Personalized genomic disease risk of volunteers

Manuel L. Gonzalez-Garay[a,1], Amy L. McGuire[b], Stacey Pereira[b], and C. Thomas Caskey[c,1]

[a]Center for Molecular Imaging, Division of Genomics and Bioinformatics, The Brown Foundation Institute of Molecular Medicine, University of Texas Health Science Center, Houston, TX 77030; and [b]Center for Medical Ethics and Health Policy, Department of Medicine and Medical Ethics, and [c]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030

Contributed by C. Thomas Caskey, August 27, 2013 (sent for review July 11, 2013)

Next-generation sequencing (NGS) is commonly used for researching the causes of genetic disorders. However, its usefulness in clinical practice for medical diagnosis is in early development. In this report, we demonstrate the value of NGS for genetic risk assessment and evaluate the limitations and barriers for the adoption of this technology into medical practice. We performed whole exome sequencing (WES) on 81 volunteers, and for each volunteer, we requested personal medical histories, constructed a three-generation pedigree, and required their participation in a comprehensive educational program. We limited our clinical reporting to disease risks based on only rare damaging mutations and known pathogenic variations in genes previously reported to be associated with human disorders. We identified 271 recessive risk alleles (214 genes), 126 dominant risk alleles (101 genes), and 3 X-recessive risk alleles (3 genes). We linked personal disease histories with causative disease genes in 18 volunteers. Furthermore, by incorporating family histories into our genetic analyses, we identified an additional five heritable diseases. Traditional genetic counseling and disease education were provided in verbal and written reports to all volunteers. Our report demonstrates that when genome results are carefully interpreted and integrated with an individual's medical records and pedigree data, NGS is a valuable diagnostic tool for genetic disease risk.

molecular medicine | disease prediction | whole exome sequencing

Sequencing the whole genome of patients with genetic disorders has become reality since the sequencing of the first individual human in 2007 (1). Further advances in massively parallel DNA sequencing are reducing the price of sequencing an entire genome or exome. The quality and speed of sequencing and analyzing a personal genome are improving at an unprecedented pace, making possible the introduction of next-generation sequencing (NGS) into the clinic on a research basis (2–7). Advancements in NGS have stimulated international research initiatives to identify genetic links to rare disorders in children, with an average diagnostic success of 20–25% and the discovery of new disease–gene associations (8–12).

The rapidly increasing number of aging adults in our society will place unprecedented demands on the health care system. To provide adults with a healthy longevity we need to develop a system to identify genetic risk and apply early intervention on pathology progression. In this report, we decided to sequence the whole exomes of a healthy adult cohort of 81 volunteers and evaluate the value of applying NGS in combination with medical history and pedigree data. In this report we plan to address three main questions. (*i*) What genetic discoveries need to be provided to the volunteers? (*ii*) What is the practical value of delivering this information to volunteers? (*iii*) What are the challenges and barriers to the adoption of this powerful technology into medical practice?

The individual genetic reports yield helpful medical risk information, suggesting that population sequencing of asymptomatic adults may prove to be valuable and useful. We provided to the participants, under our institutional review board, genetic risk findings from the analyses and genetic counseling to discuss their results.

Results

Categories of Variants to Report to Patients. Variants obtained from our workflow (described in Fig. 1) were reported using three categories. Our first variant category consists of variants identified in an individual where the alleles are found in Human Genome Mutation Database (HGMD) (13, 14) and labeled disease-causing mutations (DM). These alleles also were required to be rare [<1% allele frequency in 6,500 exomes from the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (15) and the 1,000 Genomes Project Genomes (16, 17)] and predicted to be damaging to protein function by two of three predictions algorithms [Polyphen 2.0 (18), Sift (19–24), and MutationTaster (25)] using Database of Human Non-synonymous SNVs and their functional predictions and annotations (dbNSFP) (26) as described in Fig. 2. The genome sequence data of each volunteer were reviewed and interpreted, taking into account personal medical history, a three-generation pedigree with family history of diseases, and bioinformatics analysis. The medical history of each volunteer in this cohort was rich with detail because each had a private physician used for annual examinations, and in some cases, disease therapy. Fig. 3 summarizes the results of our pipeline: we recruited 81 nonrelated volunteers and sequenced their genomic DNA using exome sequencing. We detected 65,582 unique nonsynonymous coding variants (nscv). Every nscv was interrogated for human inherited disease mutations using the HGMD (13, 14) database from Biobase (DM category consisting of 109,708 variations). We were able to detect 1,036 HGMD (13, 14) DM variations. After using the filters described in Fig. 2, the number was reduced to 275 pathogenic variants. We identified in our cohort 208 autosomal recessive (AR) alleles (169 genes), 64 autosomal dominant (AD) alleles (44 genes), and three X-linked recessive (XLR)

Significance

Replacing traditional methods for genetic testing of inheritable disorders with next-generation sequencing (NGS) will reduce the cost of genetic testing and increase the information available for the patients. NGS will become an invaluable resource for the patient and physicians, especially if the sequencing information is stored properly and reanalyzed as bioinformatics tools and annotations improve. NGS is still at the early stages of development, and it is full of false-positive and -negative results and requires infrastructure and specialized personnel to properly analyze the results. This paper will explain our experience with an adult population, our bioinformatics analysis, and our clinical decisions to assure that our genetic diagnostics were accurate to detect carrier status and serious medical conditions in our volunteers.

Author contributions: M.L.G.-G. and C.T.C. designed research; M.L.G.-G., A.L.M., S.P., and C.T.C. performed research; M.L.G.-G. analyzed data; and M.L.G.-G., A.L.M., S.P., and C.T.C. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence may be addressed. E-mail: Manuel.LGonzalezGaray@uth.tmc.edu or tcaskey@bcm.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1315934110/-/DCSupplemental.

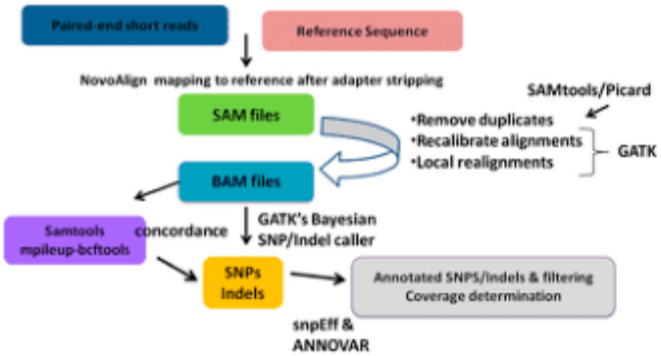


Fig. 1. Workflow for processing NGS data. Raw sequencing data are aligned against the reference sequence using Novoalign software from NovoCraft. SAM files are preprocessed using SAMtools and Picard to create BAM files and remove duplicates. The Genome Analysis Toolkit (GATK) is then used to recalibrate the alignments, perform local realignments, and identify SNPs and indels. Finally, SnpEff and ANNOVAR are used to annotate variants.

alleles (3 genes). These data resulted in an average of 3.5 disease allele reports per volunteer.

The approach for a second category of variants consisted of creating a personalized list of candidate genes from Online Mendelian Inheritance in Man (OMIM) (27, 28) known to be associated with the disorders reported in the medical literature. We detected 131 alleles (131 genes) using this approach. Each one of these variants provided a potential causation for the volunteer's disorders. Each one of the variations obtained from this approach passed our stringent pipeline. This approach added on average another 2.0 disease alleles per volunteer report.

The third approach used a family history to create a personalized list of candidate genes from OMIM (27, 28), and as before, we compared our list of candidate genes with the disorders reported in the family history.

Before reporting an allele to the volunteer, we reviewed the original publications that support the pathogenicity of all of the alleles (HGMD) and/or the evidence associating the gene with the disorder (OMIM). At this time, all three abovementioned categories of investigation were reported in full recognition; some would be found to be non–disease-producing alleles as databases improve and functional assays complement informatics predictions. We have updated clinical reports as these data emerged and counseled the patients on the options for reducing or eliminating the disease risk.

Disease Genes Identified in the Cohort. Table S1 summarizes our disease associations. Matching personal medical records to personal genome reports was informative. We elected to report findings as disease–gene associations instead of reporting findings as diagnostic because we did not included in our study traditional "surrogate markers" (analytes, proteins, and imaging) for the confirmation of a disease diagnosis. We considered potentially causative findings to be those mutations that are predicted to be damaging in addition to being reported in either HGMD (13, 14) or OMIM (27, 28) databases. These mutations are considered to be "need to know" and are reported to volunteers. There was identification of associations for vascular disease and/or hypercholesterolemia in five individuals related to LDL receptor (*LDLR*) alleles. *LDLR* mutations are causative of early onset autosomal dominant coronary artery disease (CAD) and manifest hypercholesterolemia (29, 30). Three individuals were taking statins related to their hypercholesterolemia. Two individuals were not under care but had history of personal hypercholesterolemia and in one case a son with hypercholesterolemia.

There were four volunteers detected with risk genes for diabetes mellitus (31–34). Two of the individuals were under therapy for diabetes 2, whereas two additional volunteers had elevated fasting blood sugars and were being followed by their physicians for further analytes measurements. There were two individuals with morbid obesity (body mass index of 32 and 37 kg/m^2) who carried an *MC4R* allele associated with pediatric obesity and rare heterozygotic adults (35, 36). Two ophthalmologic disease/gene associations were identified. The childhood brittle corneal syndrome type 1 occurred in a volunteer who had undergone successful corneal transplant and carried a putative compound heterozygosity in *ZNF469* (37). One volunteer was under care for macular dystrophy and carried an *ABCA4* allele (38). One sterile male volunteer was found to have an insertion in gene *USP26* (known to be responsible for infertility in men) (39). Associations for melanoma and breast cancer were identified. The two patients with melanoma carried different gene allele associations: *GRIN2A* and *BAG4* (40–42). Two volunteers diagnosed with breast cancer had different allele associations in *BRCA2* (43, 44). Single cases of early onset prostate (*LRP2*) (45) and follicular thyroid cancer (*TPR*) cancer were identified (46, 47). A volunteer with nonsyndromic deafness was found to have risk alleles in two genes associated with autosomal dominant (AD) deafness and had a three-generation positive family history of deafness (48). In each case, the volunteer was instructed to inform their physician and was requested to confirm the genomic allele identification in a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory, even when each reported allele had been sequenced twice in independent studies. The finding provided information for personal and family risk counseling not possible before gene association.

Incorporation of Three-Generation Pedigrees into the Genetic Analyses. The three-generation pedigree medical information was analyzed to identify those volunteer families who warranted additional genetic study. Table S2 lists those genetic disorders identified by pedigree/familial medical history. In each case, the volunteer was counseled for the family risk and encouraged to contact at risk family members who may benefit from focused genetic studies. Three of the families have reported that they have had their familial genetic diagnosis resolved at this time [paraganglioma (49), Prader-Willi syndrome (50, 51), and ankylosing spondylitis (AS) (52)]. One additional family is under study [Tourette syndrome (53)]. Additional familial disease risks were identified by history for atrial fibrillation (AR), bicuspid aortic valve (BAV), dyslexia (AR), Fabry's (XLR), gall stones (AD), and myotonic dystrophy (anticipation AD). Success with this approach was productive but not universally accepted because disease/gene resolution requires interaction with interested and motivated family members.

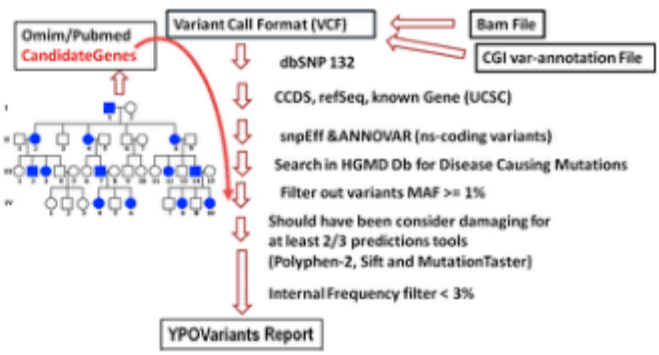


Fig. 2. Pipeline to generate variants reports. Every variant in the variant call format file is annotated using spnEff and ANNOVAR; nonsynonymous coding variants are annotated using the commercial version of the HGMD database. (*Left*) Our selection of variants by the creation of a personalized candidate gene list using medical history and family history for each volunteer. Mutations with a minor allele frequency of >1% are removed using frequencies from the NHLBI exome sequencing project (ESP), 1,000 Genomes Project. Variants that are consider benign by two of three predictions tools are removed (using dbNSFP). Finally, we remove variants that are present in our cohort more than three times.

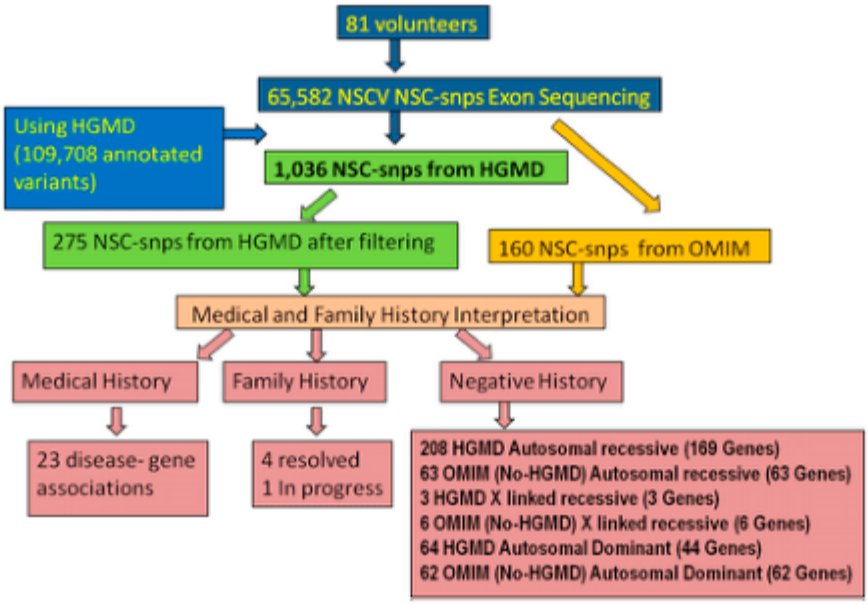


Fig. 3. Summary of results. The flowchart provides the number of variants from each step of the pipeline described in Fig. 2.

Table S3 provides a sampling of the recessive risk alleles. They constitute the majority of the observed alleles. Of the 160 offspring of the 81 volunteers, no children were affected with these disorders. All volunteers indicated their families were complete, and thus, no spousal genetic studies were recommended, but information was proposed to be provided to reproductive age descendants. Many of the genes identified are part of prenatal carrier screens and/or newborn state-sponsored screening programs [phenylketonuria, maple syrup urine disease, cystic fibrosis, Niemann-Pick disease, Gaucher disease, factor V Leiden thrombophilia, medium-chain acyl-CoA dehydrogenase (MCAD) deficiency]. Undoubtedly, NGS will expand the number of non-unreported disease alleles and scope of genes studied for couples in the pregnancy setting. The Beyond Batten Disease Foundation of Austin, TX (54), has this goal.

Table S4 shows that a category of high concern was the identification of XLR disease risk alleles among our female volunteers. One volunteer had an affected son (isolated case) with Fabry disease that was diagnosed before our study. There were four disease alleles identified, each listed in HGMD (13, 14). There was no family history of these disorders found in the three-generation pedigree of each. All were counseled to have their test confirmed and daughters studied in a CLIA-certified laboratory given the high disease risk (50% for men). Three men in our study had alleles predicted from the OMIM (27, 28) disease database to be causative for cutis laxus, Duchenne muscular dystrophy, congenital nystagmus, and hemophilia A, illustrating the challenge of predicting damaging mutations bioinformatically. None had the disorders. Counseling and family study were individualized for each disease risk. Volunteers were made aware of database errors in the reports.

Tables S5–S10 provide a third category that is very problematic, the AD group. The allele identification is as previously described, but counseling is more difficult because of variation in severity and time onset. For this age group of volunteers, the interest was high because disease prevention was frequently expressed as a goal in the face-to-face counseling meetings. A poststudy survey also reflected this objective. We focused in this paper on the three major causes of death in the United States: cancer, cardiovascular disease, and neurodegenerative disease. In our analysis of each volunteer, we reviewed the genomic and family data.

Table S5 lists the breast cancer risk results. There were 12 volunteers found to have breast cancer risk alleles of genes *BRCA1, BRCA2, PALB2, RAD51C*, and *RAD50*. Two volunteers with *BRCA2* risk alleles were diagnosed with breast cancer. One man carried a premature chain termination mutation and has a first-degree relative with breast cancer (50s). A third volunteer had a frame shift mutation (high-risk allele) but not found to have breast cancer. All alleles were predicted to be damaging. Eight volunteers had first-degree relatives with breast cancer, whereas four had a negative family history of disease. All were advised to seek confirmation via a CLIA-certified laboratory. One patient with an HGMD (13, 14) allele was confirmed but predicted to be "neutral" by a commercial laboratory. All were counseled regarding the need for regular mammograms and gynecological examinations and were requested to inform their physician of this research risk allele identification.

Table S6 displays the colon cancer alleles. There was no disease incidence of colon cancer in this group with the exception of one volunteer with a positive dysplastic polyp biopsy. Five volunteers had a positive family history of colon cancer. Five volunteers had no family history of disease. All were advised to obtain confirmatory CLIA-certified laboratory diagnosis and advise their physician of the research allele identification. Of the 10 volunteers, many had undergone colonoscopy as part of their health care.

Table S7 includes all of the remaining type of cancers. Two volunteers diagnosed with melanomas were found to have different disease gene risk alleles. We identified 10 volunteers with prostate risk alleles. One volunteer reported a diagnosis of prostate cancer at age 55 while the other nine volunteers reported no familial history of the disease. Genetic counseling for cancer risk required the greatest counseling time. The concepts of the two-hit hypothesis (55) and "somatic mutations" (56) were difficult to grasp for the volunteers, even when we discussed the subject in great detail during the education session. All volunteers were provided information regarding standard of practice approaches for early detection of the respective cancer.

Table S8 lists all of the affected volunteers with cardiomyopathies (57). Five volunteers had a medical history of cardiac dysrhythmia with identified risk alleles. One younger (50s) volunteer had first-degree relatives requiring pacemakers and carried two risk alleles. Three volunteers had either stent placements or bypass procedures related to CAD. Each was in their 70s.

Table S9 lists the 11 volunteers who had no apparent disease but had a positive family history of tachycardia, sudden death, and CAD and carried risk alleles. We provide this experience to broaden alertness to both genetic causation and risk of disease

for adult-onset cardiovascular disease (58). Of the alleles listed in Tables S8 and S9, 13 alleles were found in HGMD (13, 14). We advised volunteers to inform their physicians of these results for their long-term clinical care.

In Table S10, we listed the results for adult-onset neurodegenerative diseases. Our findings were limited but of high interest to the cohort. It was frequently asked by volunteers if they had Alzheimer's risk. We summarize our findings for Alzheimer's and Parkinson risk alleles (59, 60). The genes included *APOE, APP, PSEN1, MAPT, EIF461, GBA, GIGYF2, LRRK2, PARK2, PM20D1*, and *SNCA*. There were nine volunteers with HGMD (13, 14) listed risk alleles. Of these, two had a positive family history of Parkinson disease and one with Alzheimer's disease. One of the *PARK2* alleles occurred in a volunteer who provided a history of three second-degree relatives in a sibship affected with disease. The reminder had no family history of either disease. There were 25 alleles predicted to be damaging. One is a frameshift allele. None of these volunteers had a family history of disease.

Discussion

Exome Sequencing Is Limited. The full spectrum of disease mutation identification is not satisfied by exome sequencing alone because large deletions, copy number variations (CNVs), and triplet repeats are not reliably identified at this time. Furthermore, exon capture relies on probe design. For example, the discovery of the *MAGEL2* mutation in our Prader-Willi patient was made using whole genome sequencing (WGS) from complete genomics and missed by exome capture because of high GC content (51). The accuracy of coding allele identifications was, however, quite high and thus of great utility as a genome screening approach. CGI (61) sequencing produced higher coverage than exome sequencing; data for CNV, large deletions, and regulatory elements will have utility as we analyze previously labeled "junk" DNA for disease causation (62). There is also the issue of our limited knowledge of disease alleles within the databases. One of our biggest challenges for the interpretation of human genomes is the lack of gene annotations and the errors in databases. Our knowledge base for human disorders is small. There are only ~100,000 pathogenic variants in the HGMD (13, 14) database and a fraction of them have errors. If we do not use annotated variants but instead gene annotations as our source of information, we can calculate the fraction of knowledge that we can use at this time. For example, the number of genes associated with human disorders reported by HGMD (13, 14), OMIM (27, 28), UniProtKB (63), Gene Atlas (64), etc. is 4,622. From the 4,622 genes, only 1,955 genes have high-quality data because they are part of the GeneTest (65) database. GeneTest (65) is a database originally created by the National Center for Biotechnology Information to track all of the laboratories worldwide that offer a genetic test for a gene. With this information, we know that the fraction of genes that we can use for the interpretation of a human genome of a successful high-quality whole exome or whole genome dataset is ~7–18% when using the high confidence set of 1,955 genes or a set of 4,622 genes. Despite these limitations, this report documents the utility for disease associations and risk.

During the last few years, the field of NGS has developed a large number of tools that make it easier to handle the analysis of reads, variant calling, functional prediction, and annotation (66). There are also large publicly available datasets of healthy individuals that can be used as controls that can be used to remove technology specific errors or filter out common polymorphisms. As we begin to use whole genome sequencing at an increasing depth, we are discovering more variants, so these public datasets are becoming increasingly important for quality control and filtering of variants in smaller projects. One of the main limitations is the lack of access to public and private genome and exome variants. There are thousands of datasets, but the majority are inaccessible to the scientific community. We recognize the existence of the 1,000 Genomes project, the NHLBI Exome Sequencing Project (ESP), Exome variant server, and the 69 sets of whole genomes from CGI (15–17, 67). However, we need larger datasets from very carefully phenotyped patients to assist in the interpretation of the variants in our patients. The million genome project of the US Department of Veterans Affairs (68) has the potential to provide such data, as well as private health plans considering adaptation of genome sequencing.

Genetic Discoveries Provided to Volunteers. There are several approaches to disclose the results to volunteers. Groups like Patel et al. use the statistics and epidemiology approach in reporting the polygenic risk assessment using common SNPs that have been previous associated with genetic disorders from genome-wide association studies (69). The PGP-10 project uses an automated tool or Genome Environment Trait Evidence (GET-Evidence) system, wich is a system that is collaboratively edited (70). For this project, we decided to focus on reporting only high-quality variants that are rare in the population and considered damaging by two of three commonly used predictions algorithms. In addition, the variant has to be either reported in HGMD under category DM or the gene has to have been previously associated with a genetic disorder (OMIM). The group of volunteers consisted of adults with complete medical and family history so we personalized the reports as described in Fig. 2 to specifically try to identify molecular explanations for the maladies reported in their medical or family history. This approach generated reports that were easy to explain and accepted by the patients during the genetic counseling session.

Medical Histories and Family Pedigrees Complement Sequencing Results. The utility of genome data was significantly enhanced when integrating standard medical care features of personal and family disease diagnosis. The significant number of 23 disease associations in all likelihood represents a bias of our volunteers to seek answers to their personal disease history. This observation may hold a key to how we obtain maximal use of genome sequencing—sequence the disease index cases. Our experience would suggest a high value for that utilization. This approach has been clearly documented to be successful for pediatric genetic disorders but not exploited for adult-onset disease. The practical value of this study is summarized in Tables S1 and S2 and fell into two general categories: (*i*) new knowledge of the genetic risk and heritability for themselves and family; and (*ii*) options for therapy (CAD) or imaging (cancer) for personal and extended family care. By using the medical and family history, we were able to clarify the genetic risk in 6 of the 81 cases. One of the cases yielded a new discovery of a gene associated with Prader-Willi syndrome, which is described in another paper (51).

Prenatal vs. Adult Genetic Screening. The technology and this report beg the question of whether we are prepared to offer adult disease risk screening. Currently, prenatal and newborn screening for a selected set of frequently occurring disease alleles (not genome sequencing) is a standard of practice. There are questions that deserve medical and ethical review before adult screening becomes a standard of practice. First, for reproductive and newborn diagnosis, typically only actionable childhood diseases are explored, which respects the future autonomy of the child and preserves her right to an open future (71, 72). Because adult screening decisions would be made by an autonomous individual for her own health decisions, broader conceptions of utility, including personal utility, need to be considered (73). It is a clear and simple decision to provide patients with actionable genetic information from a WES study; on the other hand, it is challenging and it raises a difficult ethical question to decide what to do with incidental genetic findings that are not actionable and could lead to physiological distress to the patient (e.g. APO-E for Alzheimer disease). Despite this ethical dilemma our group of volunteers elected to receive information even if the genetic information might not be actionable. Only 3% of the volunteers were uncertain about receiving nonactionable information (*SI Poststudy Survey*).

Volunteer Response to Clinical Reports. From our poststudy survey, we found that 72% of the responders reported speaking with their physician about their results. This raises important questions about whether nongeneticists are adequately prepared to counsel patients based on WES results and whether such follow-up will lead to iatrogenic harm or unjustified use of health care resources (74). Twenty-five percent reported changing their behaviors because of the results, which is surprising given that previous reports found no significant behavior change resulting from adult risk screening in a direct-to-consumer setting (75). Despite that all of the participants were clearly informed that their results originated from two independent sequencing experiments and that we advised them to have their results clinically validated in a CLIA-certified laboratory, 78% reported that they did not have the results confirmed. This low percentage of confirmatory results from the volunteers raises the question of whether it is sufficient to counsel research participants to have results clinically confirmed or if investigators should be required to confirm results before disclosure.

It was apparent for some volunteers that they were seeking information related to familial diseases. Resolution of these questions required family member interest and motivation because, in all cases, we had sequenced the nonrisk family member. We followed up each case with a referral to a qualified genetics program with diagnostic capacity for the suspected genetic disease.

Our efforts to analyze cancer, cardiovascular, neurodegenerative, and obesity/diabetes risk were successful but needed considerable education/counseling to avoid confusion over risk vs. diagnosis. Second, there are standard of care options for those with risk alleles for cancer, cardiovascular disease, and diabetes for disease modification or early diagnosis. Thus, sequencing serves as a new screening risk detection approach toward the objective of improved health. It is expected that genomic studies will increase surveillance studies (e.g., colonoscopy, gynecologic examinations, mammograms, cardiovascular markers and scanning studies) but has the possibility of more precisely identifying the patients who may benefit from disease prevention surveillance.

The area of adult-onset neurologic disorders is an increasing concern worldwide as our population ages, thus exposing disease incidence not seen earlier. The genetic disease discoveries are limited. Confirmatory diagnostics such as image analysis and biomarkers/surrogate markers are just emerging, and prevention therapeutic options are nonexistent. Although one might question the utility of screening for these disorders at this time, the experience with Huntington disease (76) screening taught valuable lessons on how to proceed with studying and counseling families at risk. Furthermore, there are new therapeutic trials in disease prevention for Alzheimer's (58) and Parkinson disease based on the genetic cause of disease. These clinical trials use genetic diagnosis to select participants, which is also a successful approach in cancer drug development (77–79).

Barriers to the Adoption of Genetic Screening via Sequencing. Although the above comments would present the case for the value of adult genetic screening via whole genome sequencing, there are major issues to be addressed. In our opinion, the least is sequencing technology and cost. Bioinformatics focused on the practical extraction of medical relevant/actionable data are a challenge. We relied heavily on HGMD alleles for "need to know" information to patients. This approach is flawed in three ways: (*i*) databases contain errors; (*ii*) highly validated disease databases are scattered, private, and limited; and (*iii*) the future will provide more disease risk alleles by sequencing than by patient reports in the literature. Our current limitation for interpretation of a genome is not the quality of the data of the coverage of the genome but our disease knowledge database. R. Cotton's Human Variome Project (62) together with Beijing Genome Institute are proposing to create a highly validated disease allele database.

New technological advances such as structure-based prediction of protein–protein interactions on a genome wide scale (80), 3D structure of protein active and contact sites (81), high-throughput functional assays of damaging alleles (81–83), and new approaches that combine analytes, metabolomics and genetic information from a single individual (84) are just a few examples of the new technologies that will help us to generate better interpretation of genomic data.

The delivery of the genome risk information will need to be carried out by a new cadre of physicians and counselors skilled in medicine, genetics, and education/counseling. These experts will need to integrate into medical care as well as has been done for newborn screening, prenatal diagnosis, and newborn genetic disease diagnosis.

The approach of adult screening is in its early phase but from our data appears very promising. We conclude that the genomic study of adults deserves intensified effort to determine if "need to know" genome information has the utility for improved quality of health for our aging population.

Materials and Methods

The oversight of this research was under two institutional review boards: (*i*) HSC-IMM-08-0641 (University of Texas Health Science Center at Houston) and (*ii*) H-30710 (Baylor College of Medicine).

Cohort Description. The cohort consists of members and spouses in the Houston Chapter of the Young Presidents Organization (YPO) (85). The entire description of the cohort can be found in *SI Materials and Methods*.

WES Sequencing. Standard NGS was performed using illumina HighSeq; an extended explanation can be found in *SI Materials and Methods*.

Sequencing Analysis. Fig. 1 illustrates our pipeline, and Fig. 2 describes our pipeline to detect known pathogenic variations. Additional details can be found in *SI Materials and Methods*.

Counseling. Genome counseling was conducted by a board-certified internist and medical geneticist by both individual meetings and two written summaries over a period of 12 mo. Additional information can be found in *SI Materials and Methods*.

ACKNOWLEDGMENTS. This work was supported by the Cullen Foundation for Higher Education and the Governing Board of the Greater Houston Community Foundation. The funding organizations made the awards to the University of Texas Health Science Center at Houston and Baylor College of Medicine. C.T.C. was the principal investigator of both grants.

1. Levy S, et al. (2007) The diploid genome sequence of an individual human. *PLoS Biol* 5(10):e254.
2. Bamshad MJ, et al. (2011) Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12(11):745–755.
3. Tabor HK, Berkman BE, Hull SC, Bamshad MJ (2011) Genomics really gets personal: How exome and whole genome sequencing challenge the ethical framework of human genetics research. *Am J Med Genet A* 155A(12):2916–2924.
4. Lander ES (2011) Genome-sequencing anniversary. The accelerator. *Science* 331(6020): 1024.
5. Lander ES (2011) Initial impact of the sequencing of the human genome. *Nature* 470(7333):187–197.
6. Biesecker LG, Burke W, Kohane I, Plon SE, Zimmern R (2012) Next-generation sequencing in the clinic: Are we ready? *Nat Rev Genet* 13(11):818–824.
7. Hennekam RC, Biesecker LG (2012) Next-generation sequencing demands next-generation phenotyping. *Hum Mutat* 33(5):884–886.
8. Anonymous Finding of rare disease genes in Canada (forge Canada). Available at http://www.genomebc.ca/portfolio/projects/health-projects/finding-of-rare-disease-genes-in-canada-forge-canada/. Accessed September 19, 2013.
9. Gahl WA, et al. (2012) The National Institutes of Health undiagnosed diseases program: Insights into rare diseases. *Genet Med* 14(1):51–59.
10. Gahl WA, et al. (2012) The National Institutes of Health undiagnosed diseases program: Insights into rare diseases. *Genet Med* 14(1):51–59.
11. Gahl WA, Tifft CJ (2011) The NIH undiagnosed diseases program: Lessons learned. *JAMA* 305(18):1904–1905.
12. Koenekoop RK, et al.; Finding of Rare Disease Genes (FORGE) Canada Consortium (2012) Mutations in NMNAT1 cause Leber congenital amaurosis and identify a new disease pathway for retinal degeneration. *Nat Genet* 44(9):1035–1039.
13. Stenson PD, et al. (2012) The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Curr Protocol Bioinform* 39:1.13.1-1.13.20.

14. Stenson PD, et al. (2009) The Human Gene Mutation Database: 2008 update. *Genome Med* 1(1):13.
15. Anonymous NHLBI exome sequencing project (ESP) exome variant server. Available at http://evs.gs.washington.edu/EVS/. Accessed September 19, 2013.
16. Clarke L, Zheng-Bradley X, et al. (2012) The 1000 Genomes Project: Data management and community access. *Nat Methods* 9(5):459–462.
17. Abecasis GR, et al.; 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature* 467(7319):1061–1073.
18. Adzhubei IA, et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7(4):248–249.
19. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4(7):1073–1081.
20. Sim NL, Kumar P, et al. (2012) SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 40(Web Server issue):W452–W457.
21. Hu J, Ng PC (2012) Predicting the effects of frameshifting indels. *Genome Biol* 13(2):R9.
22. Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Res* 11(5):863–874.
23. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31(13):3812–3814.
24. Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* 7:61–80.
25. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 7(8):575–576.
26. Liu X, Jian X, Boerwinkle E (2011) dbNSFP: a lightweight database of human non-synonymous SNPs and their functional predictions. *Hum Mutat* 32(8):894–899.
27. Anonymous Online Mendelian Inheritance in man OMIM. Available at http://omim.org. Accessed September 19, 2013.
28. Anonymous NCBI OMIM Online Mendelian Inheritance in Man. Available at http://www.ncbi.nlm.nih.gov/omim. Accessed September 19, 2013.
29. Huijgen R, Kindt I, Defesche JC, Kastelein JJ (2012) Cardiovascular risk in relation to functionality of sequence variants in the gene coding for the low-density lipoprotein receptor: A study among 29,365 individuals tested for 64 specific low-density lipoprotein-receptor sequence variants. *Eur Heart J* 33(18):2325–2330.
30. Boekholdt SM, et al. (2012) Association of LDL cholesterol, non-HDL cholesterol, and apolipoprotein B levels with risk of cardiovascular events among patients treated with statins: A meta-analysis. *JAMA* 307(12):1302–1309.
31. Waeber G, et al. (2000) The gene MAPK8IP1, encoding islet-brain-1, is a candidate for type 2 diabetes. *Nat Genet* 24(3):291–295.
32. Mosca L, et al. (2011) Genetic variability of the fructosamine 3-kinase gene in diabetic patients. *Clin Chem Lab Med* 49(5):803–808.
33. da Silva Xavier G, et al. (2011) Per-arnt-sim (PAS) domain-containing protein kinase is downregulated in human islets in type 2 diabetes and regulates glucagon secretion. *Diabetologia* 54(4):819–827.
34. MacDonald PE, Rorsman P (2011) Per-arnt-sim (PAS) domain kinase (PASK) as a regulator of glucagon secretion. *Diabetologia* 54(4):719–721.
35. O'Rahilly S (2009) Human genetics illuminates the paths to metabolic disease. *Nature* 462(7271):307–314.
36. van den Berg L, et al. (2011) Melanocortin-4 receptor gene mutations in a Dutch cohort of obese children. *Obesity (Silver Spring)* 19(3):604–611.
37. Al-Owain M, Al-Dosari MS, Sunker A, Shuaib T, Alkuraya FS (2012) Identification of a novel ZNF469 mutation in a large family with Ehlers-Danlos phenotype. *Gene* 511(2):447–450.
38. Fritsche LG, et al. (2012) A subgroup of age-related macular degeneration is associated with mono-allelic sequence variants in the ABCA4 gene. *Invest Ophthalmol Vis Sci* 53(4):2112–2118.
39. Zhang J, et al. (2012) [Polymorphism of Usp26 correlates with idiopathic male infertility]. *Zhonghua Nan Ke Xue* 18(2):105–108.
40. Wei X, et al.; NISC Comparative Sequencing Program (2011) Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat Genet* 43(5):442–446.
41. Howell PM, Jr., Li X, Riker AI, Xi Y (2010) MicroRNA in melanoma. *Ochsner J* 10(2):83–92.
42. Xi Y, et al. (2008) Global comparative gene expression analysis of melanoma patient samples, derived cell lines and corresponding tumor xenografts. *Cancer Genomics Proteomics* 5(1):1–35.
43. Nelson HD, Huffman LH, Fu R, Harris EL; U.S. Preventive Services Task Force (2005) Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: Systematic evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med* 143(5):362–379.
44. Anonymous National Cancer Institute BRCA1 and BRCA2. Available at http://www.cancer.gov/cancertopics/factsheet/Risk/BRCA. Accessed September 19, 2013.
45. Holt SK, et al. (2008) Association of megalin genetic polymorphisms with prostate cancer risk and prognosis. *Clin Cancer Res* 14(12):3823–3831.
46. Frank-Raue K, et al. (2013) Prevalence and clinical spectrum of nonsecretory medullary thyroid carcinoma in a series of 839 patients with sporadic medullary thyroid carcinoma. *Thyroid* 23(3):294–300.
47. Mak HH, et al. (2007) Oncogenic activation of the Met receptor tyrosine kinase fusion protein, Tpr-Met, involves exclusion from the endocytic degradative pathway. *Oncogene* 26(51):7213–7221.
48. Ruel J, et al. (2008) Impairment of SLC17A8 encoding vesicular glutamate transporter-3, VGLUT3, underlies nonsyndromic deafness DFNA25 and inner hair cell dysfunction in null mice. *Am J Hum Genet* 83(2):278–292.
49. van Hulsteijn LT, Dekkers OM, Hes FJ, Smit JW, Corssmit EP (2012) Risk of malignant paraganglioma in SDHB-mutation and SDHD-mutation carriers: A systematic review and meta-analysis. *J Med Genet* 49(12):768–776.
50. Jiang Y, Tsai TF, Bressler J, Beaudet AL (1998) Imprinting in Angelman and Prader-Willi syndromes. *Curr Opin Genet Dev* 8(3):334–342.
51. Schaaf CP, et al. (2013) Truncating mutations of MAGEL2 cause autism and Prader-Willi syndrome (PWS) or PWS-like phenotypes. *Nat Genet*, in press.
52. Rashid T, Ebringer A (2011) Gut-mediated and HLA-B27-associated arthritis: An emphasis on ankylosing spondylitis and Crohn's disease with a proposal for the use of new treatment. *Discov Med* 12(64):187–194.
53. Deng H, Gao K, Jankovic J (2012) The genetics of Tourette syndrome. *Nat Rev Neurol* 8(4):203–213.
54. Anonymous Beyond Batten Disease Foundation. Available at http://beyondbatten.org/. Accessed September 19, 2013.
55. Knudson AG (1996) Hereditary cancer: Two hits revisited. *J Cancer Res Clin Oncol* 122(3):135–140.
56. Nik-Zainal S, et al.; Breast Cancer Working Group of the International Cancer Genome Consortium (2012) The life history of 21 breast cancers. *Cell* 149(5):994–1007.
57. Alcalai R, Seidman JG, Seidman CE (2008) Genetic basis of hypertrophic cardiomyopathy: from bench to the clinics. *J Cardiovasc Electrophysiol* 19(1):104–110.
58. Rader DJ, Cohen J, Hobbs HH (2003) Monogenic hypercholesterolemia: New insights in pathogenesis and treatment. *J Clin Invest* 111(12):1795–1803.
59. Martin I, Dawson VL, Dawson TM (2011) Recent advances in the genetics of Parkinson's disease. *Annu Rev Genomics Hum Genet* 12:301–325.
60. Selkoe DJ (2012) Preventing Alzheimer's disease. *Science* 337(6101):1488–1492.
61. Anonymous Complete Genomics Inc. Available at http://www.completegenomics.com. Accessed September 19, 2013.
62. Anonymous Human variome project. Available at http://www.humanvariomeproject.org. Accessed September 19, 2013.
63. Anonymous UniProtKB. Available at http://www.uniprot.org/uniprot. Accessed September 19, 2013.
64. Anonymous Gene atlas. Available at http://www.geneatlas.org/gene/main.jsp. Accessed September 19, 2013.
65. Anonymous Genetic Testing Registry (GeneTests). Available at http://www.genetests.org. Accessed September 19, 2013.
66. Anonymous SEQanswers. Available at http://seqanswers.com. Accessed September 19, 2013.
67. Anonymous 69 genomes data. Available at http://www.completegenomics.com/public-data/69-Genomes/. Accessed September 19, 2013.
68. Anonymous The million veteran program. Available at http://www.va.gov/opa/pressrel/pressrelease.cfm?id=2090. Accessed September 19, 2013.
69. Patel CJ, et al. (2013) Whole genome sequencing in support of wellness and health maintenance. *Genome Med* 5(6):58.
70. Ball MP, et al. (2012) A public resource facilitating clinical use of genomes. *Proc Natl Acad Sci USA* 109(30):11920–11927.
71. American Academy of Pediatrics Committee on Bioethics (2001) Ethical issues with genetic testing in pediatrics. *Pediatrics* 107(6):1451–1455.
72. Davis DS (1997) Genetic dilemmas and the child's right to an open future. *Hastings Cent Rep* 27(2):7–15.
73. Wolf SM, Lawrenz FP, *et al.* (2008) Managing incidental findings in human subjects research: Analysis and recommendations. *J Law Med Ethics* 36(2):219–248.
74. McGuire AL, Burke W (2008) An unwelcome side effect of direct-to-consumer personal genome testing: Raiding the medical commons. *JAMA* 300(22):2669–2671.
75. Bloss CS, Schork NJ, Topol EJ (2011) Effect of direct-to-consumer genomewide profiling to assess disease risk. *N Engl J Med* 364(6):524–534.
76. Wexler NS (2012) Huntington's disease: Advocacy driving science. *Annu Rev Med* 63:1–22.
77. Caskey CT (2007) The drug development crisis: Efficiency and safety. *Annu Rev Med* 58:1–16.
78. Caskey CT (2010) Using genetic diagnosis to determine individual therapeutic utility. *Annu Rev Med* 61:1–15.
79. Miller G (2012) Alzheimer's research. Stopping Alzheimer's before it starts. *Science* 337(6096):790–792.
80. Zhang QC, et al. (2012) Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490(7421):556–560.
81. Edwards AM, Bountra C, Kerr DJ, Willson TM (2009) Open access chemical and clinical probes to support drug discovery. *Nat Chem Biol* 5(7):436–440.
82. Namovic MT, Jarvis MF, Donnelly-Roberts D (2012) High throughput functional assays for P2X receptors. *Curr Protocol Pharmacol* Jun;Chapter 9:Unit 9.15.
83. Trivedi S, Liu J, Liu R, Bostwick R (2010) Advances in functional assays for high-throughput screening of ion channels targets. *Expert Opin Drug Discov* 5(10):995–1006.
84. Suhre K, et al.; CARDIoGRAM (2011) Human metabolic individuality in biomedical and pharmaceutical research. *Nature* 477(7362):54–60.
85. Anonymous Membership criteria YPO. Available at http://www.ypo.org/join-ypo/. Accessed September 19, 2013.

Supporting Information

Gonzalez-Garay et al. 10.1073/pnas.1315934110

SI Materials and Methods

Cohort Description. The cohort consists of members and spouses in the Houston Chapter of the Young Presidents Organization (YPO). Criteria for membership into the YPO includes corporate and community leadership (1). This cohort is well educated and of higher socioeconomic status. All 450 YPO members were invited to attend an 8-h educational program incorporating technology, human genetics, anticipated outcomes, ethical considerations, discussion groups, and technology demonstrations and printed materials. Of the 150 attendees, 81 volunteered to participate in this study: 46 men and 35 women, with an average age of 54 y. All 81 elected under the terms of the University of Texas Health Science Center at Houston's institutional review board to receive "need to know" genomic disease risk results. Each volunteer provided a detailed medical and drug use history reviewed by our physician-researcher (C.T.C.). A three-generation medical pedigree was acquired on each volunteer. One volunteer could provide no family history.

Whole exome sequencing (WES) Sequencing. Genomic DNA was extracted using a DNA kit (Promega wizard genomic DNA purification kit) following Promega's instructions (2). The cohort was sequenced twice: the first whole exome sequencing experiment (2011) was performed using Illumina's HiSeq and the Genome Analyzer IIx system (3) after enrichment with Nimblegen V2 kit (44 Mb) (4) (outsourced to the national center for genome resources). Our second WES experiment (2013) was performed using Illumina's newest machines HiSEq. 2500 (3) after enrichment with Agilent SureSelect target enrichment V5+UTRs (targeting coding regions plus UTRs) (5) (outsourced to Axeq Technologies). Genome sequencing of a small subset (24 subjects) for validation purposes was carried out by Complete Genomics Inc. (CGI) (6).

Sequencing Analysis. Our analysis pipeline consists of Novoalign (7), Samtools (8), Picard (9), and The Genome Analysis Toolkit (GATK) (10), followed by variant annotation (11–14) using multiple databases from the University of California Santa Cruz (UCSC) Genome bioinformatics site (15). Fig. 1 illustrates our pipeline. Fig. 2 describes our pipeline to detect known pathogenic variations. We detected known variants associated with human diseases using the Human Genome Mutation Database (HGMD) database from Biobase (16, 17) and genes known to be associated with human disorders from Online Mendelian Inheritance in Man (OMIM) (18, 19) and GeneTests (20). Functional effects of each nonsynonymous coding variant were evaluated using three different functional prediction algorithms [Polyphen 2.0 (21), Sift (22–27), and MutationTaster (28)] using the Database of Human Non-synonymous SNVs and their functional predictions and annotations (dbNSFP) (29). Filtration of common polymorphisms was accomplished using frequencies from the National Heart, Lung, and Blood Institute (NHLBI) exome sequencing project (ESP) (30), 1,000 Genomes (31, 32), and internally by removing any variant that appeared more than three times in our cohort. In addition, a group of candidate genes was obtained from OMIM (18, 19) for each volunteer after a careful analysis of the family and personal health history of each volunteer. Variations in those OMIM (18, 19) candidate genes were identified and submitted to the same frequency and functional effects filter as described before.

Variant Validation. Every variant identified in our pipeline was evaluated for quality control, and the variant's read alignments in the BAM file [Binary version of a SAM (Sequencing Alignment Map) file] file were visualized using Integrative Genomics Viewer (IGV) (33). The purpose of this step was to try to remove the remaining false positives.

Each genetic variant was validated using the following steps: (*i*) retrieve reads over variant sites for each individual; (*ii*) make SamTools (8) genotype calls (an alternate calling algorithm); (*iii*) retrieve quality scores for all reads; (*iv*) keep track of the directional depth and require at least two variant reads in the 5′ and 3′ orientation for a variant to be considered true; and (*v*) filter out variants if the SamTools (8) genotype call disagrees with the GATK (10) call or if the quality scores or directional depth values do not exceed minimum values.

Establishing Criteria for Highly Reliable Variant Calling from Exome Sequencing. Our first objective was to define the methods needed to identify a set of "highly reliable" variants from the Illumina sequencing and apply these methods to variant calling on all of our samples. To meet our definition of a highly reliable variant, each variant had to be detected under two independent orthogonal sequencing technologies and been considered as high quality. Because there is not a common definition of what a high-quality variant is, we decided to take advantage of the confidence category scores provided from complete genomics; variants with a score of VQHIGH are consider high quality (masterVarbeta files version 2.0) and develop an equivalent value in our illumina sequencing data. To accomplish our first objective, a dataset of variants was generated from a set of 24 samples that we sequenced using Illumina (3) and an orthogonal sequencing technology (CGI) (6). CGI has their own proprietary workflow from alignment to data annotation (34), Fig. 1 describes our analysis workflow for exome sequencing data. Fig. S2*A* shows the intersection between the nonsynonymous coding variants (NSCVs) detected by CGI (6) and Illumina (3) exome sequencing. We extracted variants from CGI with a score of VQHIGH and that were also detected in the corresponding illumina's vcf file (Fig. S2*B*). This subset of highly reliable variants represents an average of 72% of the variants detected by CGI. By using our dataset, we were able to systematically test for conditions and software setting in our pipeline that generate the majority of the highly reliable variants and reduce the probability of selecting variants not present in our dataset. We reached the conclusions that by using two variant callers tools, GATK UnifiedGenotyper and mplileup/bcftools (samtools), and selecting an overlapping set of variants, we obtained variants of the highest quality. In addition, a postcalling filter enforces that each variant has to have a mapping quality >30, a base quality >20, and a coverage ≥10, with at least a 3:7 ratio of variant to reference (Het) and the presence of the variant in reads from both orientations. By using these postcalling filters, we eliminated the majority of false-positive calls (FP).

Counseling. Genome counseling was conducted by a board-certified internist and a medical geneticist by both individual meetings and two written summaries over a period of 12 mo. The summary reports were prepared and jointly endorsed by a bioinformatician and a physician. Additional counseling was conducted by phone calls and appointments with their physician as requested by the volunteers.

Counseling of Results. Both causative and problematic alleles were reported verbally and in two written reports over an 18-mo period.

The first comprehensive report was updated ~1 y after (*i*) larger control databases downgraded some problematic alleles with more than a 1% frequency; (*ii*) private consultation with disease experts; and (*iii*) validation with original publications and small disease center databases. Several new disease–gene associations were discovered for the reported familial diseases found by pedigree and personal medical histories. Volunteers were informed that these were research results and instructed to consult with their personal physician so that they could have the results validated in a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory. Volunteers whose family members warranted genetic study were referred to the Baylor College of Medicine genetics program as a medical referral because this function was outside the institutional review board scope and Baylor College of Medicine offered both clinical genetic and CLIA Laboratory expertise. Our study preceded the publication of the incidental findings guidelines in clinical WES and whole genome sequencing (WGS) of the American College of Medical Genetics and Genomics (ACMG) (35). However, we have reviewed their list of 57 genes and 24 actionable conditions, and we found that we included all their genes in our analysis.

Poststudy Survey

We conducted an online survey to assess volunteers' experiences of participating in this project under a Baylor College of Medicine instituational review board. The survey consisted of 82 items and focused on how the volunteers felt about taking part in the research project, as well as their perspectives on genetic information in health care and genomic research in general. Study participants were told the survey was completely voluntary and that they could skip any question they preferred not to answer and could end their participation at any time.

All 81 study volunteers were invited via e-mail to participate in the anonymous online survey within 12 mo after receiving their individual genome reports. Forty-two participants responded to the online survey (response rate, 51.9%; 38 responses were complete). Of those who responded, 59% were men, 41% were women, and 95% had biological children. Ninety-seven percent described their race as white, and 5% chose "other" (participants could choose all that applied); 5% also identified themselves as Hispanic or Latino. All participants had earned a college degree, and 63% had completed at least some graduate work. All participants reported having had a routine medical check-up within the last 2 y, and when asked how they would rate their health, 58% reported excellent, 29% reported very good, 11% reported good, and 3% reported fair.

Poststudy survey results. This study had as its objective to deliver helpful medical genetic information. The mandatory education program informed volunteers that unexpected risks were to be expected. Our institutional review board required volunteers to have the options of declining this information. None chose that option.

The results of the anonymous online survey showed that, overall, participants were motivated to take part in the project to receive their genetic results and learn about their personal risk of disease. Seventy-nine percent of respondents reported that the opportunity to receive their personal genetic results was the most important factor in their decision to take part in the project, whereas another 10% cited a personal interest in genetics in general. When asked to choose which factor was most important in their decision to receive their personal genetic results, most respondents (52%) reported that their interest in finding out their personal risk for diseases was the most important factor; other important factors included the desire to get information about risk of health conditions for their children (17%), the desire to learn more about the medical conditions in their family (10%), and curiosity about their genetic makeup (10%).

Ninety-seven percent of respondents agreed or strongly agreed that they were glad that they decided to participate in this study and receive their personal results, leaving only 3% undecided. Most respondents (72%) spoke with their primary care provider about their results, and 50% reported that they spoke with other medical professionals, including cardiologists, oncologists, and obstetricians/gynecologists, among others; 22% reported that they had their twice-confirmed research results confirmed in a CLIA-certified laboratory.

Twenty-five percent of respondents reported that the test results motivated them to make changes to their health care (i.e., undergoing tests, seeing a specialist, taking vitamins or herbal supplements), exercise, medications, or insurance (Table S11).

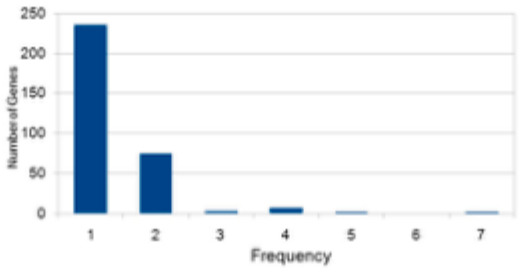
Respondents generally felt that researchers should offer personalized results to research participants: 54% felt that researchers are obligated to offer results, 22% felt that researchers are obligated to offer results only if the researcher is a physician, and the remaining 24% did not think researchers were obligated to offer results. Respondents were pleased with the methods by which they were given their results in this study, with 95% agreeing or strongly agreeing that they were glad the researchers sent them a personalized results report, and 100% agreeing or strongly agreeing that they found the in-person consultation about their results very helpful. When asked, 94% said they would also want an electronic record of their entire genome if it were available.

When asked about genetic testing in health care, 83% reported that they felt that genetic testing should be a regular part of health care and 97% agreed or strongly agreed that they felt comfortable using these results to make decisions about their health. Nevertheless, respondents were evenly split when asked if they thought these results should be part of their medical record.

In summary, our poststudy surveys indicated that volunteers were motivated to gain personal and family health knowledge, satisfied with the translation of the genetic information, and had a divided opinion about incorporating their genetic information into their medical records.

1. Anonymous Membership criteria YPO. Available at http://www.ypo.org/join-ypo/. Accessed September 19, 2013.
2. Anonymous Wizard. Available at http://www.promega.com/resources/protocols/technical-manuals/0/wizard-genomic-dna-purification-kit-protocol/. Accessed September 19, 2013.
3. Anonymous Illumina. Available at http://www.illumina.com. Accessed September 19, 2013.
4. Anonymous NimbleGen Roche. Available at http://www.nimblegen.com/products/seqcap/ez/v2/index.html. Accessed September 19, 2013.
5. Agilent Technologies Agilent SureSelect array. Available at http://www.genomics.agilent.com/en/Exome-Sequencing/SureSelect-Human-All-Exon-Kits/?cid=cat240002&tabId=AG-PR-1204. Accessed September 19, 2013.
6. Anonymous Complete Genomics Inc. Available at http://www.completegenomics.com. Accessed September 19, 2013.
7. Novocraft.com (2012) Available at http://www.novocraft.com. Accessed September 19, 2013.
8. SAMtools. Available at http://samtools.sourceforge.net/. Accessed September 19, 2013.
9. Picard. Available at http://picard.sourceforge.net/. Accessed September 19, 2013.
10. McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303.
11. Cingolani P snpEff: SNP effect predictor. Available at http://snpeff.sourceforge.net/. Accessed September 19, 2013.
12. Cingolani P, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6(2):80–92.
13. San Lucas FA, Wang G, Scheet P, Peng B (2012) Integrated annotation and analysis of genetic variants from next-generation sequencing studies with variant tools. *Bioinformatics* 28(3):421–422.
14. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38(16):e164.
15. Kuhn RM, Haussler D, Kent WJ (2013) The UCSC genome browser and associated tools. *Brief Bioinform* 14(2):144–161.
16. Stenson PD, et al. (2012) The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Curr Protoc Bioinform*.
17. Stenson PD, et al. (2009) The Human Gene Mutation Database: 2008 update. *Genome Med* 1(1):13.

18. Anonymous NCBI OMIM Online Mendelian Inheritance in Man. Available at http://www.ncbi.nlm.nih.gov/omim. Accessed September 19, 2013.
19. Anonymous Online Mendelian Inheritance in Man OMIM. Available at http://omim.org. Accessed September 19, 2013.
20. Anonymous Genetic Testing Registry (GeneTests). Available at http://www.genetests.org. Accessed September 19, 2013.
21. Adzhubei IA, et al. (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7(4):248–249.
22. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4(7):1073–1081.
23. Sim NL, et al. (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 40(Web Server issue):W452–W457.
24. Hu J, Ng PC (2012) Predicting the effects of frameshifting indels. *Genome Biol* 13(2):R9.
25. Ng PC, Henikoff S (2001) Predicting deleterious amino acid substitutions. *Genome Res* 11(5):863–874.
26. Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31(13):3812–3814.
27. Ng PC, Henikoff S (2006) Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* 7:61–80.
28. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 7(8):575–576.
29. Liu X, Jian X, Boerwinkle E (2011) dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* 32(8):894–899.
30. Anonymous NHLBI Exome Sequencing Project (ESP) exome variant server. Available at http://evs.gs.washington.edu/EVS/. Accessed September 19, 2013.
31. Clarke L, Zheng-Bradley X, et al. (2012) The 1000 Genomes Project: Data management and community access. *Nat Methods* 9(5):459–462.
32. Abecasis GR, et al.; 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. *Nature* 467(7319):1061–1073.
33. Robinson JT, et al. (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26.
34. Complete genomics (data file format standard pipeline version 2.0). Available at http://www.completegenomics.com/customer-support/documentation/100357139.html. Accessed September 19, 2013.
35. Green RC, Berg JS, et al. (2013) ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 15(7):565–574.



7	LRRK2	BRCA2									
5	CFTR	CDH23									
4	GIGYF2	LDLR	LRP2	MEFV	MYOM1	PAH	SCN5A				
3	APC	EIF4G1	TTN								
2	ABCA1	ABCA3	ABCA4	ABCB4	ABCC8	ACADM	ACAN	ACTN2	ACY1	APOB	..
1	ABCG5	ACAD9	ACAT1	ACE	ACVRL1	AGL	AGT	AKAP9	ALMS1	ALOXE3	..

Fig. S1. Grouping genes by occurrence. Frequency of genes with nonsynonymous coding mutations in our cohort. This graphic provides a summary of the number of times alleles were observed for an individual gene. In each of these cases, the allele was either part of HGMD or OMIM, rare, and carried a high polyphen2 score. An example of a gene with frequent risk alleles include *Titin*, the largest genes in our genome and recently reported to be causative of dilated cardiomyopathy. A second example of a smaller gene with a large number of variations is *CFTR*, where the disease database is deep, and it is known to be one of the most common autosomal recessive diseases in whites. This graphic supports that we did not select polymorphic genes but unique mutations in each volunteer.

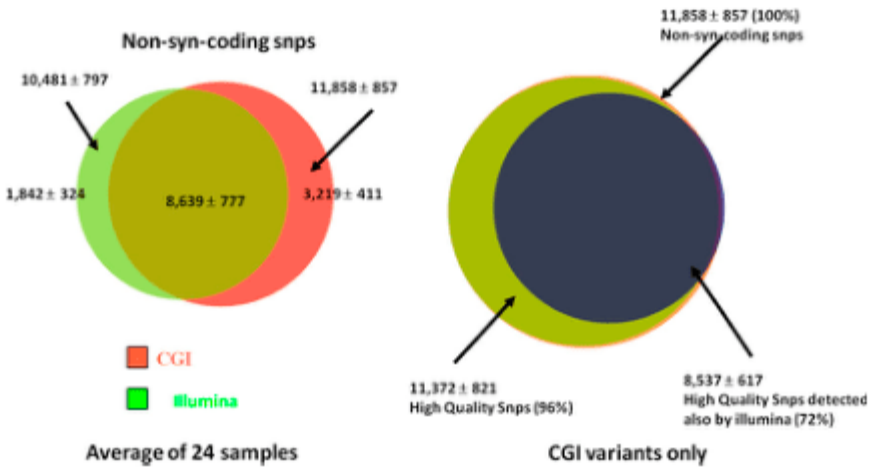


Fig. S2. Variants detected using Complete Genomics Inc. (CGI) and Illumina. (*Left*) Comparison of nonsynonymous coding SNPs (NSCS) obtained from Complete Genomics (red) and Illumina (green). Twenty-four human samples were sequenced using both technologies, and NSCS were compared in each sample. The average results were calculated and graphed as a venn diagram. The intersection represents the set of NSCS detected by both technologies. On average, 73% of the NSCS detected by CGI were also detected by Illumina, while 82% of the NSCS detected by Illumina were also detected by CGI. (*Right*) Using the same samples we calculated that 96% of all the CGI NSCS are considered "High Quality" according to the CGI proprietary quality matrix. An average of 72% of all the NSCS detected by CGI was also detected by Illumina (blue). Since two orthogonal sequence technologies detected the same set of NSCS, this group of variants most likely represents a set of real variants which we refer to as "Highly reliable NSCS." The set of "Highly reliable NSCS" were used to establish quality criteria in our Illumina's variant detection pipeline.

Table S1. Disease associations with alleles

Case	Disease	Risk gene	Allele	HGMD	OMIM gene ID
3937	Hypercholesterolaemia	*LDLR*	p.P526H	CM100938	606945
3890	Hypercholesterolaemia	*LDLR*	p.T726I	CM920469	606945
3910	Hypercholesterolaemia	*LDLR*	p.T726I	CM920469	606945
3900	Hypercholesterolaemia	*LDLR*	p.V827I	CM920471	606945
3915	Hypercholesterolaemia	*LDLR*	p.V827I	CM920471	606945
3923	Obesity	*MC4R*	p.I251L	CM030483	155541
3923	Diabetes mellitus, type II	*MAPK8IP1*	p.D386E	NA	604641
3973	Obesity	*MC4R*	p.C326R	CM070992	155541
3937	Diabetes mellitus type 2 (MODY)	*FN3K*	p.H146R	NA	608425
3937	Diabetes mellitus type 2 (MODY)	*PASK*	p.P1256L	NA	607505
3923	Macular degeneration, age related	*ABCA4*	p.G863A	CM970003	601691
3898	Brittle cornea syndrome type 1 (BCS1) keratoconus	*ZNF469*	p.D2902Y	NA	612078
3889	Male infertility	*USP26*	p.T123_Q124insT	NA	300309
3942	Melanoma	*BAG4*	p.W103X	NA	603884
3959	Melanoma	*GRIN2A*	p.N1076K	NA	138253
3896	Breast or ovarian cancer	*BRCA2*	p.I505T	CM010167	600185
3959	Breast or ovarian cancer	*BRCA2*	p.S384F	CM065036	600185
3897	Breast or ovarian cancer	*BRCA2*	p.T2515I	CM994287	600185
3950	Follicular thyroid cancer (age 41)	*TPR*	p.R105C	NA	189940
3960	Prostate cancer	*LRP2*	P.N479H	NA	600073
3960	Prostate cancer	*LRP2*	P.G4417D	NA	600073
3934	Nonsyndromic deafness	*MYH14*	p.M161I	NA	608568
3934	Nonsyndromic deafness	*SLC17A8*	p.R75C	NA	607557

NA, not available.

Table S2. Familial diseases and associations

Case	Disorder	Gene	Volunteer relatedness	Association	
				Volunteer	Affected relative
3949	Praeder Willie	*MAGEL2*	2°	—	+
3947	Paraganglioma	*SDHB*	1°	—	+
3930	Ankylosing spondylitis	*HLA-B27*	1°	—	+
3930	Tourettes	*TBD*	1°(3)	IP	IP
3928	Parkinson	*LRRK2*	1°	—	+

—, negative; IP, research in progress.

Table S3. Recessive disorders

Cases	Disease	Risk gene	Allele	HGMD	OMIM
3958	Niemann-Pick type C2 disease	*NPC2*	p.N111K	CM081368	601015
3896, 3900, 3915, 3895	Antitrypsin α1 deficiency	*SERPINA1*	p.R247C, p.E366K (3)	CM910298, CM830003	107400
3894	Glycogen storage disease 0	*GYS2*	p.Q183X	CM023388	138571
3889	Glycogen storage disease 1a	*G6PC*	p.R83C	CM930261	613742
3901	Glycogen storage disease 3	*AGL*	p.R477H	CM104343	610860
3945	Glycogen storage disease 4	*GBE1*	p.Y329S	CM960705	607839
3898	Glycogen storage disease 6	*PYGL*	p.D634H	CM078418	613741
3941, 3952	Glycogen storage disease 9B	*PHKB*	p.Q650K	CM031327	172490
3915, 3919, 3943, 3954	Fanconi anemia	*FANCA*	p.T126R, p.S858R (3)	CM043494, CM992317	607139
3936, 3934	Familial Mediterranean fever	*MEFV*	p.E148Q, p.P369S, p.R408Q	CM981240, CM990837, CM990838	608107
395, 439, 243, 953	Cystic fibrosis	*CFTR*	p.D1152H, p.S1235R,	CM950256, CM930133	602421
3933	Sandhoff disease	*HEXB*	p.A543T	CM970723	606873
3940	Fuchs endothelial dystrophy	*ZEB1*	p.Q824P	CM100242	189909
3908	Factor V deficiency	*F5*	p.P1816S	CM095204	612309
3952	Hepatic lipase deficiency	*LIPC*	p.T405M	CM910258	151670
3962	Krabbe disease	*GALC*	p.T112A	CM960678	606890
3954	Macular corneal dystrophy, type 2	*CHST6*	p.Q331H	CM055930	605294
3891, 3947, 3959, 3924, 3895, 3897	Usher syndrome 1d	*CDH23*	p.A366, p.D1806E, p.R1060W	CM050545, CM105104, CM021537	605516
3900, 3910	Phenylketonuria	*PAH*	p.A300S, p.R53H	CM920555, CM981427	612349
3933, 3946	MCAD (medium-chain acyl-coA dehydrogenase deficiency)	*ACADM*	p.K329E (2)	CM900001	607008
3914	Adrenal hyperplasia	*HSD3B2*	p.R249X	CM950655	613890
3926	17-α-hydroxylase/17,20-lyase deficiency	*CYP17A1*	p.R449C	HM0669	609300

Table S4. X-linked recessive

Case	Disorder	Risk gene	Allele	Sex	HGMD	OMIM
3891	ATRX syndrome	*ATRX*	p.N1860S	Female	CM950125	300032
3930	Fabry disease	*GLA*	p.A143T	Female	CM972773	300644
3901	Mucopolysaccharidosis II	*IDS*	p.D252N	Female	CM960865	300823

Table S5. Breast cancer risk

Case	Disease	Risk gene	Allele	Family history	Sex	Age (y)	HGMD	OMIM gene ID
3959	Breast cancer	*BRCA2*	p.S384F	Affected (44)	Female	44	CM065036	600185
3896	Breast cancer	*BRCA2*	p.I505T	Affected	Female	49	CM010167	600185
3955	Breast cancer	*BRCA2*	p.E1625fs	Negative	Female	42	CD011121	600185
3962	Breast cancer	*PALB2*	p.V1103M	First, second, third degree (2) (49–60s)	Female	51	CM118272	610355
3936	Breast cancer	*BRCA1*	p.Y856H	First degree (sister 40s)	Male	62	CM042673	113705
3936	Breast cancer	*BRCA2*	p.K2729N	First degree (sister 40s)	Male	62	CM021957	600185
3963	Breast cancer	*BRCA2*	p.R2034C	First degree (60s)	Male	48	CM994286	600185
3897	Breast cancer	*BRCA2*	p.T2515I	First degree (80)	Female	51	CM994287	600185
3934	Breast cancer	*RAD51C*	pT287A	First degree (uterine)	Female	50	NA	602774
3939	Breast cancer	*RAD50*	p.R1069X	First degree breast (60s)/second colon (60s)	Male	56	NA	604040
3912	Breast cancer	*RAD51C*	p.A126T	Negative	Male	77	CM1010201	602774
3923	Breast cancer	*RAD51C*	pT287A	Negative	Male	60	CM1010198	602774
3956	Breast cancer	*RAD51C*	pT287A	Negative	Male	59	CM1010198	602774

NA, not available.

Table S6. Colon cancer risk

Case	Disease	Risk gene	Allele	Family history	Sex	Age (y)	HGMD	OMIM gene ID
3896	Colon cancer	*MLH1*	p.K618A	First degree	Female	49	CM973729, CM950808	120436
3891	Colon cancer	*MLH3*	p.E1451K	First degree (70s)	Female	62	CM013011	604395
3897	Colon cancer	*APC*	p.A2690T	First and second degree cancer	Female	51	CM045404	611731
3904	Colon cancer	*MSH2*	p.G315V	Second degree	Male	49	CM995220	609309
3897	Colon cancer	*MSH2*	p.G12D	Negative	Female	51	CM950813	609309
3962	Colon cancer	*APC*	p.S2621C	Negative	Female	51	CM921028	611731
3955	Colon cancer	*APC*	p.R2505Q	Negative	Female	42	NA	611731
3933	Colon cancer	*MUTYH*	p.G382D	Negative	Female	69	CM020287	604933

NA, not available.

Table S7. Other cancer risk

Case	Disease	Risk gene	Allele	Family history	Sex	Age (y)	HGMD	OMIM gene ID
3959	Melanoma	*GRIN2A*	p.N1076K	Affected	Female	44	NA	138253
3942	Melanoma	*BAG4*	p.W103X	Affected	Male	70	NA	603884
3950	Follicular thyroid cancer	*TPR*	p.R105C	Affected	Male	48	NA	189940
3960	Prostate cancer	*LRP2*	p.N479H	Affected	Male	65	NA	600073
3946	Prostate cancer	*LRP2*	p.M4601I	Negative	Female	59	NA	600073
3957	Prostate cancer	*LRP2*	p.N1797S	First degree (father)	Male	44	NA	600073
3957	Prostate cancer	*DLC1*	p.D89N	First degree (father)	Male	44	NA	604258
3932	Prostate cancer	*CHEK2*	p.E64K	Negative	Male	47	CM030414	604373
3935	Prostate cancer	*ELAC2*	p.R781H	Negative	Female	70	CM010221	605367
3902	Prostate cancer	*MSR1*	p.H441R	Negative	Female	46	CM023581	153622
3900	Prostate cancer	*MSR1*	p.R293X	Negative	Male	45	CM023579	153622
3954	Prostate cancer	*RNASEL*	p.E265X	Negative	Male	72	CM020300	180435
3954	Prostate cancer	*RNASEL*	p.G59S	Negative	Male	72	CM031342	180435
3963	Retinoblastoma	*RB1*	p.R656W	Negative	Male	48	CM030511	614041
3896	Pituitary cancer	*ACVRL1*	p.A482V	Negative	Female	46	CM994582	601284
3896	Pituitary cancer	*ACVRL1*	p.A482V	Negative	Female	46	CM994582	601284
3930	Esophageal cancer	*WWOX*	p.G178S	Negative	Female	52	NA	605131
3973	Esophageal cancer	*WWOX*	p.R120W	Negative	Male	71	CM016224	605131
3916	Esophageal cancer	*WWOX*	p.R120W	Negative	Male	70	CM016224	605131
3941	Gastric cancer	*MET*	p.A347T	Negative	Male	46	NA	164860

NA, not available.

Table S8. Cardiomyopathy-affected volunteers

Case	Disease	Risk gene	Allele	Clinical	Age (y)	HGMD	OMIM gene ID
3925	Dilated cardiomyopathy	*MYH6*	p.A1443D	Atrial fibrillation	65	CM107536	160710
3926	Cardiomyopathy arrhythmogenic right ventricular	*DSG2*	p.V158G	Arrhythmia	65	CM070921	125671
3935	Dilated cardiomyopathy	*MYH6*	p.R1398Q	Cardiac dysrhythmia	70	NA	160710
3935	Cardiomyopathy, dilated, 1EE	*MYH6*	p.R1398Q	Cardiac dysrhythmia	70	NA	160710
3935	Arrhythmogenic right ventricular cardiomyopathy	*TTN*	p.P3751R	Cardiac dysrhythmia	70	NA	188840
3955	Dilated cardiomyopathy	*ACTN2*	p.Q349L	1° pacemaker	53	NA	102573
3955	Familial hypertrophic cardiomyopathy 12	*CSRP3*	p.R100H	1° pacemaker	53	CM091458	600824
3916	Dilated cardiomyopathy type 1A	*LAMA2*	p.T821M	Stent placement	71	NA	156225
3887	Cardiomyopathy, hypertrophic	*MYBPC3*	p.R326Q	Stent placement (3)	73	CM020155	600958
3887	Cardiomyopathy familial hypertrophic (CMH)	*MYLK2*	p.V402F	Stent placement (3)	73	NA	606566
3953	Brugada syndrome (arrhythmia)	*KCNE3*	p.M65T	Two bypass, stent, and familial history of CAD	71	NA	604433
3953	Arrhythmogenic right ventricular cardiomyopathy	*TTN*	p.P5237T	Two bypass, stent, and familial history of CAD	71	NA	188840
3937	Hypercholesterolaemia	*LDLR*	p.P526H	Three generations of early MI, elevated LDL, cholesterol, triglycerides, and treated with statins	53	CM100938	606945
3890	Hypercholesterolaemia	*LDLR*	p.T726I	1° early MI	57	CM920469	606945
3910	Hypercholesterolaemia	*LDLR*	p.T726I	1° aortic occlusion, elevated cholesterol	51	CM920469	606945
3900	Hypercholesterolaemia	*LDLR*	p.V827I	1° early MI	45	CM920471	606945
3915	Hypercholesterolaemia	*LDLR*	p.V827I	Three generations of elevated cholesterol, treated with statins	70	CM920471	606945

CAD, coronary artery disease; MI, myocardial infarction; NA, not available.

Table S9. Cardiomyopathy unaffected but family history

Case	Disease	Risk gene	Allele	Clinical	Age (y)	HGMD	OMIM gene ID
3943	Arrhythmogenic right ventricular cardiomyopathy	*TTN*	p.G1345D	Familial history of arrhythmia	44	NA	188840
3896	Dilated cardiomyopathy	*SYNE1*	p.L3057V	Familial history	45	NA	608441
3896	Arrhythmogenic right ventricular dysplasia/cardiomyopathy	*JUP*	p.V648I	Familial history	45	NA	173325
3944	Hypertrophic cardiomyopathy	*OBSCN*	p.K1671N	Father	45	NA	608616
3931	Dilated cardiomyopathy	*MYH6*	p.R1398Q	Familial history	46	NA	160710
3907	Cardiomyopathy, hypertrophic	*ACTN2*	p.T495M	Father	47	CM101366	102573
3950	Cardiomyopathy	*MYOM1*	p.G1162S	Familial history	48	NA	603508
3919	Romano-Ward syndrome (arrhythmia)	*SCN5A*	p.S1769N	Familial history	51	CM002391	600163
3889	Romano-Ward syndrome (arrhythmia)	*SCN5A*	p.S1769N	Mother	51	CM002391	600163
3917	Cardiomyopathy	*MYOM1*	p.R1573Q	Familial history + father	51	NA	603508
3960	Dilated cardiomyopathy	*NEBL*	p.K60N	Son CAD	66	CM106905	605491
3976	Cardiomyopathy	*MYOM1*	p.E704K	Older brother	72	NA	603508
3976	Early onset myopathy	*MYH2*	p.V970I	Older brother	72	CM051560	160740

Table S10. Neurodegenerative risk

Case	Disease	Risk gene	Allele	Family history	Age (y)	HGMD	OMIM
3908	Alzheimer's disease	*APOE*	p.C130R	Negative	44	CM900020	107741
3916	Alzheimer's disease	*APOE*	p.L46P	Parkinson 1° (72)	71	CM990167	107741
3954	Alzheimer's disease	*APP*	p.R469H	Negative	72	NA	104760
3942	Frontotemporal dementia	*MAPT*	p.S427F	Negative	71	NA	157140
3954	Frontotemporal dementia	*MAPT*	p.V224G	Negative	72	NA	157140
3895	Parkinson disease	*EIF4G1*	p.G686C	Negative	49	CM117028	600495
3916	Parkinson disease	*EIF4G1*	p.R1205H	Parkinson 1° (78)	64	CM117009	600495
3951	Parkinson disease	*EIF4G1*	p.S1596T	Negative	64	NA	600495
3931	Parkinson disease 11	*GIGYF2*	p.P1222fs	Negative	44	NA	612003
3946	Parkinson disease 11	*GIGYF2*	p.H1171R	Negative	59	NA	612003
3957	Parkinson disease 11	*GIGYF2*	p.M48I	Negative	44	NA	612003
3930	Parkinson disease 11	*GIGYF2*	p.S1035C	Negative	52	NA	612003
3933	Parkinson disease 11	*GIGYF2*	p.S1035C	Negative	68	NA	612003
3928	Parkinson disease	*LRRK2*	p.A419V	Tremor 1° Parkinson 2°	68	CM125746	609007
3903	Parkinson disease	*LRRK2*	p.D972G	Negative	54	NA	609007
3919	Parkinson disease	*LRRK2*	p.D972G	Negative	51	NA	609007
3889	Parkinson disease	*LRRK2*	p.G2019S	Negative	51	CM050659	609007
3951	Parkinson disease	*LRRK2*	p.L119P	Negative	50	NA	609007
3918	Parkinson disease	*LRRK2*	p.L286V	Negative	64	NA	609007
3907	Parkinson disease	*LRRK2*	p.P1542S	Alzheimer's 2°	47	NA	609007
3935	Parkinson disease	*LRRK2*	p.P1542S	Negative	70	NA	609007
3893	Parkinson disease	*LRRK2*	p.R1514Q	Negative	45	CM057190	609007
3943	Parkinson disease	*LRRK2*	p.R1514Q	Negative	50	CM057190	609007
3949	Parkinsonism, juvenile, autosomal recessive	*PARK2*	p.R275W	2° three siblings	52	CM991007	602544
3924	Parkinsonism, juvenile, autosomal recessive	*PARK2*	p.R334C	Negative	54	CM003865	602544
3927	Parkinson	*PM20D1*	p.A332V	Negative	73	NA	613164
3886	Parkinson	*PM20D1*	p.P281Q	Negative	62	NA	613164

Table S11. Percentage of survey respondents reporting having made behavioral changes specifically motivated by their test results

Type of behavior change	Yes	No
Changes to diet	4 (10%)	36 (90%)
Changes to health care (such as undergoing tests or seeing a specialist)	4 (10%)	36 (90%)
Changes to use of vitamins/herbal supplements	4 (10%)	36 (90%)
Changes to exercise	3 (8%)	37 (92%)
Changes to medications	1 (2%)	39 (98%)
Changes to insurance coverage	1 (2%)	39 (98%)
Number of respondents making at least one of the above behavior changes	10 (25%)	