



## Disclosing pathogenic genetic variants to research participants: Quantifying an emerging ethical responsibility

Christopher A Cassa, Sarah K Savage, Patrick L Taylor, et al.

*Genome Res.* published online December 6, 2011

Access the most recent version at doi:[10.1101/gr.127845.111](https://doi.org/10.1101/gr.127845.111)

---

**Supplemental Material** <http://genome.cshlp.org/content/suppl/2011/12/06/gr.127845.111.DC1.html>

**P<P** Published online December 6, 2011 in advance of the print journal.

**Accepted Preprint** Peer-reviewed and accepted for publication but not copyedited or typeset; preprint is likely to differ from the final, published version.

**Email alerting service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

---

---

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

---

To subscribe to *Genome Research* go to:  
<http://genome.cshlp.org/subscriptions>

---

## *Disclosing pathogenic genetic variants to research participants: Quantifying an emerging ethical responsibility*

Christopher A. Cassa PhD, Sarah K. Savage MS, Patrick L. Taylor JD, Robert C. Green MD MPH, Amy L. McGuire JD PhD, and Kenneth D. Mandl MD MPH

Children's Hospital Informatics Program, Children's Hospital Boston, Center for Biomedical Informatics, Harvard Medical School, Center for Medical Ethics and Health Policy, Baylor College of Medicine, Division of Genetics, Brigham and Women's Hospital, Partners Healthcare Center for Personalized Genetic Medicine, and the Petrie-Flom Center of Harvard Law School

Corresponding Author: Christopher A. Cassa PhD, 1 Autumn St #721 Informatics, Boston, MA 02215, [cassa@alum.mit.edu](mailto:cassa@alum.mit.edu)

Running Title: Researchers may need to share many variants with subjects

### **Abstract**

There is an emerging consensus that when investigators obtain genomic data from research participants, they may incur an ethical responsibility to inform at-risk individuals about clinically significant variants discovered during the course of their research. With whole exome sequencing becoming commonplace and the falling costs of full genome sequencing, there will be an increasingly large number of variants identified in research participants that may be of sufficient clinical relevance to share. An explicit approach to triaging and communicating these results has yet to be developed, and even the magnitude of the task is uncertain. To develop an estimate of the number of variants that might qualify for disclosure, we apply recently published recommendations for return of results to a defined and representative set of variants and then extrapolate these estimates to genome scale. We find that the total number of variants meeting the threshold for recommended disclosure ranges from 3,955-12,579 (3.79-12.06%, 95% CI) in the most conservative estimate to 6,998-17,189 (6.69-16.48%, 95% CI) in an estimate including variants with variable disease expressivity. Additionally, if the growth rate from the previous four years continues, we estimate that the total number of disease-associated variants will grow 37% over the next four years.

## Introduction

In genomics, there is an emerging consensus that when investigators obtain genetic data from research participants, they may incur an ethical responsibility to inform at-risk individuals about clinically significant variations discovered during the course of their research (Bookman et al. 2006; Wolf et al. 2008a; Wolf et al. 2008b; Fabsitz et al. 2010). Perhaps the largest obstacle to reviewing and communicating incidental findings in genomics research is the sheer magnitude of the task. When performing whole genome sequencing, we expect to observe hundreds of variants in each individual participant from over 100,000 variants that have been previously associated with disease (Ashley et al. 2010). The proportion of these that would meet criteria for disclosure has not previously been considered.

Because of the complexities inherent in exposing participants to predictive genetic information discovered outside of the clinical context (Kohane et al. 2006; Meltzer 2006; Wolf et al. 2008a; Johnson et al. 2010), and concerns about blurring the lines between research and clinical care (Caulfield et al. 2008), there has been extensive debate about whether it is appropriate to communicate results derived from genetic research to study participants (Fernandez et al. 2003; Bookman et al. 2006; Fernandez and Weijer 2006; MacNeil and Fernandez 2006; Meltzer 2006; Kozanczyn et al. 2007; Wolf et al. 2008a; Wolf et al. 2008b; Fabsitz et al. 2010; 2010). Several bodies have presented recommendations for the return of individual genetic results to participants, including the National Bioethics Advisory Commission (NBAC), the Centers for Disease Control and Prevention (CDC), and the National Heart, Lung, and Blood Institute (NHLBI) (Beskow et al. 2001; White and Gamm 2002; Bookman et al. 2006; Teutsch et al. 2009; Fabsitz et al. 2010) (Table 1). Although these recommendations vary, they all recognize an obligation to return at least some research findings. While there is no set of universally accepted standards for disclosure, there is some similarity among their criteria for returning individual research results to participants. Each of these groups places priority on the scientific validity of the reported association, the clinical significance of the associated phenotype, and the availability of beneficial medical interventions.

In support of translational medicine, hundreds of thousands of participants have provided samples to research biorepositories (Kohane et al. 2007; McGuire 2008; Roden et al. 2008; Blow 2009). Inevitably, analyses of these samples will identify both novel and previously-discovered variants that confer disease risk. Considering that there are already over 100,000 genetic variants cited in the medical literature (Hindorff LA), how many of these variants will meet an ethical obligation for disclosure to participants?

To answer this question we estimated the proportion of known genomic variants that would meet the expectation to report to participants based on published guidelines, using the most recently-published recommendations (Fabsitz et al. 2010) as a model. These recommendations were selected because they incorporate most of the commonly articulated criteria. The NHLBI 2010 working group concluded that individual genetic results should (with conditions) be offered to

study participants in a timely manner if they meet all of the following criteria (Recommendation 1):

- 1) The genetic finding has important health implications for the participant and the associated risks are established and substantial.
- 2) The genetic finding is actionable; that is, there are established therapeutic or preventive interventions or other available actions that have the potential to change the clinical course of the disease.
- 3) The test is analytically valid and complies with all applicable laws.
- 4) During the informed consent process or subsequently, the study participant has opted to receive his/her individual genetic results.

We selected a representative sample of disease-associated genetic variants from the scientific literature and reviewed each variant to assess characteristics that would help determine whether investigators would be expected to report it according to these 2010 recommendations. We then extrapolated these results to the genome scale, estimating the total number of variants that would satisfy these disclosure criteria.

## Results

An expert committee of three certified genetic counselors reviewed and scored a representative sample of 160 disease-associated variants that were randomly sampled from databases that curate genetic findings from the scientific literature. Each variant was scored on multiple characteristics, including the validity of its disease association, the severity of the associated disease with and without treatment, and the potential to improve medical outcome with intervention. We then applied disclosure criteria from the 2010 Guidelines (Fabsitz et al. 2010) to the sample under the strictest possible interpretation and then again with an interpretation that allowed for greater variability of disease expression. Finally, we extrapolated these results to the genome scale, estimating the total number of variants that would satisfy these disclosure criteria (Table 2).

Under the strictest interpretation, which required possible expression of severe disease, we identified that 6.9% of the variants reviewed would meet criteria for disclosure to research participants. An additional 3.8% of all sampled variants are associated with variable disease expression or uncertain risk for severe disease and would meet the criteria for disclosure if disease expression were at least “severe” in some cases. With these additional variants included, 10.6% met the criteria for disclosure.

Among the 104,304 disease-associated variants genome-wide from published research considered in our study (a sample of 160 randomly sampled from curated pathogenic variant databases: HGMD, OMIM, and NHGRI Catalog of GWAS) (Hindorff LA ; Amberger et al. 2009; Stenson et al. 2009), we conservatively estimate that investigators could be urged to share 7,171 [3,955-12,579, 95%CI] of those variants

with participants. If variants with variable disease expression are included (which meet all other required criteria and also may result in severe disease,) then 11,082 [6,998-17,189, 95%CI] variants would be shared, genome-wide (Table 3).

We further project an increase in the number of disease-associated variants that will be identified in publications. Using four years of historical disease-associated variant data from HGMD and the NHGRI Catalog of Published GWAS (98% of our variant data set) at a quarter-year resolution, we identified an average growth rate of 10,437 variants per annum ( $R^2=0.9977$ ) in a conservative linear growth scenario and increasing growth rates in more aggressive scenarios (Figure 1).

It is expected that by 2015, there will be over 150,000 published disease-associated variants using a conservative linear estimate, and over 190,000 if extrapolating using an exponential growth model. Estimates are presented for growth in the number of variants that would meet the recommendation to share with participants if a conservative, steady growth (linear model,  $R^2=0.9976$ ) is selected (Table 4), when extrapolating the observed 10.6% variant disclosure rate from our sample using Fabsitz et al. Recommendation 1.

Additionally, to evaluate the number of variants we might encounter in research subjects, we analyzed 36 whole genome sequences of asymptomatic individuals that were publicly available from Complete Genomics (Drmanac et al. 2010; Genomics 2011). We identified that each of these samples carried an average of 2,120 substitution variants from our study knowledge base (Table 5). This demonstrates that a substantial number of previously identified variants are likely to be observed in the whole genome sequence analysis of asymptomatic individuals, although it is not yet possible to accurately estimate the actual number of these variants that will be reportable to each individual.

## Discussion

This is the first study to estimate the number of published disease-associated variants that met the criteria for disclosure to research participants, according to the recommendations of a recent consensus group (Fabsitz et al. 2010). Extrapolating to genome scale, investigators following these guidelines may be responsible for disclosing over 11,000 variants today and over 16,000 variants by the year 2015.

The return of incidental research findings has been the subject of debate for many years. Still contentious and unresolved, for example, is the optimal approach to disclosing suspicious incidental findings to subjects of neuroimaging studies (Illes et al. 2004; Illes et al. 2010; Palmour et al. 2011). Finding a lesion on the computed tomography scan of an asymptomatic research subject creates a complex dilemma; the meaning of the lesion may be unclear when the prior probability of disease is low (Sadatsafavi et al. 2010; Issues 2011). In genomics, the core ethical issues are similar, but the information management task becomes daunting at the genome scale.

Without new tools for investigators and institutions, emerging requirements for the return of results in genomics may become rapidly unmanageable in the setting of a growing corpus of known disease-associated variants. The scientific review process to measure the validity and communicability criteria for each of these variants would be protracted, and even once the estimated 11,000-16,000 qualified variants are identified, there is the process of evaluating and prioritizing those identified in each participant. This is a drastically larger burden on investigators than the review of a single neuroimaging study per participant – the impact on genomics research could be quite serious. Are we setting the stage for participants to be disappointed and for researchers to become legally liable for something they cannot realistically accomplish?

While published policy recommendations are certainly well intentioned, they are difficult to apply when such a large number of potential variants are worthy of communication to participants. We need more realistic standards that balance the ethical arguments in favor of disclosing results with the reality of what is feasible to review and communicate. Investigators and leaders of biorepositories will not be able to meet this potential ethical responsibility without substantial, evolving knowledge bases and enhanced processes. Additional questions from the current guidelines remain: Is CLIA certification really equivalent to analytic validity? What is the appropriate predictive risk cutoff for having clinical significance or ‘important health implications’? Does a GWAS-identified allele that increases relative risk for a rare disease by 10% warrant communication to participants?

Our estimates are based on an ambitious approach of investigating previously identified disease-associated variants that have been described in the scientific literature. To execute return of results at this level, researchers and research subjects would necessarily rely on yet-to-be developed tools for identifying and presenting variants of importance across the genome. In the mean time, as these genome scale decision and risk tools are developed, there are simple approaches to reduce the burden inherent in analyzing incidental findings in the clinical interpretation of whole genome sequences. One is to limit reportable findings to those actually discovered during the normal course of research, likely including only a small set of genes. Alternatively, when the scope of the primary research is genome-wide, it may be more appropriate to ask investigators to check for a panel of well-known variants that meet a high standard of clinical significance and actionability, throughout the genome. Unfortunately, no such list yet exists.

The National Institutes of Health (NIH) is calling for empirically informed guidelines for the return of results ((NHGRI) 2011a; (NHGRI) 2011b; (NHGRI) 2011c); in parallel there should be support for new communication strategies and dynamic data sources to support this disclosure. While there are data sources emerging that measure the clinical and syntactic validity of previously identified variants (Khoury et al. 2009; Genomes 2010; Tong et al. 2011), there is presently no knowledge base that includes the necessary information to reach disclosure decisions. One way to prospectively build such a database would be to create an obligation for

investigators to report novel variant associations identified in research studies. However, many investigators are not necessarily qualified to create clinical annotations that assess risk/benefit tradeoffs and other reportability criteria.

Further, how will the participant's clinical context (Beskow and Burke) and preferences be taken into account? As the science and practice of full genome interpretation matures (Ormond et al. ; Ashley et al.), the next challenge becomes personalization of the communication strategy for very large numbers of variants, considering risks and benefits for each individual (Fabsitz et al. 2010; Kohane and Taylor 2010) that include the clinical validity and utility of each variant (Holtzman et al. 1997). While the disclosure of a small number of individual genetic risk variables has been measured in empirical trials (Green et al. 2009; Teutsch et al. 2009), these approaches are limited because of the sheer volume of variants that are and will become available, along with the changing nature of both the variant information and the therapeutic considerations that impact clinical utility. Although there is mounting evidence that participants are eager to receive a broad range of genetic risk information upon contribution of materials and data to biorepository researchers (Murphy et al. 2008; Murphy et al. 2009), many institutions have not recorded participant disclosure preferences. We have not explored such preferences in this study, but rather, started with the assumption that participants have consented to receive all possible results.

When there are results that participants have requested that meet the ethical criteria for disclosure, there is not yet a consensus on the proper mechanism for participant notification. There will be challenges in coordinating and funding this careful communication with research subjects, as many pure science investigators do not have ready access to genetic counseling staff or other clinicians, and there is no broadly established mechanism for funding support for these endeavors.

There are several limitations to this study. The variant sample size is small in comparison to the number of known variants. While this creates broad confidence intervals, we are able to produce estimates that inform feasibility of the overall task. Also, in order to create a genome-wide estimate, our sampling strategy was inclusive of all possible variants in the knowledge bases we considered, rather than sampling the most frequently occurring variants. While these variants have been previously associated with disease in the scientific literature, they largely have been derived from small disease cohorts with limited control populations such that a reassessment of the evidence for pathogenicity is required. This process involves manual review of the evidence for each variant; however, we anticipate that a reasonable subset of variants will be filtered out as likely benign based upon the expanded frequency data that is emerging from whole genome studies.

Additionally, for asymptomatic individuals, there is currently no authoritative source for relative risk of disease for the majority of variants in the knowledge bases we used; most manuscripts review a disease in the context of a small number of individuals, so there may be limitations to external generalizability that will be

uncovered as new whole genome data becomes available. When such data is reliably available, future studies should evaluate these estimates in the context of the likelihood of encountering each variant in participants. Additionally, there is no established quantitative standard in these published guidelines about what constitutes an “established and substantial” associated risk for disease. For this study, we relied on a human review process that was necessarily subjective, although we used a consensus-based process that reviewed a primary associated publication and clinical research data sources to increase objectivity. These reviews were conducted for research purposes, outside of the clinical context of specific patients, which presents limitations for robustness and external generalizability. The growth estimates are also based on the assumption that the pace of discovery will continue at current rates for the coming four years, however this rate of discovery may change over time.

Our findings have implications not only for genome researchers, but also for clinicians. Microarrays and targeted sequencing are already used diagnostically, and it is anticipated that whole genome sequencing will eventually be integrated into clinical care (Green and Guyer 2011). While suspicious lesions discovered during imaging research are routinely investigated clinically, our findings suggest that the same standard will not be feasible in genomic medicine. Issues to address in future research include diagnostic and intervention costs (both at present and downstream) and the decision support systems for prioritizing and communicating large numbers of variants, in conjunction with family history and/or clinical presentation.

## Methods

### *Sources of disease-associated variant data and creation of a study sample*

The set of variants for clinical annotation and review were randomly sampled, using a stratified methodology, from high-quality, curated databases (Table 6) that include a variety of variant types, including rare mutations, large insertions and deletions, rearrangements, and polymorphisms. These include the Human Gene Mutation Database (HGMD) (Stenson et al. 2009), the National Human Genome Research Institute (NHGRI) Catalog of Published Genome Wide Association Studies (Hindorff LA) and the Online Mendelian Inheritance in Man (OMIM) (Amberger et al. 2009). When variants were listed in more than one repository, duplicate entries were removed.

All variant records in the HGMD and NHGRI databases include an associated reference publication or LSDB entry that reports an association with a phenotype. All OMIM records have a corresponding OMIM entry and many also have a reference publication.

We adopted a stratified sampling strategy to select a representative subset of variants that was drawn proportionally from each data source (Table 6). Each data source and variant category was evaluated separately (within HGMD, there are 10



variant types we consider in Table 6), and we sub-stratified the sample among the sub-categories in proportion to their frequencies. We first calculated the proportion of each variant type with respect to the entire set of variants we considered (104,304 total variants). We then multiplied that subset percentage by 160, and took the integer portion of that decimal number plus a simulated coin flip multiplied by the remaining (non-decimal) portion of the number to determine whether an extra variant would be sampled for that category. This resulted in an integer number of variants to review for each category. Though the large majority of variants are derived from HGMD, we proportionally sampled from all three databases.

To calculate the order of review for variants within each category, we assigned a randomly generated, unique number to each variant. The sampling strategy then rank-ordered each variant by its random number and included the determined number of variants from that category into the final variant sample.

#### *Exclusion Criteria*

To avoid a bias toward selecting higher quality studies, we evaluated and independently annotated as many of the sampled variants as possible. Two of 160 sampled variants were not analyzed because the primary reference articles were not available in English, and another was excluded because there was no reference available. Excluded variants, where we were unable to create clinical annotations, were categorized as 'Unknown'.

#### *Creating scores for each of the NHLBI 2010 recommendation criteria*

For each variant in the study sample, we created scores for the key factors cited in the NHLBI 2010 recommendations. We combined the areas of clinical review from NHLBI 2010 into three major axes: clinical impact of phenotype, clinical 'actionability', and association validity (Table 7).

'Clinical impact of phenotype' focuses on phenotype-specific characteristics including reproductive impact and age of onset. 'Clinical actionability' focuses on available medical interventions and preventative behaviors which may be applied, given knowledge of the phenotypic risk, including the efficacy of available interventions, and the impact on a patient's life when undergoing those interventions. The 'association validity' focuses on the strength and validity of the reported genotype-phenotype associations, including a relative risk value (when available) and a scaled validity score.

Within each category or subcategory, we allowed for a range of scores, most often using a five-point rating scale (Supplementary Materials S1). Finally, we created written guidance to consistently categorize the scores for each clinical annotation component. We categorized the validity of the association as low, moderate, or high and supplemented that with an explanatory comment field that addressed the type of mutation, existence of consistent functional studies, and familial segregation, among other factors.

To generate consensus scores for each category, three certified genetic counselors reviewed supporting data from both clinical and research resources (including GeneReviews (Pagon 2006), OMIM (Amberger et al. 2009), and eMedicine (2002), among others) to contextualize the phenotype, determine the strength of the association between the variant and the disease, and to identify any potential treatments. Additionally, we thoroughly reviewed the primary citation associated with each variant in the knowledge bases. Notably, each validity score is based only on the above data, an approach motivated by the significant time and resources that would be required to conduct an exhaustive literature review for a sample set of this size. Therefore, the validity scores are based on careful consideration of the presented data, but not derived from any specific algorithm, and are categorized broadly as low, moderate, or high validity.

The genetic counselors clinically annotated each variant association, and when all three counselors were in agreement, the score was considered a final consensus score. Upon review of the available information, the genetic counselors were able to reach consensus in almost all but one case. However, where there was not a true consensus, the score was determined by the majority opinion, with two of the three counselors in agreement.

Occasionally, a specific characteristic in a variant annotation was not scorable; some examples include a highly variable disease severity or age of onset, which we would mark as 'Variable' with a comment, or if insufficient information was available to reach a score, the characteristic was scored as 'Unknown'. We marked items as 'Not Applicable' when the score was not applicable to a specific variant or phenotype.

#### *Evaluating the responsibility for communication to a participant*

Once the representative sample of variants was clinically annotated, we evaluated whether each variant would be appropriate for communication to research participants using communication strategies derived from the NHLBI 2010 Working Group.

We matched each characteristic from the disclosure guidelines to a characteristic from the clinical annotations created by the genetic counselors, including strength of association, phenotypic severity, and improvement with treatment. If the score for each characteristic met the definition for the associated disclosure criterion, the variant qualified for the recommendation to disclose to participants. This is illustrated as the green portion of Table 2. If one or more criteria were missing or unknown, the responsibility to disclose was designated as 'Unknown'. If all disclosure criteria were met but the phenotypic severity was considered variable, but had the potential to meet the threshold to disclose, then that variant was labeled 'Variable'.

To focus this study on identifying the number of variants recommended for disclosure to research participants, we excluded a number of external factors from consideration. We specified that all participants would have consented to learning

about variants they carried, that the genetic assay would have met analytic validity standards, and that the disclosure of results to participants was IRB-approved and met all state and federal laws.

*Estimating the number of variants genome-wide that met the threshold for recommendation to communicate under NHLBI 2010*

To estimate the number of variants, genome-wide, that would meet the threshold for recommended disclosure, we used the observed proportion of sampled variants that fulfilled the requirements for recommended reporting (6.88%, plus an additional 3.75% under the less strict interpretation). To reach the estimate, we used the sampling fraction, the proportion of the variants that we sampled from the total variant knowledge base that was considered in this study. We divided the number of variants that met the threshold for recommended or potential communication under each disclosure strategy by the sampling fraction to reach an estimate of variants from the total variant knowledge base that would meet each set of criteria.

In addition to the estimate that was generated, we used the Agresti-Coull binomial confidence interval method (Brown et al. 2001) to calculate 95% confidence intervals for each disclosure result. Because the sample was drawn from a finite population of variants, we completed a finite population correction factor, and the effect was negligible because the sample size was small in comparison to the total number of variants.

*Potential growth of number of variants that would be disclosed to participants under NHLBI 2010*

We estimated the potential growth of disease-associated variants using data from the HGMD and the NHGRI Catalog of Published GWAS (which represented over 98% of the variants in our knowledge base.) We applied standard regression analysis to quarterly reports of total variant counts from the previous four years, calculating logarithmic, linear, and exponential regression models, and estimated the future growth of variants following a linear trend over the following four years.

*Number of previously identified substitution variants detected in the whole genome analysis of 36 asymptomatic individuals*

We compared the whole genome sequence data from 36 asymptomatic individuals that are publicly available from Complete Genomics with all substitution variants from the knowledge base used in this study (Drmanac et al. 2010; Genomics 2011). We then calculated the maximum, minimum, and average number of variants that were identified in each sample, and also recorded whether each variant was heterozygous or homozygous in each individual.

## **Acknowledgements**

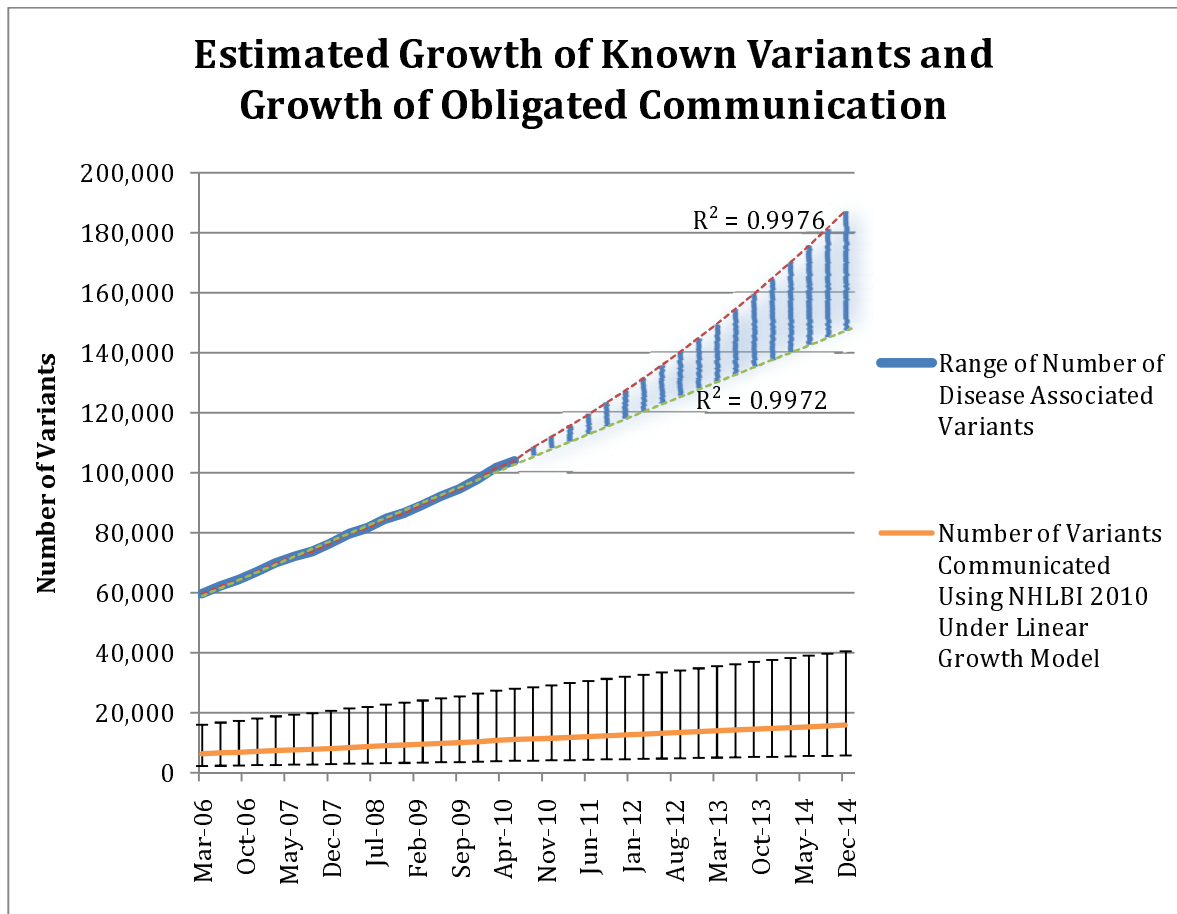
This research was supported by grant LM010470-01 from the National Library of Medicine and the Manton Center for Orphan Diseases at Children's Hospital Boston (Dr. Mandl), by training grant HD040128 from the National Institute of Child Health and Human Development (Dr. Cassa), by HG02213, HG005092 and AG027841 (Dr. Green), and by the BCM Clinical and Translational Research Program and the Baylor Annual Fund (Dr. McGuire). We thank our genetic counselors, Meghan Connolly and Catherine Clinton, for their service on the expert committee. We would also like to thank Dr. Mark Tong for his assistance in the whole genome analysis of the Complete Genomics samples.

## Figure Legends

Figure 1: Estimated growth of the knowledge base of disease-associated variants and the number of variants that may meet the threshold for recommended communication to research participants. The quarterly totals of variants from the Human Gene Mutation Database and the NHGRI Catalog of Published GWAS over four years were the basis for creating logarithmic (lower line) and exponential (upper line) regressions ( $R^2=0.9976$ ,  $0.9972$ ). The range of likely growth is highlighted in blue between these two lines. These regressions were extrapolated to estimate the possible growth rates of disease-associated genetic variants in the following four years. Linear growth rate data ( $R^2=0.9977$ ) were also used to extrapolate the estimated number of variants that would be shared with research participants under the 2010 Guidelines for disclosure. 95% confidence intervals for each quarterly estimate are shown as bars.

## Figures

Figure 1:



## Tables

**Table 1: Summary of key U.S. guidelines on the return of genetic research results.\* While there is no universally accepted set of standards for the return of genetic research results, many of the groups that**

have visited this issue have placed priority on the scientific validity, clinical significance, and availability of medical interventions.

	Disclosure Policy	To Whom Information Can be Disclosed
NBAC (1999)	Individual data only if valid, confirmed, has significant health implications, can treat or ameliorate	Research participant
CDC (2001)	Aggregate and individual data only if likely to lead to evidence-based intervention	Research participant
RAND (2003)	Aggregate data only	Public - via internet, newsletter, scientific meeting
NHLBI (2004)	Individual data only if analytically valid, replicable, and significant, has severe health implications, can treat or prevent	Research participant
NCI (2007)	Aggregate and individual data	Research participant, participant's health care provider, family
PRIM&R (2007)	Individual data only if compelling rationale	Research participant**
NIH GWAS (2007)	Individual data only in rare circumstances	Downstream users disclose to contributing investigator
NHGRI (2008)	Right to access individual data unless results are of unproven clinical validity and judged by IRB to be of no benefit to subjects	Research participant
NHLBI (2010)	Individual data only if analytically valid, replicable, and significant, has important health implications, can treat or prevent	Research participant
NCI (2011)	Aggregate and individual data if research participant has been consented to receive research results and if results are analytically valid, clinically significant or serious, and clinically actionable	Research participant

\*This is a representative list of U.S. guidelines. It is not intended to be exhaustive.

\*\* Research participant must be informed about plans to disclose so investigator can disclose to others with participant consent.

**Table 2: Applying return of results recommendations to our set of genetic variant annotations. Each variant in this study was reviewed and annotated by an expert committee of three genetic counselors and was given a score for each of the criteria cited in the NHLBI 2010 recommendations. If each of the scores for the variant met the recommendations (in the zone highlighted in green in each row) then that variant was judged to have met all criteria to warrant disclosure. If the score for even one criterion is in a zone in blue, there is no responsibility to disclose the variant to a participant.**

NHLBI 2010 Recommendation 1 as applied to our variant annotations						
Constraint	Criterion	Value (ranges from low to high, left to right)				
All must be met	Validity of association	Low	Moderate		High	
	Phenotypic severity	1	2	3	4	5
	Improves with treatment	0	1	2	3	4
	Analytic validity and laws	We assume legal testing in a CLIA approved lab				
	Participant consent	We assume that the participant has consented				

**Table 3: The number of variants from our sample that would be shared with participants according to NHLBI 2010 Recommendation 1 is shown in the upper row. The lower row shows an extrapolation from the sample to the whole genome. Cells include the number of results to be communicated in each category, the associated percentage of disease-associated variants, and 95% confidence intervals.**

Variant Sample	Ethical obligation to disclose the variant	Variant may be disclosed (variable disease expression)	No ethical obligation to disclose variant	Unknown or insufficient information

NHLBI 2010 Sample	11 (6.88%) [3.80-12.06]	6 (3.75%) [1.58-8.17]	140 (87.50%) [81.37-91.80]	3 (1.88%) [0.42-5.68]
NHLBI 2010 Genome-wide	7,171 [3,955-12,581]	3,911 [1,649-8,526]	91,266 [84,871-95,753]	1,956 [435-5,925]

**Table 4: Estimated growth of disease-associated variants and the estimated growth of disease-associated variants that would meet the threshold for recommended disclosure over the next four years. We used data from two major sources of disease-associated variants (the Human Gene Mutation Database and the NHGRI GWAS Catalog) over the previous four years to estimate the growth of disease-associated variants and the number of those variants that would meet the threshold for disclosure to research participants. These estimates are based on the observed 10.6% variant disclosure rate from our sample when following NHLBI Recommendation 1. 95% confidence intervals are in brackets for NHLBI disclosure estimates.**

Date	Expected Total Known Variants	Expected NHLBI 2010 Potentially Communicated Variants [95% CI]
Mar-06	59,666	6,339 [3,989, 9,835]
Mar-07	69,851	7,421 [4,670, 11,514]
Mar-08	79,651	8,462 [5,326, 13,129]
Mar-09	89,352	9,493 [5,974, 14,728]
Mar-10	101,686	10,804 [6,799, 16,761]
Mar-11	110,708	11,762 [7,403, 18,248]
Mar-12	121,091	12,865 [8,097, 19,960]
Mar-13	131,447	13,966 [8,789, 21,667]
Mar-14	141,803	15,066 [9,482, 23,374]
Mar-15	152,159	16,166 [10,174, 25,081]

Table 5: Aggregate results from the whole genome interpretation of 36 Complete Genomics genomes of asymptomatic individuals. Whole genome sequence data from 36 publicly available genomes of asymptomatic individuals (Drmanac et al. 2010; Genomics 2011) was analyzed, using the substitution variants from the knowledge base in this study. The total number of variants identified in each sequence is reported, along with the subset of those that are homozygous and heterozygous.	Total Study Variants Identified	Homozygous Study Variants	Heterozygous Study Variants
<b>Minimum</b>	1812	623	1028
<b>Maximum</b>	2252	835	1371

<b>Average</b>	2120	737	1214
<b>Std. Dev.</b>	70	54	74



**Table 6: Variant types and data sources. Variants were sampled from three different databases that curate genotype-phenotype associations derived from the scientific and medical literature. A stratified sampling method was used.**

Data source (date of access: type of variants included)	Number of published variants	Number of sampled variants	Percentage of variants in sample
HGMD (Mar, 2010: Mutations, Insertions, Deletions, Other)	100,329	154	96.25
HGMD missense & nonsense	56,457	86	53.75
HGMD small deletions	15,805	24	15.00
HGMD splicing	9,600	15	9.38
HGMD small insertions	6,513	10	6.25
HGMD gross deletions	6,201	10	6.25
HGMD regulatory	1,766	3	1.88
HGMD small indels	1,473	2	1.25
HGMD gross insertions	1,260	2	1.25
HGMD complex rearrangements	947	1	0.63
HGMD repeat variations	307	1	0.63
NHGRI <i>not</i> in HGMD (Jan, 2010: Genome Wide Association Study SNPs)	2,131	3	1.88
OMIM <i>not</i> in HGMD (Jun, 2010: Single Nucleotide Substitutions)	1,844	3	1.88
<b>Total</b>	<b>104,304</b>	<b>160</b>	<b>100</b>

**Table 7: Clinical Annotation Characteristics.** A team of three certified genetic counselors used a consensus-based method to create clinical annotations for each variant in three categories: clinical impact of phenotype, clinical actionability, and association validity. For each of these categories, a set of characteristics was scored for each variant, most often using a rating scale.

Categories	Characteristics
<b>Clinical impact of phenotype</b>	Severity with treatment, severity without treatment, age of onset, and reproductive issues
<b>Clinical actionability</b>	Efficacy of available treatment, invasiveness and difficulty of treatment, frequency and duration of treatment
<b>Association validity</b>	Association validity and relative risk (when applicable)

## Supporting Online Materials

### Supporting Online Materials Table 1:

Categories	Characteristics	Scoring method	Description
Clinical Impact of Phenotype	<b>Age of onset:</b>	Prenatal, 0-2, 3-7, 8-13, 14-18, 19-30, 31-50, 50+	Age or age range during which phenotype presents and/or begins to have impact on individual's health
	<b>Severity with treatment</b>	Scored on a Likert scale, ranged 1 to 5, low to high	Phenotypic severity after treatment (if available), considering impact on day to day life and longterm morbidity and risk
	<b>Severity without treatment</b>	Scored on a Likert scale, ranged 1 to 5, low to high	Phenotypic severity before or without any treatment, considering impact on day to day life and longterm morbidity and risk
	<b>Reproductive issues:</b>	No impact	Possible reproductive considerations resulting from phenotype
		Physical or medical problem(s) directly affecting reproduction	
Decreased reproductive fitness based on lifespan issues			
Decreased reproductive fitness based on physical or mental disability			
Clinical Actionability	<b>Efficacy</b>	Scored on a Likert scale, ranged 1 to 5, low to high	How effective are available treatments and interventions, how much can these interventions alleviate or prevent symptoms/presentations of the condition
	<b>Invasiveness/Challenge</b>	Scored on a Likert scale, ranged 1 to 5, low to high	How physically invasive are available interventions, how much risk is involved, how challenging might these interventions be (physically, emotionally, socially, etc.)
	<b>Frequency/Duration</b>	Scored on a Likert scale, ranged 1 to 5, low to high	How often must an affected individual undertake treatments and interventions or engage in management behaviors
	<b>Medical benefit of presymptomatic knowledge</b>	Scored on a Likert scale, ranged 1 to 3, low to high	If available, how beneficial are presymptomatic interventions, including screening and monitoring, prophylactics measures, etc.
	<b>Medical harm of presymptomatic knowledge</b>	Scored on a Likert scale, ranged 1 to 3, low to high	If available, how potentially harmful are presymptomatic interventions, including screening and monitoring, prophylactics measures, etc.
	<b>Comments</b>	Standardized list of intervention types	Details concerning presymptomatic medical interventions
Association Validity	<b>Relative Risk</b>	Scored on a Likert scale, ranged 1 to 4, low to high	If available, what is the relative risk (or odds ratio) value for the association
	<b>Validity</b>	Scored on a Likert scale, ranged 1 to 3, low to high	How strong is the reported association, how likely is it that the variant is causative or a risk factor for the listed phenotype
	<b>Validity comments</b>	Free text field	Details concerning the validity score, including study factors and association data that contributed to validity score

Supporting Online Materials Figure 1: The complete criteria used by the consensus group of expert by genetic counselors to annotate each sampled variant.

**Supporting Online Materials Table 2: Annotation scores for the 160 reviewed variants. Each row includes the phenotype of the reviewed variant, in addition to the severity of associated phenotype with and without treatment, as well as the validity of each phenotype-variant association.**

Phenotype	Severity without treatment	Severity with treatment	Change in Severity	Validity
22q11.2 deletion	4	3	1	Low
22q11.2 deletion syndrome (VCFS)	4	3	1	Moderate
Acute intermittent porphyria	3	3	0	Low
AD Spastic paraplegia	4	4	0	Low
Agammaglobulinemia (XL)	4	3	1	Low
Age of natural menopause	2	1	1	Unknown
Alexander disease	4	4	0	High
Alport syndrome (XL)	5	Variable	Variable	Low
Altered nAChR function	Unknown	Unknown	Unknown	Low
Androgen sensitivity in prostate cancer cells as a result of somatic mutation	Unknown	Unknown	Unknown	Low
Aniridia	3	2	1	High
Aniridia	3	2	1	Low
Aniridia	3	2	1	Low
Aplastic anemia	Variable	3	Variable	Low
AR Polycystic kidney and hepatic disease	4	3	1	Low
Association with myocardial infarction*	4	4	0	Low
Ataxia telangiectasia	4	4	0	High
Autosomal recessive deafness	1	1	0	High
Autosomal recessive osteopetrosis	5	4	1	Moderate
Bardet-Biedl syndrome	3	3	0	Low
Beckwith-Wiedemann syndrome	4	3	1	Moderate
Benign recurrent intrahepatic cholestasis	3	3	0	Low
Beta thalassemia - Hb Korea	2	2	0	Low
Bethlem myopathy	3	3	0	High
Bifid nose, renal agenesis, and anorectal malformation syndrome (BNAR)	Variable	Unknown	Unknown	High

<b>Catecholaminergic polymorphic ventricular tachycardia</b>	4	3	1	Low
<b>Central Core disease</b>	Variable	Variable	Variable	High
<b>Cerebral cavernous malformation type1</b>	3	3	0	High
<b>Cerebral cavernous malformations</b>	3	3	0	Unknown
<b>Charcot-Marie-Tooth disease 4C</b>	4	4	0	Low
<b>Charcot-Marie-Tooth disease 4h</b>	4	4	0	Moderate
<b>Charcot-Marie-Tooth disease type 1b</b>	4	4	0	Moderate
<b>Chronic granulomatosis disease (AR)</b>	4	3	1	Moderate
<b>Colorectal cancer (NOS?)</b>	Unknown	Unknown	Unknown	Unknown
<b>Complement C7 deficiency</b>	3	3	0	Unknown
<b>Complement C7 deficiency</b>	3	3	0	High
<b>Congenital hypothyroidism</b>	2	1	1	Low
<b>Congenital lipoid adrenal hyperplasia</b>	5	1	4	High
<b>Coronary artery disease</b>	4	4	0	Moderate
<b>Currarino syndrome</b>	5	2	3	High
<b>Cystic fibrosis</b>	4	4	0	Moderate
<b>Cystic fibrosis</b>	4	4	0	Low
<b>Cystic fibrosis</b>	4	4	0	High
<b>Cystinuria</b>	3	2	1	Low
<b>Duchenne muscular dystrophy</b>	5	5	0	Low
<b>Dystrophic epidermolysis bullosa</b>	5	5	0	High
<b>Early onset sarcoidosis</b>	3	3	0	Moderate
<b>Ectodermal dysplasia</b>	4	3	1	Low
<b>Ectopia lentis</b>	2	2	0	Low
<b>Emery Dreifuss muscular dystrophy</b>	4	4	0	High
<b>Enhanced S Cone syndrome</b>	2	2	0	Moderate
<b>Epidermolysis bullosa simplex (Dowling-Meara type)</b>	4	3	1	Unknown
<b>Erythrokeratoderma variabilis</b>	2	2	0	Moderate
<b>Fabry disease</b>	5	4	1	Low

<b>Factor XI deficiency (Hemophilia C)</b>	3	3	0	High
<b>Familial adenomatous polyposis</b>	5	3	2	Moderate
<b>Familial adenomatous polyposis</b>	5	3	2	Low
<b>Familial adenomatous polyposis</b>	5	3	2	Low
<b>Familial intrahepatic cholestasis</b>	5	4	1	Low
<b>FAP (attenuated)</b>	4	2	2	Low
<b>FAP (attenuated)</b>	4	2	2	High
<b>Gastric cancer susceptibility</b>	5	4	1	Low
<b>Gitelman syndrome</b>	1	1	0	Low
<b>Glioma</b>	Variable	Variable	Variable	Low
<b>Glucose transporter 1 deficiency syndrome</b>	4	4	0	Moderate
<b>Glucose-6-phosphate dehydrogenase deficiency</b>	3	3	0	Low
<b>Glucosephosphate isomerase deficiency</b>	3	3	0	Moderate
<b>H-antigen deficiency (Bombay)</b>	1	1	0	Moderate
<b>Harlequin Ichthyosis</b>	5	4	1	Moderate
<b>Harlequin ichthyosis (AR)</b>	5	4	1	Moderate
<b>Hemoglobin Tacoma</b>	1	1	0	High
<b>Hereditary angioedema</b>	3	3	0	Moderate
<b>Hereditary angioedema</b>	3	3	0	Low
<b>Hereditary Breast and Ovarian Cancer syndrome</b>	5	3	2	High
<b>Hereditary Breast and Ovarian Cancer syndrome</b>	5	3	2	High
<b>Hereditary Breast and Ovarian Cancer syndrome</b>	5	3	2	Low
<b>Hereditary Breast and Ovarian Cancer syndrome</b>	5	3	2	Moderate
<b>Hereditary hemorrhagic telangiectasia</b>	3	3	0	High
<b>Hereditary Hemorrhagic Telangiectasia</b>	3	3	0	High
<b>Hereditary Spastic Paraplegia</b>	4	3	1	Moderate
<b>Hirschsprung disease</b>	4	2	2	Moderate

<b>Hirschsprung disease</b>	4	2	2	Low
<b>HLA-B null allele</b>	1	1	0	Moderate
<b>HNPCC</b>	5	4	1	Low
<b>Homocystinura cblD type</b>	4	3	1	Moderate
<b>Hypercholesterolemia</b>	Variable	3	Variable	Low
<b>Hypercholesterolemia</b>	Variable	3	Variable	Low
<b>Hypercholesterolemia</b>	Variable	3	Variable	High
<b>Hypertension (essential)</b>	3	2	1	Low
<b>Hypoprothrombinemia</b>	3	3	0	Moderate
<b>Incomplete congenital stationary night blindness (type 2)</b>	1	1	0	Low
<b>Increased risk for type II diabetes</b>	4	2	2	High
<b>Juvenile intestinal polyposis</b>	4	3	1	Low
<b>Limb Girdle muscular dystrophy type 2a</b>	3	3	0	Low
<b>Long QT syndrome</b>	4	2	2	Low
<b>Macular corneal dystrophy type 1</b>	3	3	0	Low
<b>Marfan syndrome</b>	4	4	0	Low
<b>Marinesco-Sjogren syndrome</b>	4	4	0	High
<b>Maturity onset diabetes of the young (MODY) type II</b>	Variable	1	Variable	Low
<b>Maturity onset diabetes of the young (MODY) type III</b>	Variable	1	Variable	High
<b>Modifier of risk for obesity</b>	3	1	2	Low
<b>Molybdenum cofactor deficiency</b>	5	Unknown	Unknown	Unknown
<b>MPS II - Hunter syndrome</b>	Variable	Variable	Variable	Low
<b>Mucopolysaccharidosis type 2 (I-cell disease)</b>	5	5	0	High
<b>Multiple endocrine neoplasia type 1</b>	4	3	1	Low
<b>Multiple epiphyseal dysplasia</b>	3	3	0	Low
<b>Nemaline myopathy</b>	Variable	Variable	Variable	High
<b>Nephrogenic diabetes insipidus</b>	4	2	2	Low
<b>Nephropathic cystinosis</b>	5	4	1	High

<b>Neurofibromatosis type 1 (NF1)</b>	Variable	Variable	Variable	High
<b>Neurofibromatosis type 1 (NF1)</b>	Variable	Variable	Variable	Moderate
<b>Nevoid basal cell carcinoma</b>	5	3	2	Moderate
<b>Niemann-Pick disease type A</b>	5	5	0	High
<b>Niemann-Pick type C</b>	5	5	0	Moderate
<b>Nocturnal asthma</b>	3	2	1	High
<b>Oculocutaneous albinism type 1A</b>	2	1	1	Unknown
<b>Oculocutaneous albinism, type 1A</b>	2	1	1	Unknown
<b>Oral white sponge nevus</b>	1	1	0	Moderate
<b>Ornithine transcarbamylase deficiency</b>	5	2	3	Low
<b>Osteogenesis imperfecta</b>	4	4	0	Moderate
<b>Osteoporosis</b>	3	2	1	Moderate
<b>PKU (phenylketonuria)</b>	5	1	4	High
<b>Polycystic kidney disease 2</b>	5	4	1	Moderate
<b>Polycystic kidney disease type 1 (AD)</b>	4	3	1	Moderate
<b>Polymorphic ventricular tachycardia</b>	4	3	1	Moderate
<b>Primary congenital glaucoma</b>	4	2	2	Moderate
<b>Primary open-angle glaucoma</b>	3	1	2	Low
<b>Progressive external ophthalmoplegia</b>	Variable	Variable	Variable	Moderate
<b>Properdin deficiency</b>	4	1	3	Moderate
<b>Protein S deficiency</b>	3	2	1	Low
<b>Protoporphyrria</b>	3	3	0	Unknown
<b>Pulmonary hyptertension</b>	Unknown	Unknown	Unknown	Low
<b>Renal glucosuria</b>	1	1	0	Low
<b>Retinitis Pigmentosa - autosomal dominant</b>	2	2	0	Low
<b>Retinol deficiency</b>	2	Unknown	Unknown	Moderate
<b>Retinoschisis</b>	2	2	0	Moderate
<b>Rett syndrome</b>	5	5	0	Moderate
<b>Rett syndrome</b>	5	5	0	Low
<b>Rett syndrome</b>	5	5	0	High



<b>Salla disease</b>	5	5	0	High
<b>Shprintzen-Goldberg syndrome (or related disorders)</b>	4	4	0	Moderate
<b>Slow channel myasthenic syndrome</b>	3	3	0	High
<b>Spherocytosis</b>	Variable	Variable	Variable	Low
<b>Spondyloepiphyseal dysplasia</b>	3	3	0	High
<b>Stargardt disease</b>	2	2	0	Unknown
<b>Susceptibility to TB</b>	3	3	0	Moderate
<b>Tropical pancreatitis</b>	3	3	0	Moderate
<b>Tuberous sclerosis</b>	4	4	0	Moderate
<b>Tuberous sclerosis</b>	4	4	0	Unknown
<b>Tuberous sclerosis</b>	4	4	0	Moderate
<b>Tyrosinemia type 1</b>	5	2	3	High
<b>Ullrich congenital muscular dystrophy</b>	4	4	0	Low
<b>Variegate porphyria</b>	Variable	3	Variable	Low
<b>Von Hippel-Lindau</b>	4	4	0	Low
<b>Walker Warburg syndrome</b>	5	5	0	Moderate
<b>Wilson disease</b>	4	2	2	Moderate
<b>Wilson disease</b>	4	2	2	Moderate
<b>Xeroderma pigmentosum (variant type)</b>	5	4	1	Moderate
<b>XL lymphoproliferative syndrome</b>	5	4	1	Moderate
<b>XL SCID (severe combined immunodeficiency)</b>	5	2	3	Moderate

## References

- (NHGRI) NHGRI. 2011a. Clinical Sequencing Exploratory Research (U01). Vol 2011.
- . 2011b. Development of a Preliminary Evidence Base to Inform Decision-making about Returning Research Results to Participants in Genomic Studies (R01). Vol 2011.
- . 2011c. Ethical, Legal, and Social Implications of Returning Research Results to Genomic Research Participants (R21). Vol 2011.
- Amberger J, Bocchini CA, Scott AF, Hamosh A. 2009. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* **37**(Database issue): D793-796.

- Ashley EA, Butte AJ, Wheeler MT, Chen R, Klein TE, Dewey FE, Dudley JT, Ormond KE, Pavlovic A, Morgan AA et al. 2010. Clinical assessment incorporating a personal genome. *Lancet* **375**(9725): 1525-1535.
- Beskow LM, Burke W. 2010. Offering individual genetic research results: context matters. *Sci Transl Med* **2**(38): 38cm20.
- Beskow LM, Burke W, Merz JF, Barr PA, Terry S, Penchaszadeh VB, Gostin LO, Gwinn M, Houry MJ. 2001. Informed consent for population-based research involving genetics. *JAMA* **286**(18): 2315-2321.
- Blow N. 2009. Biobanking: freezer burn. *Nature Methods* **6**: 173-178.
- Bookman EB, Langehorne AA, Eckfeldt JH, Glass KC, Jarvik GP, Klag M, Koski G, Motulsky A, Wilfond B, Manolio TA et al. 2006. Reporting genetic results in research studies: summary and recommendations of an NHLBI working group. *Am J Med Genet A* **140**(10): 1033-1040.
- Boston Water and Sewer Commission. 1996. Boston planimetric and topographic data.
- Brown L, Cai T, DasGupta A. 2001. Interval Estimation for a Binomial Proportion. *Statist Sci* **16**(2): 101-133.
- Caulfield T, McGuire AL, Cho M, Buchanan JA, Burgess MM, Danilczyk U, Diaz CM, Fryer-Edwards K, Green SK, Hodosh MA et al. 2008. Research ethics recommendations for whole-genome research: consensus statement. *PLoS Biol* **6**(3): e73.
- Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, Carnevali P, Nazarenko I, Nilsen GB, Yeung G et al. 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**(5961): 78-81.
- Fabsitz RR, McGuire AL, G.L. B, al. e. 2010. Ethical and Practical Guidelines for Reporting Genetic Research Results to Study Participants: Updated Guidelines From a National Heart, Lung, and Blood Institute Working Group. *Circulation: Cardiovascular Genetics* **3**: 574-580.
- Fernandez CV, Kodish E, Weijer C. 2003. Informing study participants of research results: an ethical imperative. *IRB* **25**(3): 12-19.
- Fernandez CV, Weijer C. 2006. Obligations in offering to disclose genetic research results. *Am J Bioeth* **6**(6): 44-46; author reply W10-42.
- Genomes P. 2010. GET-Evidence: Variant Impact Score.
- Genomics C. 2011. Sample Sequence Data.
- Green ED, Guyer MS. 2011. Charting a course for genomic medicine from base pairs to bedside. *Nature* **470**(7333): 204-213.
- Green RC, Roberts JS, Cupples LA, Relkin NR, Whitehouse PJ, Brown T, Eckert SL, Butson M, Sadovnick AD, Quaid KA et al. 2009. Disclosure of APOE genotype for risk of Alzheimer's disease. *N Engl J Med* **361**(3): 245-254.
- Hindorff LA JH, Hall PN, Mehta JP, and Manolio TA. A Catalog of Published Genome-Wide Association Studies.
- Holtzman NA, Murphy PD, Watson MS, Barr PA. 1997. Predictive genetic testing: from basic research to clinical practice. *Science* **278**(5338): 602-605.

- Illes J, Rosen AC, Huang L, Goldstein RA, Raffin TA, Swan G, Atlas SW. 2004. Ethical consideration of incidental findings on adult brain MRI in research. *Neurology* **62**(6): 888-890.
- Illes J, Tairyan K, Federico CA, Tabet A, Glover GH. 2010. Reducing barriers to ethics in neuroscience. *Front Hum Neurosci* **4**.
- Issues TPCftSoB. 2011. Current Issues in Neuroscience and Neuroimaging. Transcript from Feb. 28, 2011, in Washington, D.C.
- Johnson AD, Bhimavarapu A, Benjamin EJ, Fox C, Levy D, Jarvik GP, O'Donnell CJ. 2010. CLIA-tested genetic variants on commercial SNP arrays: potential for incidental findings in genome-wide association studies. *Genet Med* **12**(6): 355-363.
- Khoury MJ, Feero WG, Reyes M, Citrin T, Freedman A, Leonard D, Burke W, Coates R, Croyle RT, Edwards K et al. 2009. The Genomic Applications in Practice and Prevention Network. *Genetics in Medicine* **11**(7): 488-494.
- Kohane IS, Mandl KD, Taylor PL, Holm IA, Nigrin DJ, Kunkel LM. 2007. Medicine. Reestablishing the researcher-patient compact. *Science* **316**(5826): 836-837.
- Kohane IS, Masys DR, Altman RB. 2006. The incidentalome: a threat to genomic medicine. *JAMA* **296**(2): 212-215.
- Kohane IS, Taylor PL. 2010. Multidimensional results reporting to participants in genomic studies: getting it right. *Sci Transl Med* **2**(37): 37cm19.
- Kozanczyn C, Collins K, Fernandez CV. 2007. Offering results to research subjects: U.S. Institutional Review Board policy. *Account Res* **14**(4): 255-267.
- MacNeil SD, Fernandez CV. 2006. Offering results to research participants. *BMJ* **332**(7535): 188-189.
- McGuire AL. 2008. 1000 Genomes on the Road to Personalized Medicine. *Per Med* **5**(3): 195-197.
- Meltzer LA. 2006. Undesirable implications of disclosing individual genetic results to research participants. *Am J Bioeth* **6**(6): 28-30; author reply W10-22.
- Murphy J, Scott J, Kaufman D, Geller G, LeRoy L, Hudson K. 2008. Public expectations for return of results from large-cohort genetic research. *Am J Bioeth* **8**(11): 36-43.
- . 2009. Public perspectives on informed consent for biobanking. *Am J Public Health* **99**(12): 2128-2134.
- Office of Biorepositories and Biospecimen Research NCI, National Institutes of Health. 2010. Workshop on Release of Research Results to Participants in Biospecimen Studies. Office of Biorepositories and Biospecimen Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland.
- Ormond KE, Wheeler MT, Hudgins L, Klein TE, Butte AJ, Altman RB, Ashley EA, Greely HT. Challenges in the clinical application of whole-genome sequencing. *Lancet* **375**(9727): 1749-1751.
- Pagon RA. 2006. GeneTests: an online genetic information resource for health care providers. *J Med Libr Assoc* **94**(3): 343-348.
- Palmour N, Affleck W, Bell E, Deslauriers C, Pike B, Doyon J, Racine E. 2011. Informed consent for MRI and fMRI research: analysis of a sample of Canadian consent documents. *BMC Med Ethics* **12**: 1.
- Q MRS. 2002. eMedicine. *Med Ref Serv Q* **21**(2): 47-54.

- Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsler JR, Masys DR. 2008. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* **84**(3): 362-369.
- Sadatsafavi M, Marra C, Li D, Illes J. 2010. An ounce of prevention is worth a pound of cure: a cost-effectiveness analysis of incidentally detected aneurysms in functional MRI research. *Value Health* **13**(6): 761-769.
- Stenson PD, Ball EV, Howells K, Phillips AD, Mort M, Cooper DN. 2009. The Human Gene Mutation Database: providing a comprehensive central mutation database for molecular diagnostics and personalized genomics. *Hum Genomics* **4**(2): 69-72.
- Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, Dotson WD, Douglas MP, Berg AO. 2009. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet Med* **11**(1): 3-14.
- Tong MY, Cassa CA, Kohane IS. 2011. Automated validation of genetic variants from large databases: Ensuring that variant references refer to the same genomic locations *Bioinformatics*.
- White MT, Gamm J. 2002. Informed consent for research on stored blood and tissue samples: a survey of institutional review board practices. *Account Res* **9**(1): 1-16.
- Wolf SM, Lawrenz FP, Nelson CA, Kahn JP, Cho MK, Clayton EW, Fletcher JG, Georgieff MK, Hammerschmidt D, Hudson K et al. 2008a. Managing incidental findings in human subjects research: analysis and recommendations. *J Law Med Ethics* **36**(2): 219-248, 211.
- Wolf SM, Paradise J, Caga-anan C. 2008b. The law of incidental findings in human subjects research: establishing researchers' duties. *J Law Med Ethics* **36**(2): 361-383, 214.

## Figures

Figure 1:

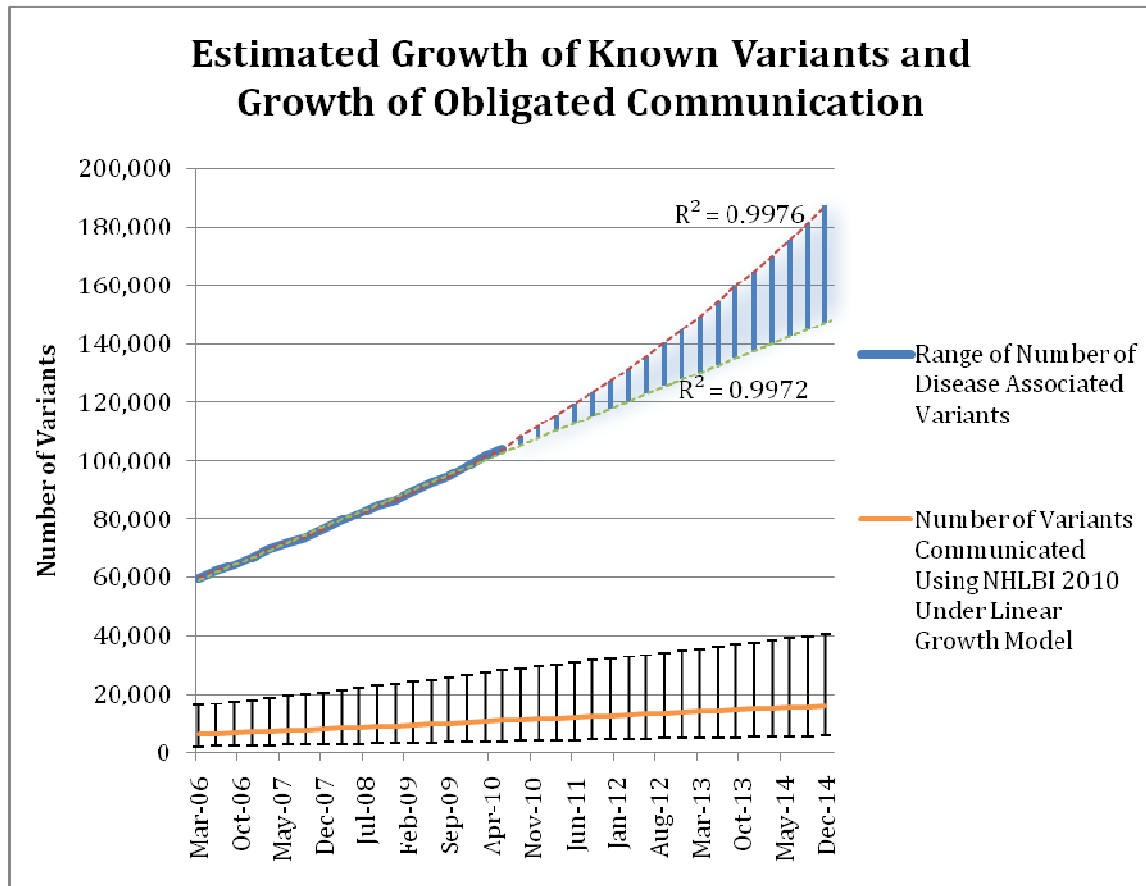


Figure 1: Estimated growth of the knowledge base of disease-associated variants and the number of variants that may meet the threshold for recommended communication to research participants. The quarterly totals of variants from the Human Gene Mutation Database and the NHGRI Catalog of Published GWAS over four years were the basis for creating logarithmic (lower line) and exponential (upper line) regressions ( $R^2=0.9976$ ,  $0.9972$ ). The range of likely growth is highlighted in blue between these two lines. These regressions were extrapolated to estimate the possible growth rates of disease-associated genetic variants in the following four years. Linear growth rate data ( $R^2=0.9977$ ) were also used to extrapolate the estimated number of variants that would be shared with research participants under the 2010 Guidelines for disclosure. 95% confidence intervals for each quarterly estimate are shown as bars.