions of health-related genomic tests in Web pages, newspaper articles, and blogs. After eliminating duplicates and any tests that were not health-related (e.g., tests used for genealogy research or social networking), we manually extracted relevant information about each test. We adapted EGAPP definitions to describe four

characteristics that, when combined, could be used to define a unique test: the disease or condition (e.g., irritable bowel syndrome), the type of test (e.g., expression of 32 genes in tumor tissue), the target population (e.g., women with a family history of ovarian cancer), and the intended use (e.g., to determine appropriate warfarin dosing).<sup>3</sup> In addition, we noted the source query, the initial Web link, and trade and company names.

For future searchability, we further classified tests according to their proposed applications (screening, diagnosis, prognosis, pharmacogenomics, risk prediction, or other) and status (research and development, commercially available, marketed directly to consumers, Food and Drug Administration license claimed, and unknown). By the end of March 2010, we had finalized search, review, and curation procedures and developed an online database. Since then, daily horizon scanning, weekly review, and curation have continued. We analyzed data collected through June 2010 to describe the numbers and characteristics of genetic tests identified by this process.

## RESULTS

The final set of Google Alert queries returned approximately 400 items each week (average 433, range 401–493 in June 2010). These often included large numbers of duplicates, which typically occurred when a scientific publication or news release describing a new test was covered by many different news outlets and blogs. We attempted to identify and document the primary source (often a test developer's or licensee's Web site) and excluded the others. The time required to screen Google Alert results varied from week to week, ranging from one to several hours; finding and extracting relevant information and entering it into the database required approximately 30 minutes per test.

We identified 151 tests during the pilot phase (October 12, 2009, to March 31, 2010) and 37 more during the first 3 months

**Table 1** Distribution of new genomic tests by disease or condition, October 12, 2009, to June 30, 2010

	Number	Percent
Cancer	122	65
Colorectal	25	
Lung	21	
Breast	17	
Prostate	13	
Ovary	6	
Leukemia/lymphoma	6	
Skin	6	
Bladder	5	
Mesothelioma	3	
Pancreas	3	
Gastric	2	
Other	4	
Multiple sites	3	
Not specified	8	
Cardiovascular	19	10
Coronary heart disease	7	
Clotting disorders	6	
Cardiac conduction disorders	4	
Hypertension	1	
Not specified	1	
Gastrointestinal disorders	5	3
Neurodegenerative disorders	4	2
Developmental disorders/birth defects	4	2
Age-related macular degeneration	4	2
Autoimmune disorders	2	2
Diabetes/metabolic syndrome	2	1
Psychiatric disorders	2	1
Hereditary disorders	11	6
Other	13	
Total	188	

of routine data collection, from April through June, 2010 (Fig. 1). The large numbers of tests identified during the early weeks of the pilot phase reflected a backlog of recently introduced or existing tests with ongoing mentions in blogs and news items. After capturing information about these tests, subsequent mentions were treated as duplicate information. After Week 12, the number of new tests detected averaged approximately two to three per week.

The distribution of tests by disease or condition is summarized in Table 1. Overall, nearly two thirds of tests (122/188, 65%) were related to cancer. Tests for colorectal, lung, and breast cancers accounted for about half of the cancer-related tests (63/122, 52%) and about one third of tests overall (63/188, 33%). The next largest group, tests related to cardiovascular diseases, comprised 19 tests (10%). Hereditary disorders accounted for 6% of all tests; the “Other” category (7%) included tests for conditions such as aging, early menopause, and weight management.

More than one fourth (51/188) of the tests we found were proposed for multiple uses, e.g., for detecting and classifying cancer (diagnosis) and for predicting survival (prognosis). After excluding 2 tests with 3 proposed uses, we summarized 235 proposed uses for the remaining 186 tests in Table 2. Overall, diagnostic tests constituted the largest category (66/235, 28%), followed by prognostic tests (54/235, 23%) and pharmacogenetic tests (41/235, 18%). Tests for risk prediction (i.e., for estimating the likelihood of future disease in currently asymptomatic persons) accounted for only 11% of the total. Although nearly half of all tests (88/188, including 2 tests marketed directly to consumers) were commercially available, Food and Drug Administration licensure was claimed for only 12 (6%) tests. There were 82 (44%) tests still in development and 6 for which no information about availability was provided.

The information we extracted from the Internet sources for the 188 tests included in this analysis, in addition to those identified in subsequent weeks, is available online in a searchable database, the GAPP Finder (<http://www.gappnet.org>, Fig. 2).

## DISCUSSION

We used Google Alerts to detect newly developed or launched genomic tests. The Internet’s widespread use, and frequently updated entries from test developers eager for public notice, helped us capture new announcements and extract relevant scientific information. We acknowledge that many challenges to identifying and validating new genomic tests remain.

**Table 2** Numbers of genomic tests by combinations of proposed uses, GAPP Finder, October 12, 2009, to June 30, 2010

Proposed use	Screening	Diagnostic	Prognostic	Pharmacogenetic	Risk prediction	Other
Screening	<b>23</b>	14	1	1	1	0
Diagnostic	14	<b>37</b>	12	0	1	2
Prognostic	1	12	<b>27</b>	11	3	0
Pharmacogenetic	1	0	11	<b>26</b>	2	1
Risk prediction	1	1	3	2	<b>20</b>	0
Other	0	2	0	1	0	<b>4</b>
Total	40	66	54	41	27	7

Table cells indicate the numbers of tests with each combination of proposed uses. Bolded values indicate the numbers of tests with a single proposed use. Column totals indicate the total numbers of tests proposed for each use. For example, of the 40 tests proposed for use in screening, 14 were also proposed for use in diagnosis, 1 for prognosis, 1 for pharmacogenetics, and 1 for risk prediction; 23 screening tests were proposed for screening only. The table excludes 2 tests with 3 proposed uses each.

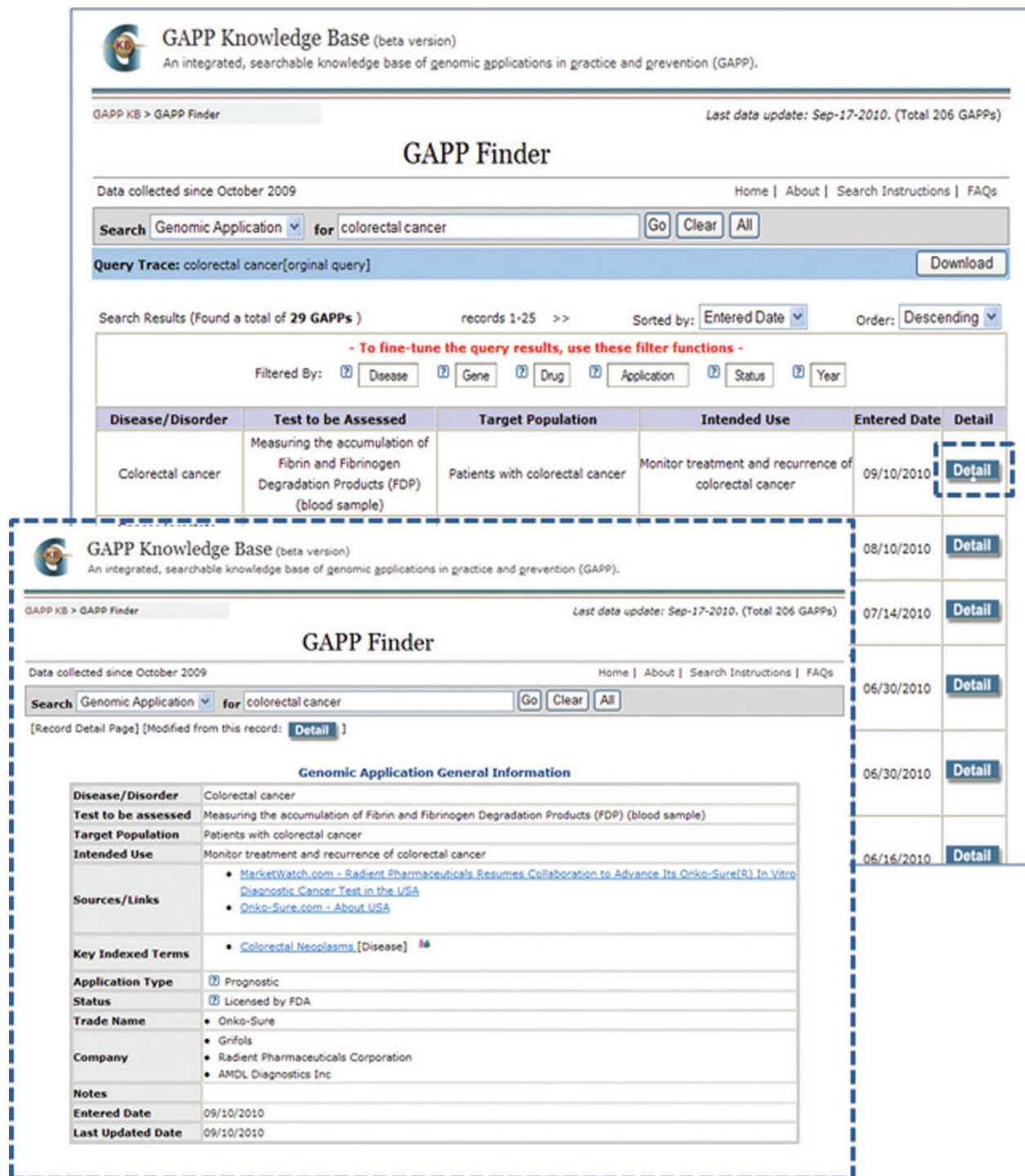


Fig. 2. Example of search results for colorectal cancer, GAPP Finder. Detail page for each test includes additional information, including trade names, Web links, and indexing terms. Trade names displayed for demonstration only.

For most new tests, the information available online is incomplete. We encountered missing data for test analyses, indications, and target populations. In some cases, test developers may not have shared what they consider to be proprietary information. Tests described as “genomic profiles,” which may include hundreds of different markers to predict risks for multiple conditions, present particular challenges. Information on the

numbers, types, and combinations of markers tested in such profiles is often not presented; changes may not be reported, and we did not make special efforts to find them.<sup>11</sup> In addition, we learned that it could be difficult to accurately identify tests when their names were changed or they were licensed to multiple companies.

After the pilot phase, when a large backlog of tests was reviewed, the amount of time required to identify new tests and

enter them into the database averaged less than one-half day per week. Recently, we changed the Google queries to run weekly; during a 4-week period, this approach returned approximately one third as many items as the same queries run daily and compiled weekly (125 vs. 400 per week, on average), without any loss in sensitivity. Although additional refinement of the Google queries could further reduce the amount of time required to identify new tests, growth in the number of new tests will increase the time required to extract and enter the data into the GAPP Finder.

Our approach to horizon scanning aimed to maximize timeliness and transparency by finding and sharing information online; we saw this as the best way to complement other efforts to monitor and evaluate genomic tests. We are currently sharing data with National Center for Biotechnology to support their development of the proposed Genetic Test Registry; at their request, we have standardized some data elements to enhance exchangeability. In the future, active online surveillance for genomic test information could complement registry data, which will be reported by test developers on a voluntary basis. Our online database, the GAPP Finder, is currently accessible and embedded in the GAPP Knowledge Base, a suite of Web-based products that promote objective synthesis and timely dissemination of information on genome-based tests and technologies. We plan to compare GAPP Finder data prospectively with the results of other horizon scans for genetic tests, including both general scans (such as those conducted by Agency for Health care Research and Quality) and scans targeted to specific diseases or test types.

It seems unlikely that any single database will be able to capture all of the information essential for evaluating genomic tests. The field is growing and evolving rapidly and not always predictably. Different types of information—gathered by different processes, from different sources—are needed to evaluate the analytic and clinical validity and utility of genomic tests. Although test developers may be able to report analytic sensitivity and specificity to a voluntary registry, they will rarely have access to the epidemiologic data needed to evaluate clinical validity.<sup>12</sup> Evidence for clinical utility is even more diverse and will continue to accrue after a test is in widespread use. We used information that is freely available online to establish our surveillance system and will share the results freely via an

online resource, the GAPP Finder. By allowing easy navigation to other online resources on genomic testing, including the newly launched knowledge collection *PLoS currents: evidence on genomic tests*,<sup>13</sup> we hope to enhance the value of our horizon scanning to a wide group of stakeholders.<sup>14</sup>

## REFERENCES

1. Pagon RA, Tarczy-Hornoch P, Baskin PK, et al. GeneTests-GeneClinics: genetic testing information for a growing audience. *Hum Mutat* 2002;19:501–509.
2. GeneTests: Medical Genetics Information Resource (database online). Copyright, University of Washington, Seattle. 1993–2010. Available at: <http://www.genetests.org>. Accessed September 21, 2010.
3. 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;11:3–14.
4. 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;12:153–159.
5. Raman G, Trikalinos TA, Zintzaras E, et al. Reviews of selected pharmacogenetic tests for non-cancer and cancer conditions. Agency for Healthcare Research and Quality Technology Assessment Report, 2009. Available at: <http://www.cms.gov/determinationprocess/downloads/id61TA.pdf>. Accessed September 21, 2010.
6. Raman G, Wallace B, Chung M, et al. Update on genetic tests for non-cancer diseases/conditions: a horizon scan. Agency for Healthcare Research and Quality Assessment Report, 2010. Available at: <http://www.cms.gov/determinationprocess/downloads/id49ta2.pdf>. Accessed September 21, 2010.
7. Genetics and Public Policy Center. DTC genetic testing companies. Compiled May 28, 2010. Available at: <http://www.dnapolicy.org/resources/Alphabetized-DTCGeneticTestingCompanies.pdf>. Accessed September 21, 2010.
8. Kuehn BM. NIH launching genetic test registry. *JAMA* 2010;303:1685.
9. Genetic Testing Registry. Available at: <http://www.ncbi.nlm.nih.gov/gtr/>. Accessed September 21, 2010.
10. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. *Nat Genet* 2008;40:124–125.
11. Janssens AC, Gwinn M, Bradley LA, Oostra BA, van Duijn CM, Khoury MJ. A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. *Am J Hum Genet* 2008;82:593–599.
12. Khoury MJ, Reyes M, Gwinn M, Feero WG. A genetic test registry: bringing credible and actionable data together. *Public Health Genomics* 2010;13:360–361.
13. Gwinn M, Dotson WD, Khoury MJ. PLoS currents: evidence on genomic tests—at the crossroads of translation [Internet]. Version 8. PLoS currents: evidence on genomic tests. 2010 Aug 31 [revised 2010 Sep 2]. Available from: <http://knol.google.com/k/marta-gwinn/plos-currents-evidence-on-genomic-tests/3f696lo68dgd/1>. Accessed September 21, 2010.
14. Khoury MJ, Feero WG, Reyes M, et al. The genomic applications in practice and prevention network. *Genet Med* 2009;11:488–494.