



July 16, 2010

Dr. Cathy Fomous
NIH Office of Biotechnology Activities
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892
phone: 301-496-9838
fax: 301-496-9839
email: CFomous@od.nih.gov.

Anne Wojcicki

23andMe, Inc.
1390 Shorebird Way
Mountain View, CA 94043
650.938.6300 phone
650.938.6302 fax

anne@23andme.com
joanna@23andme.com
www.23andme.com

Dear Dr. Fomous and colleagues,

We appreciate your soliciting input and feedback on the NIH's plan to develop the Genetic Testing Registry (GTR). We understand that you are seeking input regarding both the types of tests to include within the GTR and the data elements to include for each test. We also understand that information regarding the analytical validity, clinical validity, personal utility, and clinical utility of a test, whenever available, will make the registry most useful to a broad range of stakeholders. The following feedback corresponds to the numbered items in the <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-10-101.html> document:

#1. *(Are there any types of genetic tests that should not be included in the GTR?)*

Although genetic tests used for genealogical and ancestry discovery could be included within the GTR, results of these tests typically take on very different forms from results of tests associated with diseases and health conditions. Genetic tests used for genealogical and ancestry purposes would therefore require an entirely separate design process. Given the health and disease focus of the registry, the genealogical and ancestry tests seem to be relatively low priorities. For similar reasons, tests associated with non-disease conditions (e.g., eye color, hair curl, ear wax) seem to be low priorities.

#2. *(What are the potential uses of the GTR ...?)*

(2, 3) The GTR would help patients, health care providers, and other consumers of genetic testing understand what information they would and would not obtain from a given genetic test, and the potential applications of that information. The GTR would also help health care providers, patients, and other consumers compare genetic tests as they make decisions regarding which test to obtain.

(6) The GTR would enable genetic testing entities to provide information regarding their tests in a standardized format, facilitating transparency.

#3. (What data elements are critical to include ...?)

- **Scope of Registry:** According to the definition of "genetic test" in the RFI (section II), the GTR will need to accommodate information regarding genotype-based, sequencing-based, and copy number variant (CNV) based tests (including comparative genomic hybridization (CGH)), among others. Ideally the GTR will be flexible enough to handle the data elements associated with each of these categories. Indeed, with whole genome sequencing on the horizon, the GTR would be most valuable if it can handle tests that involve entire genomes. On a separate topic, we understand that the scope has to be limited to ensure successful execution but note that the current definition does not cover tests, such as those measuring gene expression, that may not reveal mutations, but that have related uses. A registry that accommodated such expression tests would be valuable.
- **Nomenclature:** All stakeholders will benefit if the GTR is prepared to handle varying nomenclature and present it in a standardized way to minimize the chance that, for example, synonyms for gene names and mutations, and DNA strand orientation confuse consumers into thinking that two identical tests cover different things. Furthermore, mutations should follow the recommendations for nomenclature of variations outlined by HGVS (<http://www.hgvs.org/rec.html>).
- **Quantitative Trait Loci:** To be most broadly valuable, the design of the GTR would take into account tests for quantitative trait loci (multi-marker genotyping tests that can be used to produce a numerical risk estimate).
- **Genes and Environment:** In the future more genetic tests will be components of a model of risk prediction that also includes non-genetic risk factors. Accommodation of information on non-genetic factors would enable these more complex, and in many cases more highly predictive, tests to be included in the GTR. The GAIL model for the risk of breast cancer provides one example; recent studies have examined the predictive power of models that include genetic factors plus the non-genetic factors of the standard GAIL model.
- **Variants of Unknown Significance:** Certain tests, such as those involving DNA sequencing or CNV determination, are capable of revealing variants that are either rare or unique to an individual (private). In some cases, these variants may be readily identifiable as having a large and/or causal effect in disease; in other cases the impact of the variant may not be as clear (e.g., "variants of unknown significance," (VOUS)). Consumers of genetic tests will find the GTR most helpful if the GTR is capable of communicating that some tests may produce results where accuracy cannot be evaluated or health significance is unknown.
- **Multiple Tests for One Disease or Condition:** In order to accommodate information on tests for diseases such as cystic fibrosis (CF) that have hundreds or thousands of known mutations/VOUSs, the GTR will need to include options for multiple tests, each of which could include a different number of the relevant variants. For instance, some CF genetic tests include just one mutation (deltaF508), while others include dozens of mutations, and yet others include sequence data for the entire gene.

- **RSIDs, Stranding and other Critical Information:** Researchers will find the GTR most useful if the following information is included for a test, when appropriate: RSIDs when SNPs are reported; reference to the relevant Human Genome Build (e.g., Build 37) when sequence positions are reported without an RSID; stranding (plus or minus) when alleles are named.
- **Analytical and Clinical Validity:** Increasingly the scientific literature will include results from multiple studies regarding the analytical validity, clinical validity, or clinical utility, of a particular test. Users of the test and researchers, among others, would benefit from seeing results from multiple studies, where applicable.
- **Heritability:** For some diseases, an estimate of heritability may be helpful to consumers, health care providers, and researchers.
- **Risk Modifiers:** For some diseases, known risk modifiers not explicitly tested for, whether genetic, environmental, or both, would provide valuable context.
- **Standards for Reporting:** Organizations and companies have established and made public standards for providing particular tests. For instance, 23andMe requires, for its "Established Reports," that a genetic association have been found via at least two studies with at least 750 cases and 750 controls each. Information on whether a test meets a set of publicly stated standards would allow health care providers, patients and other consumers to better evaluate the test.

#5. (*What is the best way to distinguish between data fields left blank because of an absence of data/evidence and those left blank for other reasons?*)

The distinction between data fields left blank because of absence of data and those left blank for other reasons seems critical. Users of the GTR who can quickly identify whether, for instance, a clinical validity study has been conducted, will be in a better position to make a decision regarding whether to use the test or recommend that another person use the test. In addition, the distinction could facilitate research by making gaps in knowledge more obvious. Since data submission is voluntary, the GTR need not require that all data fields applicable to a given test type be completed. However, the reason a given field is incomplete would be valuable to consumers and other users of the GTR. For example, the responses "Not Provided," "Not Applicable," "No Evidence Available" could each have a different meaning and would therefore be valid entries.

#6. (*... which of the following data elements should be included in the GTR?*)

It seems reasonable to include all the data elements mentioned in section III.6., recognizing that in some cases data may not be available for all elements for all tests.

#6g. (*Limitations of the test*)

It may be helpful to have an explicit "Subpopulations for which test has been validated" data element since this question will come up for many tests.

#6m. (*Performance characteristics*)

The characteristics listed in the RFI are geared toward single-marker DNA tests. Additional criteria (e.g., Area Under the Curve (AUC)) should be accommodated to cover the case of multiple markers combined via an algorithm -- this suggestion applies to both DNA tests and expression tests. Criteria specific to single-marker tests (e.g., clinical sensitivity/specificity) need not be required of multi-marker tests in which an algorithm produces a numerical result. Note, however, that given a threshold model, sensitivity and specificity could be provided for numerical multi-marker tests.

#6n. (*Clinical validity*)

For tests that use multiple markers, it will become increasingly difficult to pull together research cohorts that are large enough to independently validate a risk prediction algorithm. Once the number of markers used in a test reaches the point that the data become unique to an individual, the only way to produce a risk estimate or other test score will be to use modeling. In order to accommodate such tests, the GTR could allow estimates of accuracy to be specified as either "Empirical" or "Simulated". Additionally, some tests may be assembled from information available in the scientific literature (i.e., without an independently performed validation study). These estimates could be specified as "Literature-Derived."

#7. (*What types of information might be difficult for test providers to submit, and why?*)

Test providers are unlikely to be willing to submit confidential commercial or trade secret information.

#10. (*As the GTR is being designed, what are the important processes to consider to make the submission of data as easy as possible for the data provider ...?*)

A master file with data common to multiple tests would facilitate submission to the GTR.

#11. (*Which potential benefits and risks would be most likely to affect the decisions of researchers, test developers, and manufacturers on whether to submit data to the GTR ...?*)

If the GTR is seen as a trustworthy source of information, it could benefit test developers and providers whose information is made available through this source. It would also enable these developers and providers to be transparent about their tests.

(what factors will best encourage submission of complete and accurate data?)

The easier the submission process, the more likely groups will submit complete and accurate data. Extensive user testing during development would very likely improve submission rates because it would reveal the need for additional functionality that would further encourage submission. Given the complexity of the data to be submitted, flexibility will be essential so that data that don't quite fit the usual format can be entered somewhere. Specifically, a flexible submission interface that allows fields to be left blank if inapplicable and that allows for additional information to be included, even if not explicitly requested, could increase submission rates.

#12. *(What are the most effective methods to ensure continued stakeholder input into the maintenance of the GTR?)*

Regular reports of usage of the GTR by various stakeholders could encourage continued input.

#13. *(For what purpose(s) would you use the Registry to support your professional efforts?)*

23andMe could use the GTR in a variety of ways, including: (a) to provide detailed information to prospective 23andMe users and research partners and (b) to find out how other entities report gene or mutation names in order to increase consistency across reporting.

#14. *(Are there any other issues that NIH should consider in the development of the GTR?)*

We have one additional suggestion:

Sample entries for tests, including non-genetic tests, that are already widely accepted in the medical profession, would help users better understand the information provided via the GTR. For example, the sensitivity/specificity and PPV/NPV for mammograms and/or HIV tests could be provided as a baseline comparison for single-factor tests that health care providers, patients, and other consumers are already familiar with. Blood pressure and cholesterol levels are quantitative trait loci that could be used as example formats for tests that return a numerical estimate that is stratified into low-, average-, and high-risk categories. Such examples not only would put users into a better position to use the GTR, but also could make the GTR a valuable educational resource.

We look forward to the NIH's Genetic Testing Registry becoming a resource that is valuable to a broad range of stakeholders.

Sincerely,



Joanna Mountain, Senior Director of Research
on behalf of
Anne Wojcicki, Co-founder and President