Pathogenic Variants for Mendelian and Complex Traits in Exomes of 6,517 European and African Americans: Implications for the Return of Incidental Results

Holly K. Tabor,[1,2,9] Paul L. Auer,[3,4,9] Seema M. Jamal,[2] Jessica X. Chong,[2] Joon-Ho Yu,[2] Adam S. Gordon,[5] Timothy A. Graubert,[6] Christopher J. O'Donnell,[7] Stephen S. Rich,[8] Deborah A. Nickerson,[5] NHLBI Exome Sequencing Project, and Michael J. Bamshad[2,5,*]

Exome sequencing (ES) is rapidly being deployed for use in clinical settings despite limited empirical data about the number and types of incidental results (with potential clinical utility) that could be offered for return to an individual. We analyzed deidentified ES data from 6,517 participants (2,204 African Americans and 4,313 European Americans) from the National Heart, Lung, and Blood Institute Exome Sequencing Project. We characterized the frequencies of pathogenic alleles in genes underlying Mendelian conditions commonly assessed by newborn-screening (NBS, n = 39) programs, genes associated with age-related macular degeneration (ARMD, n = 17), and genes known to influence drug response (PGx, n = 14). From these 70 genes, we identified 10,789 variants and curated them by manual review of OMIM, HGMD, locus-specific databases, or primary literature to a total of 399 validated pathogenic variants. The mean number of risk alleles per individual was 15.3. Every individual had at least five known PGx alleles, 99% of individuals had at least one ARMD risk allele, and 45% of individuals were carriers for at least one pathogenic NBS allele. The carrier burden for severe recessive childhood disorders was 0.57. Our results demonstrate that risk alleles of potential clinical utility for both Mendelian and complex traits are detectable in every individual. These findings highlight the necessity of developing guidelines and policies that consider the return of results to all individuals and underscore the need to develop innovative approaches and tools that enable individuals to exercise their choice about the return of incidental results.

Introduction

Exome sequencing and whole-genome sequencing (ES/WGS) are highly effective tools for identifying the genetic basis of heritable disorders, and their use in clinical settings is rapidly increasing.[1–3] Aside from identifying the variant(s) underlying the disorder for which sequencing was performed (i.e., the primary result), ES/WGS can identify other variants (i.e., secondary or incidental results) that could be of clinical and/or personal (e.g., reproductive planning or lifestyle modification) utility.[4] The ability to identify incidental results has led to considerable controversy about whether such variants should be reported to individuals and/or families in clinical settings and research participants and, if so, how they should be offered for return.[5–10] Clinical utility is defined as the existence of established therapeutic interventions or actions that have the potential to change the clinical course of disease,[11] whereas personal utility is more broadly debated and can include reproductive decision making, lifestyle changes with a less direct impact on disease risk, and life planning.[12] Several recent studies have demonstrated that personal utility is an important consideration for individuals in clinical care and research participants in making decisions about the return of genetic results, given that the vast majority of individuals indicate that they want access to a broad range of results beyond those with a direct clinical benefit.[13–15] Recently published results of survey responses from ~900 genetic professionals suggest that genetics care providers also value the potential benefit of results of personal utility and think they should be offered for return.[16]

Frequently, the debate about the return of incidental results from ES/WGS has been framed by major gaps in knowledge: (1) the frequency with which incidental results that prompt clinical action (i.e., changes in medical management) are identified in ES/WGS data[5,6] and (2) the extent to which the process of curating and reporting such variants is beyond what is required to answer a clinical question.[5] Yet, how much weight these issues should be given has been informed to only a very limited extent by empirical data on the effectiveness of curating large ES/WGS data sets for well-known pathogenic risk variants, the proportion of individuals who have known risk variants for disease, and the number and kinds of risk variants (i.e., burden) found in each individual.[10,17–22]

Most recent attempts to estimate the number of predicted or known pathogenic variants in subsets of genes

[1]Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA 98101, USA; [2]Department of Pediatrics, University of Washington, Seattle, WA 98195, USA; [3]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [4]School of Public Health, University of Wisconsin–Milwaukee, Milwaukee, WI 53201, USA; [5]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [6]Department of Medicine, Harvard University, Boston, MA 02114, USA; [7]Cardiovascular Epidemiology and Human Genomics Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Framingham, MA 01702, USA; [8]Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908, USA
[9]These authors contributed equally to this work
*Correspondence: mbamshad@u.washington.edu
http://dx.doi.org/10.1016/j.ajhg.2014.07.006.

underlying various phenotypes in certain populations[18,23] have focused on identifying large-effect variants with high clinical impact for very rare autosomal-recessive and autosomal-dominant conditions.[10,17–23] The prior probability of finding risk variants that underlie most rare Mendelian conditions is very low. This fosters screening of relatively large numbers of genes with little expectation that variants of known pathogenic significance will be found in many individuals. For example, recommendations from the American College of Medical Genetics and Genomics (ACMG) on the return of incidental results from 56 genes captured by exome sequencing suggest that only 1% of individuals will have a result to return.[24] Moreover, the large-effect alleles that underlie Mendelian conditions represent a small fraction of the risk variants that could be of utility for an individual. In other words, risk variants underlying more common complex diseases also need to be considered for return. With this in mind, it has been emphasized that every individual most likely has an ES/WGS result(s) that could be offered for return.[25] Yet, there is scant empirical evidence from assessments of large ES/WGS data sets to support this claim.

To characterize the broad range of risk alleles that could be offered for return in individuals who undergo ES/WGS and to better frame the debate about the return of incidental results, we sought to address four major questions about the spectrum of genetic results of potential utility, both clinical and personal, in both individual exomes and collectively in a large set of exomes sampled from European American (EA) and African American (AA) individuals: for genes associated with common Mendelian conditions and common complex diseases, (1) what is the per-individual burden of pathogenic risk variants, (2) what is the cumulative burden of risk variants per individual, (3) how well do computational predictors of deleterious variants identify known disease-risk variants, and (4) what are the implications of estimates of per-individual burden of variants of utility for the management of incidental results from ES/WGS?

We assessed protein-coding variants in health-related genes in 6,517 individuals enrolled in the National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP). The ESP data set consists of high-coverage (average median depth ~ 90×) exome sequences from 2,204 AA and 4,313 EA individuals sampled from more than 20 cohorts with heart and lung phenotypes unrelated to those conditions assessed herein (Table S1, available online).[26]

In this ESP data set, we identified pathogenic variants in three broad categories that represent the spectrum of potential utility and across a range of allele frequencies and disease prevalences. The first category included 39 genes underlying 31 Mendelian conditions, most of which are inborn errors of metabolism, recommended for newborn screening (NBS) (Table S2) by the Discretionary Advisory Committee on Heritable Disorders in Newborns and Children (formerly known as the Secretary's Advisory Committee on Heritable Disorders in Newborns and Children) Recommended Uniform Screening Panel (Table S2). NBS results in adults are primarily of personal utility because they can be used for reproductive planning.

The second category included 17 genes associated with risk of age-related macular degeneration (ARMD) (Table S2). ARMD is a common complex condition with onset late in life and a relatively high frequency and is the third leading cause of vision loss worldwide.[27] Many healthy adults are likely to carry ARMD risk variants, and at present there are limited interventions to prevent or mitigate ARMD. Therefore, ARMD results are primarily of utility for potential early risk identification, improved diagnosis, and delay of onset and/or mitigation of symptoms through lifestyle modifications (e.g., diet, vitamin supplementation, and smoking cessation), as well as life planning. In the future, such results might allow for risk stratification and targeted interventions.[27,28]

The third category included 14 genes containing variants known to influence drug response (PGx) (Table S2). PGx results are primarily of utility for the avoidance of adverse reactions or lack of response to certain medications across a range of clinical conditions or settings. Our selection criteria for this set depended upon the higher levels of evidence of association in the PharmGKB database, including replicated studies with significant p values and/or medical-society-endorsed PGx guidelines. Although the relative risks for these variants vary, their clinical utility and functional significance are well documented. Consequences for these variants range from severe allergic reactions to a lack of response to a range of drugs, including blood-thinning agents, antipsychotic medications, and chemotherapeutic agents.[29–31]

We annotated all variants in each of these 70 genes with commonly available computational-prediction approaches and manually curated them to determine whether they were pathogenic. We subsequently compared the effectiveness of computational predictions of known risk variants and generated individual and population-specific estimates of the burden of risk variants for Mendelian and complex traits in the NBS, ARMD, and PGx gene sets.

Material and Methods

Exome Sequencing and Variant Calling

Library construction, exome capture, sequencing, and mapping were performed as previously described.[26] Sequencing was performed at the University of Washington (UW) and the Broad Institute of MIT and Harvard (Broad). In brief, exome capture was performed with the Roche Nimblegen SeqCap EZ Library or the Agilent SureSelect Human All Exon 50 Mb Kit. Paired-end sequencing (2 × 76 bp) was performed on Illumina GAII and HiSeq instruments. Single-nucleotide variants (SNVs) were called with a maximum-likelihood approach[32] implemented in the UMAKE pipeline at the University of Michigan, which allowed all samples to be analyzed simultaneously for both variant calling and filtering. BAM[33] files summarizing Burrows-Wheeler Aligner[34] alignments generated at the UW and the Broad were used as input.

BAM files summarized alignments mapped to the GRCh37 human reference sequence and were refined by duplicate removal, recalibration, and indel realignment with the Genome Analysis Toolkit.[35] We excluded all reads that were not confidently mapped (Phred-scaled mapping quality < 20) from further analysis. We then used SAMtools[33] to compute genotype likelihoods for exome targeted regions and 50 flanking bases while accounting for per-base alignment quality. We identified variable sites and their allele frequencies by using a maximum-likelihood model implemented in glfMultiples.[32] These analyses assumed a uniform prior probability of polymorphism at each site. The final call set was performed on 6,823 samples.

We used a support vector machine (SVM) classifier to separate most likely true-positive and false-positive variant sites with the use of various quality metrics. These included allelic balance (the proportional representation of each allele in likely heterozygotes), base-quality distribution for sites supporting the reference and alternate alleles, and the distribution of supporting evidence between strands and sequencing cycle, among others. We used variants identified by dbSNP or 1000 Genomes[36] as the positive training set and used variants that failed multiple filters as the negative training set. This method was effective at removing sequencing artifacts while preserving good-quality data, as indicated by the transition-transversion ratio for previously known and newly identified variant sites, the proportion of high-frequency variants overlapping with dbSNP, and the ratio of synonymous to nonsynonymous variants, as well as attempts at validating a subset of sites. A total of 1,908,614 SNVs passed the SVM filter.

In order to obtain high-quality genotypes for estimating allele frequency and carrier burden, we set individual genotypes to missing if the corresponding read depth was <10×. We also excluded variants with read depth > 500×. With the exception of one outlier with a low call rate, all samples had high call rates. The outlier was removed from the analysis.

Identification of Related Individuals and Ancestry Designation

We used principal-component analysis (PCA) to assign genetic ancestry. Only autosomal SNVs with a minor allele frequency (MAF) ≥ 0.1% and a call rate > 95% were included in the PCA, which we executed in PLINK[37] after pruning SNVs in linkage disequilibrium. The first two principal components clearly separated AA samples from EA samples (Figure S1). Thirty individuals of indeterminate genetic ancestry (i.e., located between the two vertical lines in Figure S1) were removed.

After designating samples to AA and EA ancestry groups, we used KING[38] to run a kinship analysis stratified by genetic-ancestry assignment to identify cryptically related individuals. Pairs of samples with a kinship-coefficient range of >0.354, [0.177, 0.354], [0.0884, 0.177], and [0.0442, 0.0884] were designated as duplicates, first-degree relatives, second-degree relatives, and third-degree relatives, respectively. Only the sample with the higher overall call rate from each duplicate or relative pair was retained. This resulted in the removal of 274 samples from analysis. We also estimated inbreeding coefficients for each pair of samples and removed one sample with a value higher than that of all other samples. This resulted in a final data set of 2,204 AA and 4,313 EA samples (total = 6,517), hereafter referred to as ESP6500.

The 6,517 ESP samples contained 422 samples from cystic-fibrosis (CF [MIM 219700])-affected individuals who were known carriers for pathogenic *CFTR* mutations. We excluded these 422 CF individuals from the *CFTR* carrier-burden and risk-variant frequency estimates described in the Results.

Annotation and Selection of Risk Variants

Variants that were polymorphic and had a call rate ≥ 90% in ESP6500 were submitted to the SeattleSeq Annotation server on May 29, 2012. We used SeattleSeq Annotation 134, the hg19 build of the human reference genome (UCSC Genome Browser), and the NCBI full genes (NM, XM) gene-model option. For variants mapping to multiple transcripts, we retained the most damaging classification (from most to least damaging: nonsense, splice, missense, synonymous, UTR, other). PolyPhen-2[39] scores were obtained from the PolyPhen-2 server with the use of the HumDiv classifier model. Combined Annotation-Dependent Depletion (CADD) scores were obtained from the CADD server.[40]

All annotated variants from the 70 genes selected for assessment were entered into a spreadsheet for curation. Each variant was manually compared to variants in the Human Gene Mutation Database (HGMD Professional 2012.3) and Online Mendelian Inheritance in Man (OMIM) by either rsID or by protein position. Variants not found in OMIM or HGMD in the PGx and ARMD gene sets and variants not found in OMIM or HGMD as either "disease mutation" (DM) or "disease mutation?" (DM?) in the NBS gene set were eliminated from further consideration (i.e., exclusion step 1; Table S3 and Figure S2). Next, we eliminated variants listed in HGMD or OMIM but associated with a phenotype that differed from the one for which screening was intended, those listed as polymorphic in one of 27 locus-specific databases (LSDBs), and those not in PharmGKB (Tables S2 and S4; exclusion step 2 in Figure S2). We then excluded variants with a call rate < 90% in either EA or AA individuals (exclusion step 3; Table S3 and Figure S2). Lastly, we excluded ten NBS variants that were homozygous in the absence of compelling evidence, via literature review, of disease causality and three NBS variants that had an allele frequency > 3% (suggesting that they were polymorphisms rather than risk variants for a rare Mendelian disorder and/or had recently been shown not to be pathogenic); excluded PGx variants that had the lowest level of evidence of functional significance (i.e., level 4) in PharmGKB; and excluded ARMD variants for which clinical testing was not available per GeneTests (exclusion step 4; Table S3 and Figure S2). We included all nonsense variants that had not been observed previously, all of which were singletons.

The institutional review boards at all participating sites approved the NHLBI ESP study protocols.

Results

Curation of Variants

We identified a total of 10,879 different variants (i.e., candidate-variant set) in 70 disease-related genes in the 6,517 ESP samples. A total of 5,155 variants in AA individuals and 5,724 variants in EA individuals were found. As anticipated, manual comparison of the set of candidate variants to HGMD and OMIM resulted in a substantial reduction in the number of variants retained for further review; 91%, 92%, and 90% of variants were excluded in the total sample, in AA individuals, and in EA individuals, respectively (Table S3). A larger fraction of variants were excluded in the ARMD (94%) and PGx (94%) gene sets

than in the NBS gene set (86%) partly because of the retention of a larger fraction of rare variants in the NBS gene set.

To evaluate the extent to which the curation process enriched the set of risk variants with those likely to confer disease risk, we compared metrics of evolutionary conservation, pathogenicity, and deleteriousness between the set of risk variants and the variants that were excluded from further analysis. A Genome Evolutionary Rate Profiling (GERP) score is a measure of evolutionary constraint, such that high GERP scores reflect the strength of past purifying selection.[41] Heuristically, variants with high GERP scores are more likely to exert phenotypic consequences than variants with low GERP scores. Mean GERP scores of variants in the risk set were significantly higher than the mean scores of the excluded variants (Figure 1) for NBS ($p = 1.45 \times 10^{-86}$) and ARMD ($p = 4.40 \times 10^{-4}$) gene sets, but not for the PGx gene set ($p = 0.066$).

PolyPhen-2 scores are a measure of the impact of an amino acid substitution on the structure and function of a protein.[39] Variants with PolyPhen-2 scores nearer to 1 are predicted to be damaging, whereas scores close to 0 are less likely to affect protein function. There were significantly more damaging variants in the NBS set of risk variants than among excluded variants ($p = 1.99 \times 10^{-19}$; Figure 1). The difference was not significant for the ARMD ($p = 0.42$) or PGx ($p = 0.25$) variant sets.

CADD scores provide another means of prioritizing variants. Larger CADD scores correspond to variants that are more likely to be deleterious in terms of overall organismal fitness.[40] CADD scores are available for all 8.6 billion possible SNVs in the human genome and are Phred scaled. There were significantly more damaging variants in the NBS set of risk variants than among excluded variants ($p = 1.38 \times 10^{-109}$; Figure 1). The difference was still significant but less striking for ARMD ($p = 9.9 \times 10^{-4}$) and PGx ($p = 1.47 \times 10^{-9}$) variant sets, consistent with the prediction that variants underlying rare, Mendelian disorders are highly deleterious whereas variants for complex phenotypes (e.g., ARMD) and phenotypes with low to modest effects on fitness (e.g., PGx) are less deleterious.

Whereas the percentage of variants retained in the ARMD and PGx variant sets was similar between AA and EA individuals, the percentage of variants retained in the NBS gene set was substantially higher in EA individuals (15%) than in AA individuals (10%). This observation raised the possibility that a greater fraction of true risk variants in AA individuals were being excluded as a result of underrepresentation of pathogenic rare variants in AA individuals in HGMD and/or OMIM.[42] If this was the case, one prediction is that compared to rare variants excluded in EA individuals, rare variants excluded in AA individuals should be enriched with predicted deleterious variants. However, the mean GERP, PolyPhen-2, and CADD scores did not differ significantly.

Because the allele frequencies of variants that cause inborn errors of metabolism assessed via NBS are generally well known, we compared the allele frequency of each NBS risk variant with a MAF > 1% in ESP6500 to estimates from 1000 Genomes data. Allele frequencies were similar except for rs334, the variant underlying sickle cell disease (MIM 603903), which had a MAF of only 4.0% in ESP6500 but an MAF of 9.0% in the 1000 Genomes data (Table S5).

We also estimated carrier frequencies of many of the conditions in the NBS gene set in order to compare our estimates to those previously published. For several conditions, our observed carrier frequencies were higher than published estimates (Table S6). Our original strategy for curating variants included HGMD, OMIM, and LSDB variants that could be identified during NBS but that might not cause a phenotype (e.g., pseudodeficiency or partial-deficiency alleles) requiring treatment. As a consequence, our observed carrier-frequency estimates for some NBS conditions were inflated. For this reason, we performed an additional step of manually reviewing variants listed in HGMD as DM? (n = 12); variants present as homozygotes in either EA or AA individuals and/or variants with a MAF > 0.5% in either EA or AA individuals (n = 15) (Table S7); all variants in *CFTR* (MIM 602421) (n = 60); and *PAH* (MIM 261600) variants listed in HGMD as "hyperphenylalaninemia" (MIM 261600) (n = 13). Hyperphenylalaninemia is detected by NBS but is not associated with clinical or symptomatic phenylketonuria (MIM 261600). Detailed methods of this additional review are described in Table S6, and the revised carrier frequencies are shown in Table S7.

For the majority of examined NBS conditions, our original and revised carrier-frequency estimates were similar to, and often lower than, population-based estimates of carrier frequencies (Figure 2). For example, our review of variants for congenital adrenal hyperplasia (MIM 201910) eliminated the only two potentially pathogenic variants in ESP6500. This resulted in a carrier frequency of 0, which is unrealistic given that the published carrier frequency is 1/61. This result suggests that our additional review to attempt to generate more accurate estimates of carrier frequencies was too strict for many conditions. This observation is consistent with the fact that pathogenic copy-number variants (CNVs), indels, and variants outside of the target were not included in the ESP6500 data set.

Distribution of Risk Variants

In the overall EA and AA samples, a total of 399 validated risk variants (i.e., risk-variant set) were identified in 6,517 individuals in the NBS gene set (n = 328), the ARMD gene set (n = 25), and the PGx gene set (n = 46) (Table S8). The majority (87%) of variants were rare (MAF ≤ 0.5%), and as expected given the predicted phenotypic consequences of risk variants in each gene set, there was a higher percentage of rare variants in the NBS gene set (96%) than in either the ARMD set (68%) or the PGx set (39%). As a corollary, the site-frequency spectrum of risk variants underlying PGx and ARMD phenotypes included a higher percentage of variants of intermediate (1% < MAF < 5%) and high (MAF ≥ 5%) frequency (Figure S3).

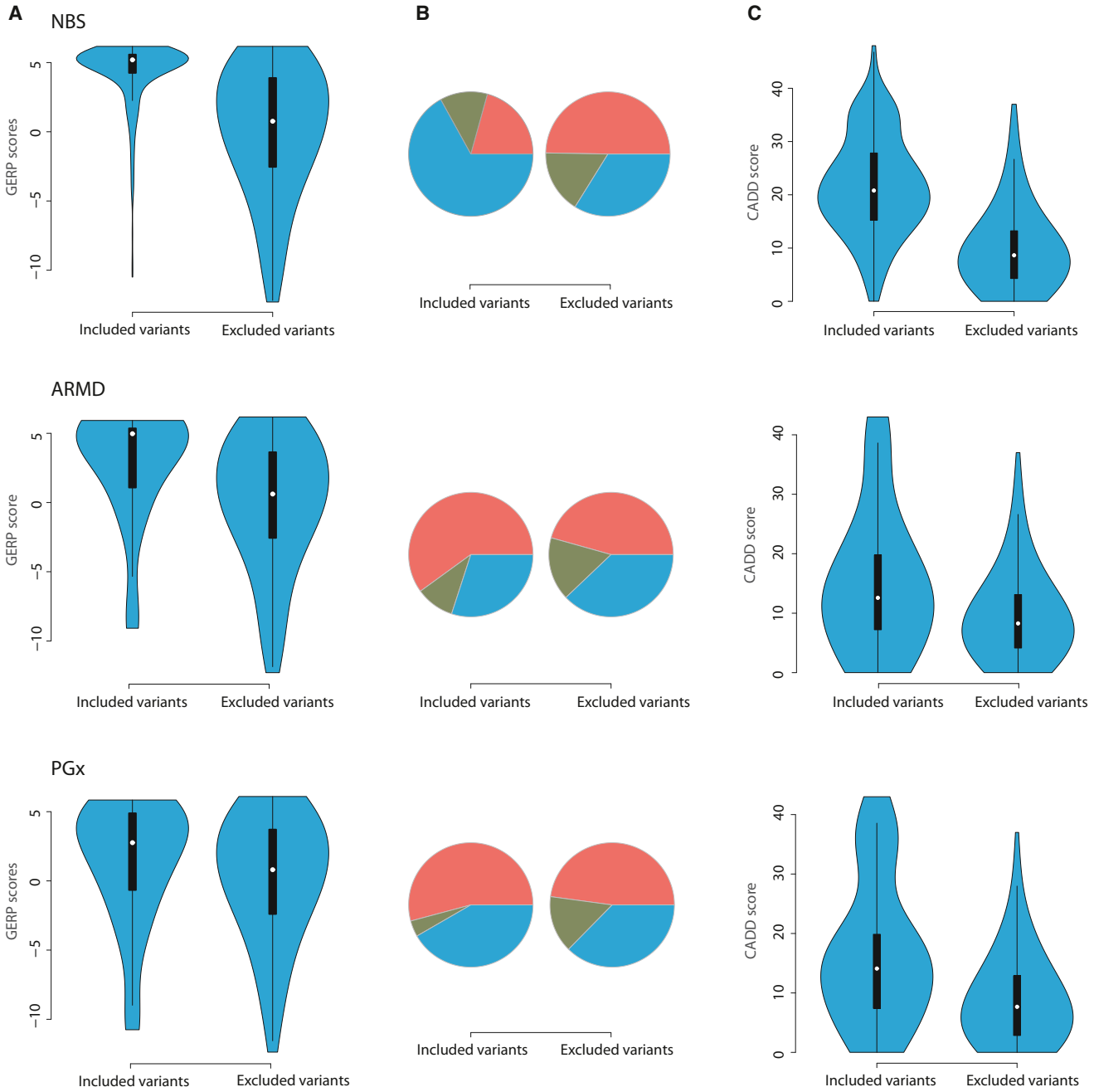

Figure 1. Extent to which the Variant-Curation Process Enriched the Risk-Variant Set with Those that Most Likely Confer Disease Risk
Violin plots showing the distribution of metrics of conservation (A), pathogenicity (B), and deleteriousness (C) between included and excluded variants in the NBS, ARMD, and PGx gene sets.
(A) Mean GERP scores of variants in the risk set were significantly higher than the mean scores of the excluded variants for the NBS ($p = 1.45 \times 10^{-86}$) and ARMD ($p = 4.40 \times 10^{-4}$) gene sets, but not for the PGx gene set ($p = 0.066$).
(B) Significantly more variants were predicted to be probably damaging (blue) by PolyPhen-2 in the NBS ($p = 1.99 \times 10^{-19}$) and ARMD ($p = 0.42$) gene sets than in the PGx gene set ($p = 0.25$). The fraction of variants predicted to be benign (red) or possibly damaging (green) is denoted.
(C) With the use of CADD scores, significantly more damaging variants were found in the NBS ($p = 1.38 \times 10^{-109}$), ARMD ($p = 9.9 \times 10^{-4}$), and PGx ($p = 1.47 \times 10^{-9}$) risk-variant sets.

A total of 309 and 197 risk variants were identified in EA and AA individuals, respectively. The higher number of risk variants in EA individuals was partly due to the fact that a larger number of EA individuals (n = 4,313) than AA individuals (n = 2,204) were sequenced. The overall distribution of risk variants of rare, intermediate, and high frequency in both EA and AA individuals was similar to that observed for the total sample. The percentage of rare risk variants in each gene set was slightly higher in EA than in AA individuals, and a higher fraction of rare

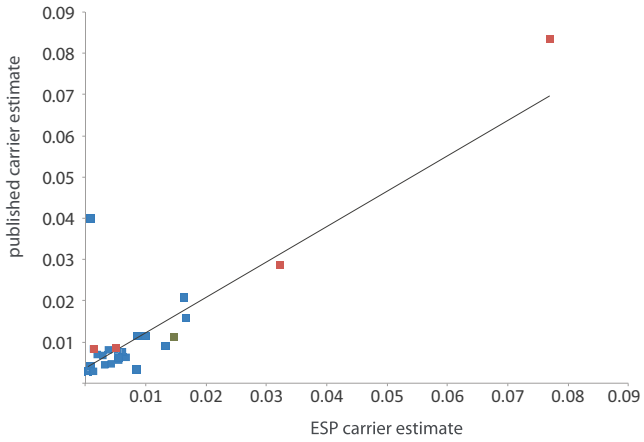

Figure 2. Comparison of Published versus Observed Carrier Estimates for NBS Conditions
Published estimates of carrier frequencies were similar to estimates on the basis of a conservative manual review of NBS risk variants. Carrier estimates for sickle cell anemia (MIM 603903), beta-thalassemia (MIM 613985), and hemoglobin C disease (MIM 141900.0038) were calculated in AA individuals separately (red), and estimates for CF (MIM 219700) were calculated in AA (red) and EA (green) individuals separately, whereas all other estimates were calculated in the total ESP6500 sample (blue).

risk variants in each gene set in EA individuals had never been observed. Most risk variants that were of rare or intermediate frequency were specific to EA or AA individuals; however, most of the risk variants of intermediate frequency for ARMD were shared by both AA and EA individuals.

Burden of Risk Alleles per Individual

The mean number of risk alleles per individual in the NBS gene set was 0.57 (0.60 in EA individuals and 0.51 in AA individuals). This diminished to 0.18 overall (0.17 in EA individuals and 0.22 in AA individuals) if we excluded all variants that are typically reported and therefore of potential personal utility but for which it is quite unclear whether they are truly pathogenic. In both EA and AA individuals, 45% had at least one and 11% had at least two pathogenic alleles in the NBS gene set (Figure 3A). The maximum number of pathogenic alleles per individual was four in AA individuals (n = 2) and five in EA individuals (n = 1). Two pathogenic variants of intermediate or higher frequency accounted for the majority (68%) of pathogenic alleles found in each individual. The most common pathogenic variant (allele frequency = 7.2% [9.4% in EA individuals and 2.9% in AA individuals]) in the NBS gene set was c.940A>G (p.Asn314Asp) (rs2070074, or the Duarte allele) in *GALT* (MIM 606999), mutations in which cause classical galactosemia (MIM 230400). This was the only risk variant with an allele frequency > 5% in EA individuals in the NBS gene set. The second most common pathogenic variant (allele frequency ~ 4.2%) in EA individuals was c.1330G>C (p.Asp444His) (rs13078881) in *BTD* (MIM 609019), mutations in which underlie biotinidase deficiency (MIM 253260). The most common variant (allele frequency = 4.0%) in AA individuals was c.20T>A (p.Glu7Val) (rs334) in *HBB* (MIM 141900), mutations in which cause sickle cell anemia (MIM 603903).

The mean number of risk alleles per individual in the ARMD gene set was 3.16 (3.00 for EA individuals and 3.48 for AA individuals). Almost every individual (i.e., 99.98%) carried at least one risk allele, and more than half carried more than three risk alleles in the ARMD gene set (Figure 3B). Four variants had allele frequencies > 20% in EA individuals, and three variants in AA individuals had allele frequencies > 20%. These four variants were located in either *C3* (MIM 120700) or *CFH* (MIM 134370) (Table S8). The maximum number of ARMD risk alleles per person was seven in both EA (n = 2) and AA (n = 1) individuals.

The mean individual burden of known functional alleles in the PGx gene set was 11.5 (11.1 for EA individuals and 12.3 for AA individuals; Figure 3C). Each individual carried at least five known functional PGx alleles, and every AA individual carried at least six. The maximum number of known functional alleles per individual in the PGx gene set was 22 in EA individuals (n = 1) and 20 in AA individuals (n = 4). The frequency spectrum of PGx risk variants was fairly flat, suggesting that both rare and common variants contributed to the overall burden of risk (Figure S3; Table S8).

Across the NBS, ARMD, and PGx gene sets, the mean burden of risk alleles per individual was 15.3 (14.7 for EA individuals and 16.3 for AA individuals; Figure 3D). The overall estimate of burden was largely unchanged with the use of the more conservative NBS variant list (14.9: 14.3 in EA individuals and 16.0 in AA individuals). The range of risk alleles per individual varied widely. The minimum was five for EA individuals (n = 1) and seven for AA individuals (n = 1), and the maximum was 26 in both EA (n = 2) and AA (n = 1) individuals. These estimates represent a "snapshot" of the total burden of risk alleles to be found in any single individual but underscore the contention that every individual has alleles of potential clinical utility.

Carrier Burden for Severe Recessive Disorders of Childhood

Over the past several decades, attempts have been made to estimate the burden of deleterious variants per individual with the use of computational and/or population-genetic approaches.[20,21,43] More recently, next-generation sequencing has enabled screening of relatively large numbers of genes that underlie disease-related traits in hundreds to thousands of individuals for estimating empirical burden.[19] One estimate that has been widely cited is that by Bell et al.,[19] who reported that individuals carry, on average, 2.8 alleles for severe (i.e., causing substantial morbidity or reduced lifespan) autosomal-recessive pediatric disorders on the basis of a sampling of 437 genes in

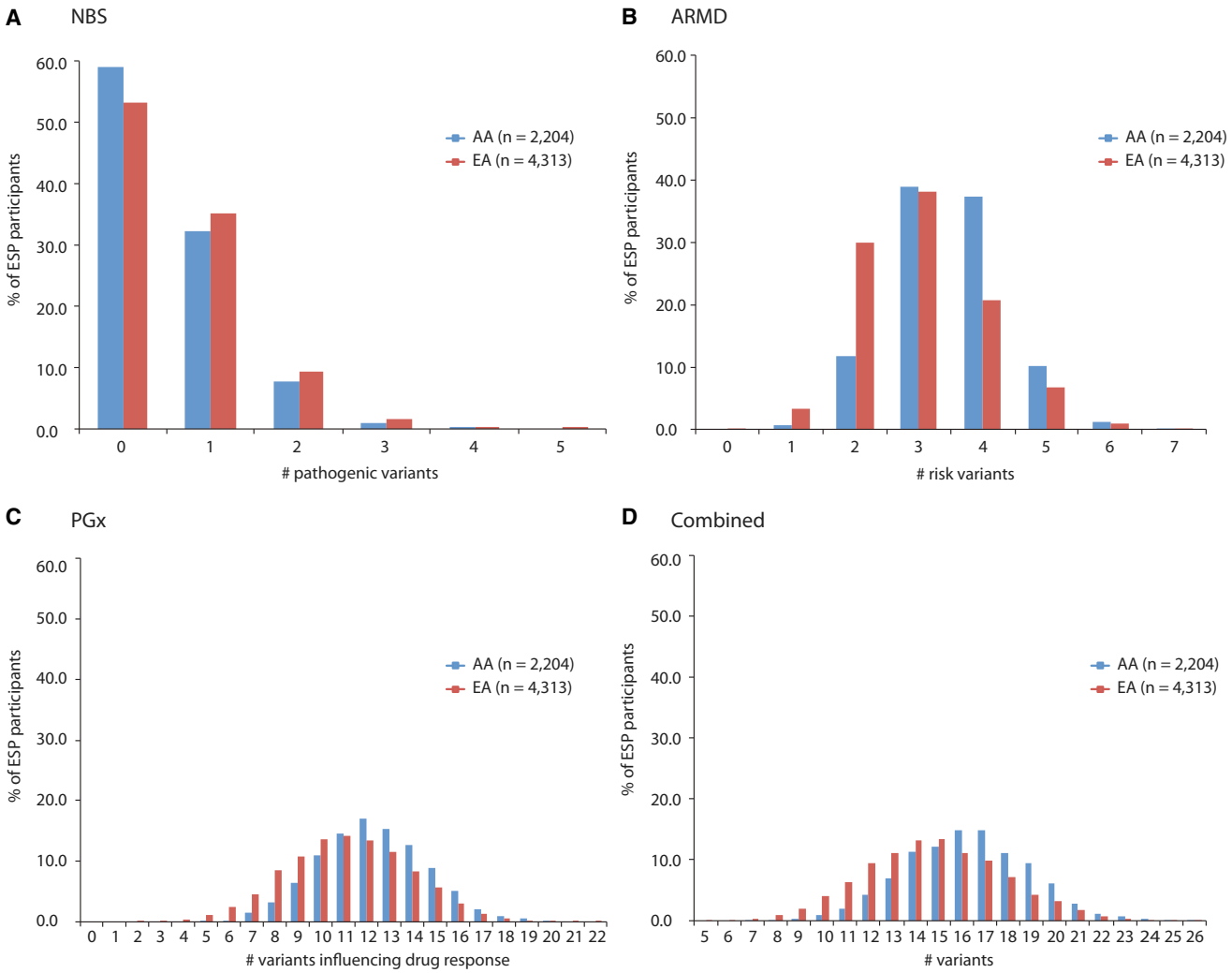

Figure 3. Plots of the Number of Pathogenic or Risk Variants per Individual
AA individuals are in blue, and EA individuals are in red. Plots are shown for NBS (A), ARMD (B), and PGx (C) gene sets and a combination across gene sets (D).

104 individuals. Their criteria for assessing whether a variant conferred risk of disease were conservative in that they excluded HGMD variants not listed as "DM," variants with MAF $\geq$ 5%, and variants observed to be homozygous. Yet, 76 of the individuals sampled by Bell et al. were either known to be carriers or affected by a severe childhood recessive disorder, and the authors did not adequately control for the excess of pathogenic alleles in those individuals.

Using the same approach as Bell et al.[19] but limiting the individuals assessed to the 26 "controls" in their sample and the 35 genes with "severe pediatric recessive mutations," we found that the burden of risk alleles, excluding indels, was 1.73 instead of 2.8. This result suggests that their original calculation was biased upward in part by their inclusion of known affected and carrier individuals. Moreover, a sample size of 26 individuals is very small and prone to sampling bias. We therefore repeated the analysis for 10,000 subsamples of 26 ESP6500 individuals, excluding CF individuals from the calculations involving *CFTR*, to generate a distribution of 10,000 estimates of the carrier burden of severe pediatric recessive mutations. None of the average carrier burdens were close to the 1.73 we obtained by using their controls. Specifically, the average number of risk alleles per individual was 0.68 (SD = 0.17), and the maximum was 1.5. Results from our curation strategy on the same 35 genes were similar. The average carrier burden was 0.69 (SD = 0.16), and the maximum was 1.5. Accordingly, we think that ~0.7 is most likely a more reliable estimate of the burden of risk alleles for severe childhood recessive disorders.

Discussion

Each of the 70 genes that we assessed contained known pathogenic variants, and every individual had several alleles (with potential clinical utility) that could be offered for return. This is an often-stated result with little, if any,

empirical support. Moreover, our results most likely underestimate the number of risk alleles in each of the three gene sets (NBS, ARMD, PGx), much less each individual, because we did not have the opportunity to include pathogenic indels (e.g., *ΔF508-CFTR*) in our analysis, and accurate curation was limited by the quality and incompleteness of existing risk-variant databases, specifically HGMD, OMIM, and LSDBs. These databases are imperfect because estimates of allele frequencies in populations of non-European ancestry are often incomplete, and it is challenging to populate databases efficiently with up-to-date curated variant information. Indeed, the development of accurate and accessible variant databases will be critical to efforts to report both primary and incidental results from ES/WGS.

Our estimate of the carrier burden for severe recessive childhood disorders was 0.57, significantly less than the estimate of 2.8 variants per person for severe autosomal-recessive disorders reported by Bell et al.[19] Although the genes we analyzed were not identical, we demonstrated that their estimate was biased upward because of their small sample size and by the inclusion of individuals who were ascertained because they were affected by a Mendelian disorder. The relatively large size of the ESP6500 data set enabled more accurate estimates of variant frequencies and improved detection of rare variants, at least in populations of European and African ancestry.

Our findings challenge the assumption that actionable incidental results, much less incidental results of potential clinical utility, in ES/WGS are rare. Indeed, they are rare only to the extent that the search is limited to those most likely affected by highly penetrant phenotypes. Empirical confirmation that incidental results of potential clinical utility are common in ES/WGS data is important because ES/WGS are being used in a variety of healthcare settings, including the evaluation of children with birth defects, developmental delay, or autism; the assessment of family risk of cancer and cancer profiling for personalizing therapeutics and delivering precision medicine; and the evaluation of adult-onset conditions, such as cardiomyopathy. Our results bring into better focus both the potential benefit of offering incidental results from ES/WGS to study participants, individuals, and families undergoing clinical care and the potential reporting burden that could be borne by clinicians. These findings, coupled with professional attitudes and concerns about the return of results,[16] highlight the potential need for future recommendations and guidelines for considering the return of results to broader groups of individuals, including "healthy" individuals.

Our findings also undermine the notion that curating and/or reporting known risk alleles is an added burden to a service lab beyond what is required to answer a clinical question. Manual curation of the variants in ESP6500 was, of course, labor intensive, but annotating ES/WGS data for the 399 validated risk variants was computationally trivial. Development of comprehensive databases of validated risk variants is a major goal of multiple national and international efforts, and using such databases to annotate ES/WGS data in order to identify a primary result is standard practice in exome service laboratories.

Existing approaches to returning genetic results are not optimal for the return of large numbers of results with the breadth and scope (i.e., that influence a wide spectrum of phenotypes across the life of an individual) that we have demonstrated could be offered for return. Whether to return all or some incidental results and, if only some, which results to return, how to most effectively return them, and how best to use them to benefit health are unknown. The lack of such information is a major gap in knowledge in empowering individuals and families to translate genomic information into improved lifestyles, medical care, and ultimately long-term health. This problem is compounded further by challenges to the traditional process of communicating and disseminating genetic results.

Traditionally, genetic and genomic test results are usually communicated from provider to individual or family via mail, phone, or face-to-face contact. However, the volume of results and information is typically far in excess of staffing allocations at many institutions, resulting in long delays for the return of result information (National Society of Genetic Counselors 2014 Professional Status Survey Executive Summary).[44,45] This delay and the resultant inability to act on the information can compromise patient safety, increase the risk of a poor clinical outcome, and disenfranchise families. Furthermore, with existing approaches to returning results, there are often few, if any, convenient and inexpensive ways for a provider to ensure that an individual and/or family has received and understood a result. This situation will continue to worsen, perhaps exponentially, as the use of ES/WGS in both research and clinical settings continues to expand. The countervailing concern is that the return of such incidental results might increase unnecessary follow-up costs.[46] Although this would certainly be a concern in the context of a policy mandating the return of results, this empirical question remains open in the context of offering secondary results that hold a range of clinical and personal utility.

A number of alternative strategies for the return of results from ES/WGS are under development and/or being tested empirically.[47,48] These strategies must meet the needs of individuals and families undergoing genetic testing and providers and institutions returning the genetic results. We have advocated for the use of individual- and family-focused web-based tools for self-guided management of genetic and genomic results such that ES/WGS results are framed as a dynamic resource of information that is managed by individuals in partnership with their healthcare providers.[47] These approaches present ES/WGS results as a resource rather than a test, recognize that the utility of results will vary over time at different stages of life, and describe results in such a way as to make them more accessible and informative without excessively reducing their medical impact.

We were able to overcome several limitations of previous studies: (1) small sample sizes, ranging from 30

to 1,000 individuals;[17,19,21,43] (2) analysis of primarily affected individuals;[19] (3) analysis of lower-coverage exomes;[21,43] (4) analysis of only known genotypes[20] or targeted sequencing;[19] and (5) analysis of only genes associated with autosomal-recessive conditions.[19,20] Nevertheless, there are several important limitations to this work. First, the ESP6500 data set is derived from exomes of individuals with heart, lung, and blood phenotypes. Therefore, our results might not be fully generalizable to other populations. In particular, the variants identified, allele frequencies, and burden calculations might be underestimates for younger, specifically newborn, populations, given that most early-onset disorders and conditions that cause early death (e.g., <18 years of age) were not included in the cohorts sequenced by the ESP. Second, because the ESP sampled mainly individuals who self-identified as EA or AA, burden estimates of variants of utility in other populations (e.g., Asian Americans, Hispanic Latinos, Ashkenazi Jews, etc.) most likely differ. Third, we were unable to assess CNVs and indels in the ESP6500 cohort. For this reason, our burden calculations are probably underestimates and are unlikely to fully capture the potential burden of pathogenic variants for these three gene categories.

In summary, we have demonstrated that every exome can be expected to contain variants of potential clinical utility. Considering only those affected by highly penetrant actionable disorders, as suggested by the ACMG, would result in the return of incidental results from ES/WGS to only 1%–3% of the population. Expanding the return to include conditions such as those studied here has the potential to increase the value of ES/WGS because of its possible large-scale consequences on the effective practice of medicine. To this end, we think that efforts should focus on providing individuals with the opportunity to make educated choices about whether they would like to receive incidental results from ES/WGS and, if so, what kinds. In turn, this will require the development and testing of tools for effectively exercising this autonomy, allowing individuals in partnership with their healthcare providers to utilize genomic information as a resource over their lifespans for disease prevention, treatment, reproductive planning, and other psychosocial benefits. Finally, a more inclusive focus on curating risk variants in populations other than EA individuals is imperative for providing these potential benefits to all populations, especially underserved minorities.

Supplemental Data

Supplemental Data include three figures, eight tables, and Supplemental Acknowledgments and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.07.006.

Acknowledgments

We acknowledge the support of the National Heart, Lung, and Blood Institute (NHLBI), the contributions of the many research institutions that participated in this study, the study investigators, field staff, and the study participants who created the Exome Sequencing Project (ESP) resource for biomedical research. Funding for the NHLBI Grand Opportunity (GO) ESP was provided by NHLBI grants RC2 HL-103010 (Heart GO), RC2 HL-102923 (Lung GO), and RC2 HL-102924 (Women's Health Initiative Sequencing Project). Exome sequencing was supported by NHLBI grants RC2 HL-102925 (Broad GO) and RC2 HL-102926 (Seattle GO). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Received: May 1, 2014
Accepted: July 14, 2014
Published: July 31, 2014

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://www.1000genomes.org/
CADD, http://cadd.gs.washington.edu/
GATK, http://www.broadinstitute.org/gatk/
GeneTests, http://www.genetests.org/
HGMD, http://www.hgmd.org/
KING, http://people.virginia.edu/~wc9c/KING/
National Society of Genetic Counselors (NSGC) 2014 Professional Status Survey Executive Summary, http://nsgc.org/p/cm/ld/fid=68
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
PharmGKB,https://www.pharmgkb.org/
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
SeattleSeq Annotation 138 (134 used in this study), http://snp.gs.washington.edu/SeattleSeqAnnotation138/
UMAKE, http://www.sph.umich.edu/csg/kang/umake/download/

References

1. Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. Nat. Rev. Genet. *12*, 745–755.
2. Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N. Engl. J. Med. *369*, 1502–1511.
3. Below, J.E., Earl, D.L., Shively, K.M., McMillin, M.J., Smith, J.D., Turner, E.H., Stephan, M.J., Al-Gazali, L.I., Hertecant, J.L., Chitayat, D., et al.; University of Washington Center for Mendelian Genomics (2013). Whole-genome analysis reveals that mutations in inositol polyphosphate phosphatase-like 1 cause opsismodysplasia. Am. J. Hum. Genet. *92*, 137–143.
4. Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. Nat. Genet. *42*, 30–35.
5. Burke, W., Matheny Antommaria, A.H., Bennett, R., Botkin, J., Clayton, E.W., Henderson, G.E., Holm, I.A., Jarvik, G.P., Khoury, M.J., Knoppers, B.M., et al. (2013). Recommendations for returning genomic incidental findings? We need to talk!. Genet. Med. *15*, 854–859.

6. Green, R.C., Lupski, J.R., and Biesecker, L.G. (2013). Reporting genomic sequencing results to ordering clinicians: incidental, but not exceptional. JAMA *310*, 365–366.
7. McGuire, A.L., Joffe, S., Koenig, B.A., Biesecker, B.B., McCullough, L.B., Blumenthal-Barby, J.S., Caulfield, T., Terry, S.F., and Green, R.C. (2013). Point-counterpoint. Ethics and genomic incidental findings. Science *340*, 1047–1048.
8. Ross, L.F., Rothstein, M.A., and Clayton, E.W. (2013). Mandatory extended searches in all genome sequencing: "incidental findings," patient autonomy, and shared decision making. JAMA *310*, 367–368.
9. Tabor, H.K., Berkman, B.E., Hull, S.C., and Bamshad, M.J. (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*, 2916–2924.
10. Dewey, F.E., Grove, M.E., Pan, C., Goldstein, B.A., Bernstein, J.A., Chaib, H., Merker, J.D., Goldfeder, R.L., Enns, G.M., David, S.P., et al. (2014). Clinical interpretation and implications of whole-genome sequencing. JAMA *311*, 1035–1045.
11. Fabsitz, R.R., McGuire, A., Sharp, R.R., Puggal, M., Beskow, L.M., Biesecker, L.G., Bookman, E., Burke, W., Burchard, E.G., Church, G., et al.; National Heart, Lung, and Blood Institute working group (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. Circ Cardiovasc Genet *3*, 574–580.
12. Parens, E., Appelbaum, P., and Chung, W. (2013). Incidental findings in the era of whole genome sequencing? Hastings Cent. Rep. *43*, 16–19.
13. Bollinger, J.M., Scott, J., Dvoskin, R., and Kaufman, D. (2012). Public preferences regarding the return of individual genetic research results: findings from a qualitative focus group study. Genet. Med. *14*, 451–457.
14. Sapp, J., Dong, D., Stark, C., Ivey, L., Hooker, G., Biesecker, L., and Biesecker, B. (2013). Parental attitudes, values, and beliefs toward the return of results from exome sequencing in children. Clin. Genet. Published online August 20, 2013. http://dx.doi.org/10.1111/cge.12254.
15. Levenseller, B.L., Soucier, D.J., Miller, V.A., Harris, D., Conway, L., and Bernhardt, B.A. (2014). Stakeholders' Opinions on the Implementation of Pediatric Whole Exome Sequencing: Implications for Informed Consent. J. Genet. Couns. *23*, 552–565.
16. Yu, J.H., Harrell, T.M., Jamal, S.M., Tabor, H.K., and Bamshad, M.J. (2014). Attitudes of genetics professionals toward the return of incidental results from exome and whole-genome sequencing. Am. J. Hum. Genet. *95*, 77–84.
17. Dorschner, M.O., Amendola, L.M., Turner, E.H., Robertson, P.D., Shirts, B.H., Gallego, C.J., Bennett, R.L., Jones, K.L., Tokita, M.J., Bennett, J.T., et al.; National Heart, Lung, and Blood Institute Grand Opportunity Exome Sequencing Project (2013). Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am. J. Hum. Genet. *93*, 631–640.
18. Johnston, J.J., Rubinstein, W.S., Facio, F.M., Ng, D., Singh, L.N., Teer, J.K., Mullikin, J.C., and Biesecker, L.G. (2012). Secondary variants in individuals undergoing exome sequencing: screening of 572 individuals identifies high-penetrance mutations in cancer-susceptibility genes. Am. J. Hum. Genet. *91*, 97–108.
19. Bell, C.J., Dinwiddie, D.L., Miller, N.A., Hateley, S.L., Ganusova, E.E., Mudge, J., Langley, R.J., Zhang, L., Lee, C.C., Schilkey, F.D., et al. (2011). Carrier testing for severe childhood recessive diseases by next-generation sequencing. Sci. Transl. Med. *3*, ra4.
20. Lazarin, G.A., Haque, I.S., Nazareth, S., Iori, K., Patterson, A.S., Jacobson, J.L., Marshall, J.R., Seltzer, W.K., Patrizio, P., Evans, E.A., and Srinivasan, B.S. (2013). An empirical estimate of carrier frequencies for 400+ causal Mendelian variants: results from an ethnically diverse clinical sample of 23,453 individuals. Genet. Med. *15*, 178–186.
21. Xue, Y., Chen, Y., Ayub, Q., Huang, N., Ball, E.V., Mort, M., Phillips, A.D., Shaw, K., Stenson, P.D., Cooper, D.N., and Tyler-Smith, C.; 1000 Genomes Project Consortium (2012). Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. Am. J. Hum. Genet. *91*, 1022–1032.
22. Gonzalez-Garay, M.L., McGuire, A.L., Pereira, S., and Caskey, C.T. (2013). Personalized genomic disease risk of volunteers. Proc. Natl. Acad. Sci. USA *110*, 16957–16962.
23. Refsgaard, L., Holst, A.G., Sadjadieh, G., Haunsø, S., Nielsen, J.B., and Olesen, M.S. (2012). High prevalence of genetic variants previously associated with LQT syndrome in new exome data. Eur. J. Hum. Genet. *20*, 905–908.
24. Green, R.C., Berg, J.S., Grody, W.W., Kalia, S.S., Korf, B.R., Martin, C.L., McGuire, A.L., Nussbaum, R.L., O'Daniel, J.M., Ormond, K.E., et al.; American College of Medical Genetics and Genomics (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet. Med. *15*, 565–574.
25. Biesecker, L.G. (2012). Opportunities and challenges for the integration of massively parallel genomic sequencing into clinical practice: lessons from the ClinSeq project. Genet. Med. *14*, 393–398.
26. Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science *337*, 64–69.
27. Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G.P., and Mariotti, S.P. (2004). Global data on visual impairment in the year 2002. Bull. World Health Organ. *82*, 844–851.
28. Porter, L.F., and Black, G.C. (2014). Personalized ophthalmology. Clin. Genet. *86*, 1–11.
29. Lee, J.W., Aminkeng, F., Bhavsar, A.P., Shaw, K., Carleton, B.C., Hayden, M.R., and Ross, C.J. (2014). The emerging era of pharmacogenomics: current successes, future potential, and challenges. Clin. Genet. *86*, 21–28.
30. Rieder, M.J., and Carleton, B. (2014). Pharmacogenomics and adverse drug reactions in children. Front. Genet. *5*, 78.
31. Filipski, K.K., Mechanic, L.E., Long, R., and Freedman, A.N. (2014). Pharmacogenomics in oncology care. Front. Genet. *5*, 73.
32. Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. *18*, 1851–1858.
33. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
34. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

35. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
36. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
37. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
38. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873.
39. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
40. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.
41. Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S., and Sidow, A.; NISC Comparative Sequencing Program (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. *15*, 901–913.
42. Boone, P.M., Campbell, I.M., Baggett, B.C., Soens, Z.T., Rao, M.M., Hixson, P.M., Patel, A., Bi, W., Cheung, S.W., Lalani, S.R., et al. (2013). Deletions of recessive disease genes: CNV contribution to carrier states and disease-causing alleles. Genome Res. *23*, 1383–1394.
43. MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al.; 1000 Genomes Project Consortium (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science *335*, 823–828.
44. Dickerson, J.A., Cole, B., Conta, J.H., Wellner, M., Wallace, S.E., Jack, R.M., Rutledge, J., and Astion, M.L. (2014). Improving the value of costly genetic reference laboratory testing with active utilization management. Arch. Pathol. Lab. Med. *138*, 110–113.
45. Casalino, L.P., Dunham, D., Chin, M.H., Bielang, R., Kistner, E.O., Karrison, T.G., Ong, M.K., Sarkar, U., McLaughlin, M.A., and Meltzer, D.O. (2009). Frequency of failure to inform patients of clinically significant outpatient test results. Arch. Intern. Med. *169*, 1123–1129.
46. Feero, W.G. (2014). Clinical application of whole-genome sequencing: proceed with care. JAMA *311*, 1017–1019.
47. Yu, J.H., Jamal, S.M., Tabor, H.K., and Bamshad, M.J. (2013). Self-guided management of exome and whole-genome sequencing results: changing the results return model. Genet. Med. *15*, 684–690.
48. Vassy, J.L., Lautenbach, D.M., McLaughlin, H.M., Kong, S.W., Christensen, K.D., Krier, J., Kohane, I.S., Feuerman, L.Z., Blumenthal-Barby, J., Roberts, J.S., et al.; MedSeq Project (2014). The MedSeq Project: a randomized trial of integrating whole genome sequencing into clinical medicine. Trials *15*, 85.

The American Journal of Human Genetics, Volume *95*

Supplemental Data

Pathogenic Variants for Mendelian and Complex Traits in Exomes of 6,517 European and African Americans: Implications for the Return of Incidental Results

Holly K. Tabor, Paul L. Auer, Seema M. Jamal, Jessica X. Chong, Joon-Ho Yu, Adam S. Gordon, Timothy A. Graubert, Christopher J. O'Donnell, Stephen S. Rich, Deborah A. Nickerson, NHLBI Exome Sequencing Project, and Michael J. Bamshad

Figure S1. Principal components analysis of NHLBI ESP6500 SNVs

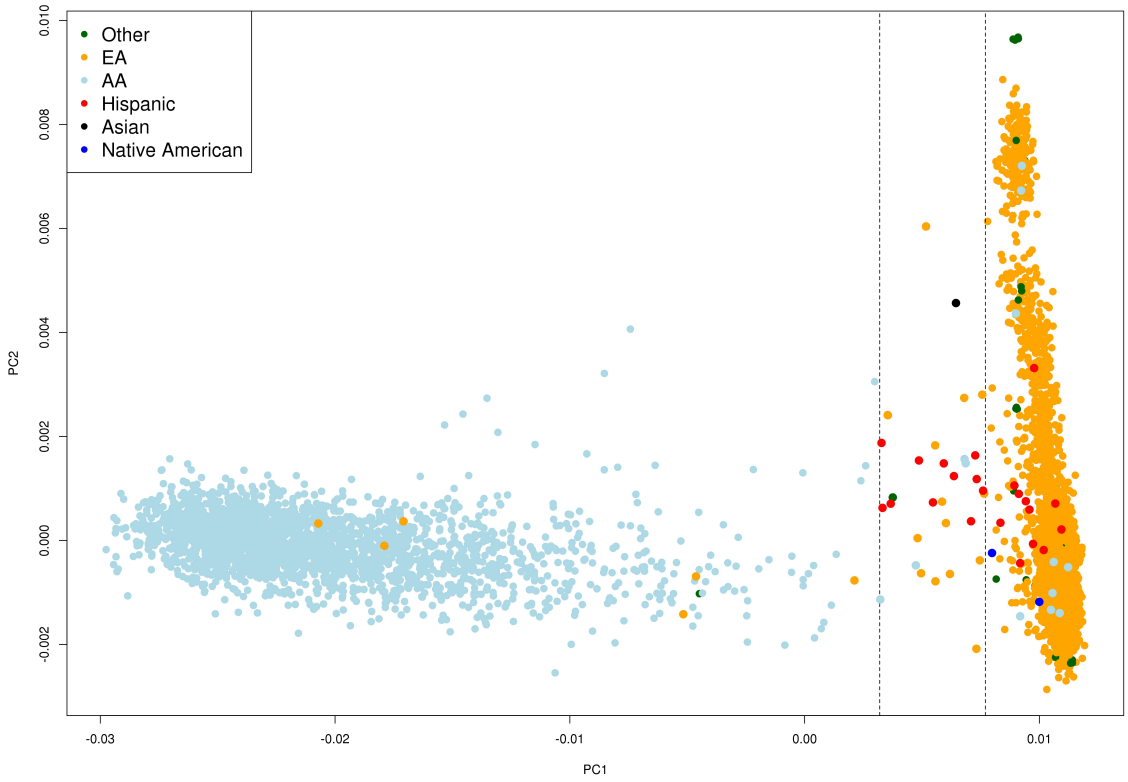

Figure S1. The first two principal components from the full 6,823 ESP samples. Self-reported European Americans (EA) are shown in orange, African Americans (AA) in light blue, Hispanics in red, Asians in black, and Native American in dark blue. Missing self-reported race is shown in green.

Figure S2. Filtering of Variants in Genes Assessed

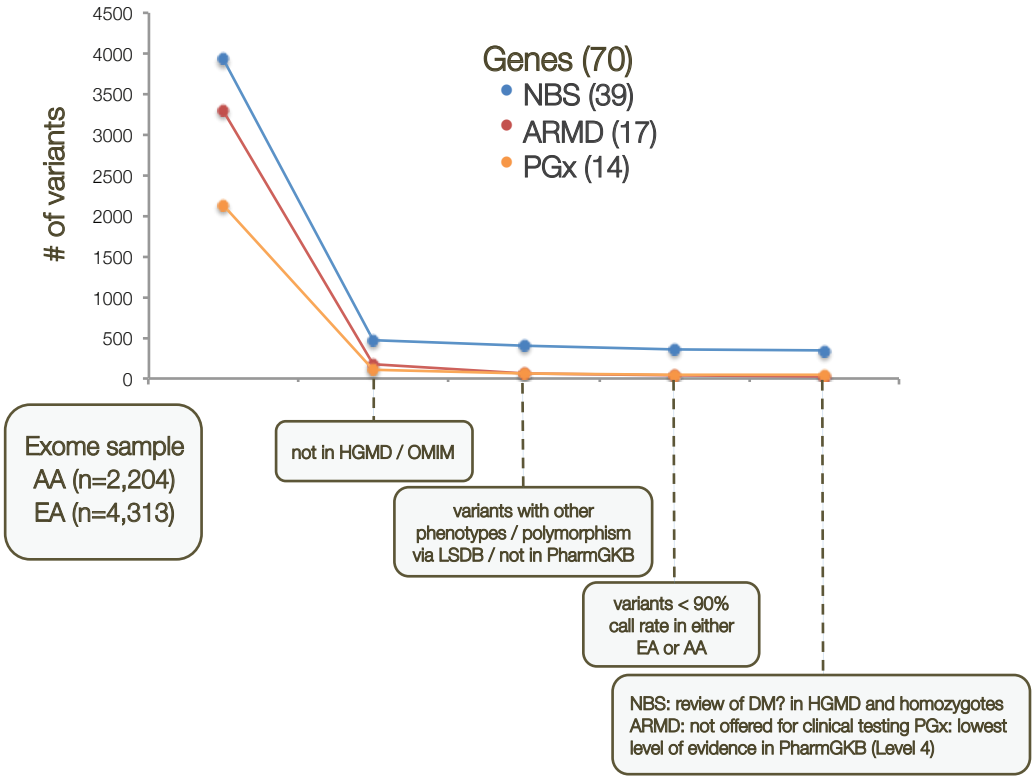

Figure S2. Plot showing the curation strategy for NBS (blue), ARMD (red), and PGx (orange) variants in the ESP6500. Exclusion step 1: variants in the PGx and ARMD gene sets that were not found in OMIM or HGMD; and variants not found in OMIM or HGMD as either "Disease Mutation" ("DM") or "Disease Mutation?" ("DM?") in the NBS gene set, were eliminated. Exclusion step 2: variants listed in HGMD or OMIM but associated with a different phenotype, those listed as polymorphic in one of twenty-seven locus-specific databases, and those not in PharmGKB were eliminated. Exclusion step 3: variants with a call rate <90% in either EA or AA were eliminated. Exclusion step 4: 10 variants in the NBS set that were homozygous in the absence, via literature review, of compelling evidence of disease causality and 3 variants that had an allele frequency >3%, suggesting they were polymorphisms rather than risk variants for a rare Mendelian disorder, and/or had been shown to not be pathogenic; in the ARMD gene set, those variants for which clinical testing was not available per GeneTests; and in the PGx gene set, had the lowest level of evidence for functional significance (i.e., level 4) in PharmGKB, were excluded.

Figure S3. Site Frequency Spectra of Variants by Gene Set

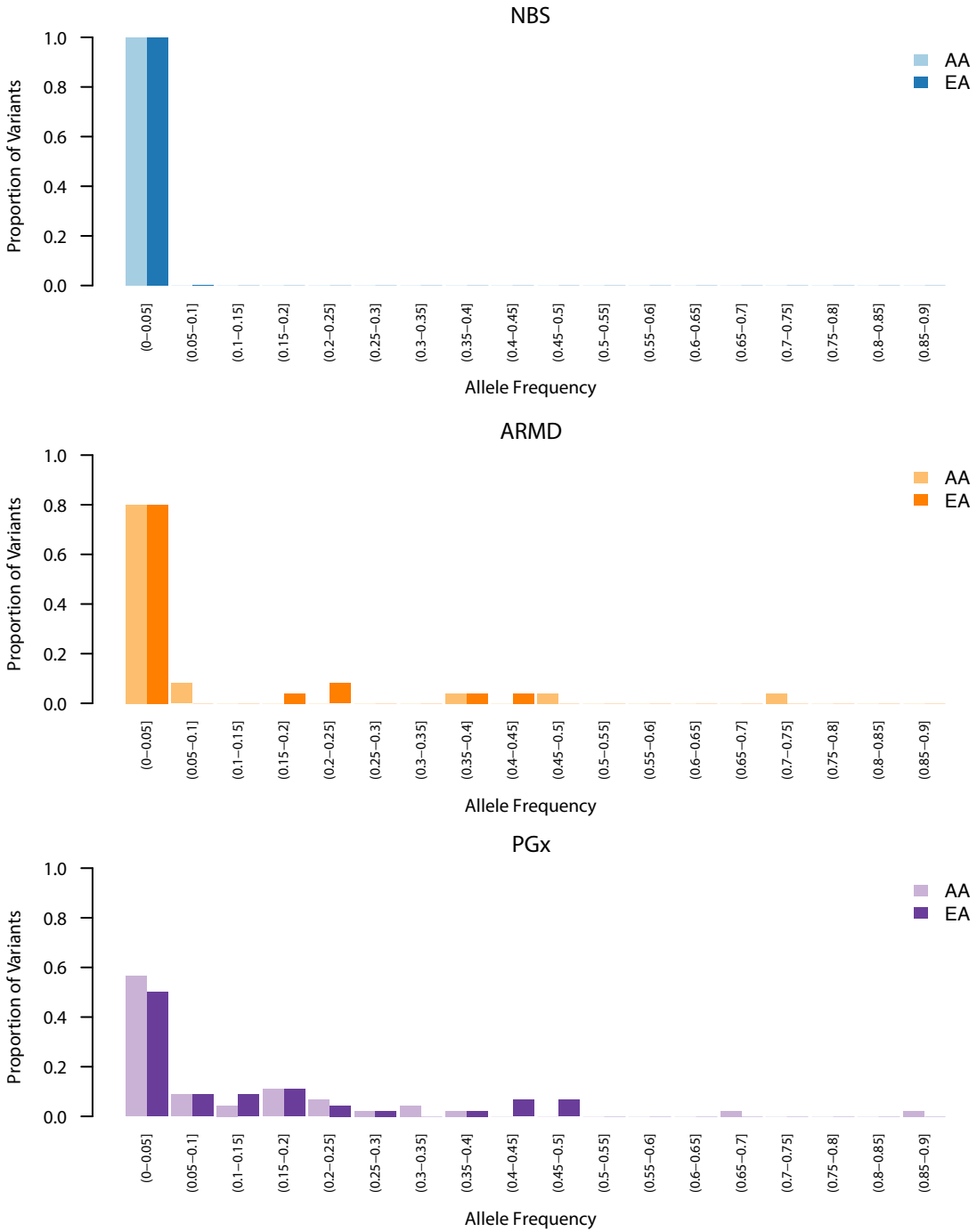

Figure S3. Histograms of the site frequency spectrum of risk variants underlying NBS phenotypes in European Americans (EA, blue) and African Americans (AA, light blue), ARMD phenotypes in EA (orange) and AA (light orange), and PGx phenotypes in EA (puprple) and AA (light purple).

Table S1: ESP Cohort descriptions

A. Overview

The ESP project funded by NHLBI included 3 cohort-focused awards. (1) Lung GO is a consortium of five lung disease related studies (Genomic Research on Asthma in the African Diaspora [GRAAD], Lung Health Study [LHS], Pulmonary Arterial Hypertension [PAH] population, Acute Lung Injury [ALI] cohort, and Cystic Fibrosis (CF) cohort). (2) HeartGO is a consortium of six cardiovascular disease-based prospective cohort studies comprised primarily of European American and African American participants: Atherosclerosis Risk in Communities [ARIC] study, Coronary Artery Risk Development in Young Adults [CARDIA], Cardiovascular Health Study [CHS], Framingham Heart Study [FHS], Jackson Heart Study [JHS], and Multi-Ethnic Study of Atherosclerosis [MESA]). (3) WHISP (Women's Health Initiative [WHI] Sequencing Project) is a prospective of women's health with a focus on coronary vascular disease (CVD). Additional studies were subsequently recruited to provide to provide additional cases of specific phenotypes [*e.g.*, early onset myocardial infarction (MI) and stroke]. Participants in all ESP studies provided written informed consent for non-commercial use of their data.

Further details regarding each of the 19 participating ESP studies can be found at their respective websites, listed below. dbGaP contains information about additional genetic information and phenotypes available on study participants (see http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap).

B. The LungGO Consortium

The goal of the NHLBI LungGO consortium is to identify disease-causing variants affecting a key set of pediatric and adult lung diseases by utilizing large cohorts characterized for a comprehensive set of clinical traits, including cystic fibrosis (CF), chronic obstructive pulmonary disease (COPD), pulmonary hypertension (PAH), asthma, and acute lung injury (ALI).

1. Cystic Fibrosis (CF):

Original cohort ascertainment: This cohort includes two longitudinal population cohorts of cystic fibrosis patients that have been well characterized for a comprehensive set of clinical traits for the study of *Pseudomonas aeruginosa* (*Pa*) acquisition. The first cohort, Early Pseudomonas Infection Control (EPIC or CFES-CF1) is the world's largest, multicenter, longitudinal, prospective cohort of early lung disease in young CF patients. CFES-CF1 consists of 1,704 CF cases who were ≤ 12 years old with no prior isolation of *Pa* or at least a two-year history of *Pa* negative cultures. The second cohort, the NHLBI-GWAS to Identify CF Modifiers (CFES-CF2), consists of 1,208 CF cases at the extremes of lung disease severity ("severe", worst 25^{th} %'tile of birth cohort vs. "mild", best 25^{th} %'tile) based on ~ 22 measures of lung function for each patient (over 5 yrs) developed at the University of North Carolina.

Ascertainment for ESP sequencing study: As part of the Lung-GO component of the NHLBI Exome Sequencing Project (ESP), exome sequencing was performed on 48 cystic fibrosis patients with early Pa infection and 48 cystic fibrosis patients with late Pa infection to identify variants influencing the time to onset of Pa infection. In Phase II, 330 additional exomes were

added to the study, to reach a total of 86 individuals with early Pa infection and 65 with late Pa infection. Additionally, 124 and 121 had mild and severe pulmonary function phenotype as determined by the survival corrected Kulich FEV percentile of Corey et al. The remaining 25 have intermediate phenotypes and/or show severe decline in lung function during childhood.

Prior GWAS: A GWAS for CF modifiers using 655,352 SNPs from the Illumina Infinium™ II HumanHap650Y BeadChip v.1.0 (Illumina Inc.) was performed on CF-2 and 3,300 additional CF cases.

Reference:
Treggiari MM, Rosenfeld M, Mayer-Hamblett N, Retsch-Bogart G, Gibson RL, Williams J, Emerson J, Kronmal RA, Ramsey BW; EPIC Study Group. Early anti-pseudomonal acquisition in young patients with cystic fibrosis: rationale and design of the EPIC clinical trial and observational study'. Contemp Clin Trials. 2009 May;30(3):256-68. Epub 2009 Jan 15. PubMed PMID: 19470318; PubMed Central PMCID: PMC2783320.

Website: none

2. COPDGene:
Original cohort ascertainment: The COPDGene Study is a multicenter study, with 21 U.S. clinical centers, designed to identify genetic variants affecting COPD susceptibility. Eligible subjects were aged 45-80 with at least 10 pack-years of cigarette smoking, without other concomitant respiratory diseases, and without contraindications to spirometry and chest CT scans. Of 10,171 eligible, enrolled subjects, 3534 are cases (GOLD stages 2-4 with forced expiratory volume over one second $FEV_1 < 80\%$ predicted and $FEV_1/FVC < 0.7$) and 4063 are controls with $FEV_1 \geq 80\%$ predicted and $FEV_1/FVC \geq 0.7$). Approximately 2/3 of COPDGene subjects are non-Hispanic European Americans , and 1/3 are African Americans. Study participants underwent a standardized study protocol including spirometry (pre/post bronchodilator), chest CT, questionnaires, six minute walk test, and phlebotomy.

Ascertainment for ESP sequencing study: As part of the Lung-GO component of the NHLBI Exome Sequencing Project (ESP) and with additional support from the COPD Foundation, 290 non-Hispanic European Americans (NHEA) COPDGene subjects underwent whole exome sequencing. COPD cases and controls were chosen from the phenotypic extremes in this cohort, focusing on COPD subjects with extensive emphysema: age < 63.0, GOLD Stage 3-4 COPD ($FEV_1 < 50\%$ predicted, $FEV_1/FVC < 0.7$), emphysema $\geq$ 15% at -950 HU, and no severe AAT deficiency. Control subjects were selected based on age $\geq$ 65.0, $FEV_1 \geq 85\%$ predicted, $FEV_1/FVC \geq 0.7$, and emphysema < 5% at -950 HU. Similar mean pack-years of smoking were included for the case and control groups.

Prior GWAS: Genome-wide SNP genotyping was obtained in all COPDGene subjects using the Illumina Omni-Express Chip which contains 730,525 SNPs..

Reference:
Regan, E. A., J. E. Hokanson, et al. 2010. Genetic epidemiology of COPD (COPDGene) study design. *COPD* **7**(1): 32-43).

Website: http://www.copdgene.org

3. Lung Health Study (LHS):
Original cohort ascertainment: The Lung Health Study (LHS) was a 14.5-year (1985-2001), multicenter (10 sites), randomized clinical trial to determine whether a program of smoking intervention and use of an inhaled bronchodilator could slow the rate of decline in pulmonary function or alter mortality among COPD patients. The LHS is a randomized multicenter clinical trial with 5,887 participants carried out from October 1986 to April 1994, designed to test the effectiveness of smoking cessation and bronchodilator administration in smokers aged 35 to 60 with mild lung function impairment. Participants were randomly assigned to one of three groups: (1) usual care, who received no intervention; (2) smoking intervention with the inhaled bronchodilator ipratroprium bromide; or (3) smoking intervention with an inhaled placebo. The effect of intervention was evaluated by the rate of decline of forced expiratory volume in one second (FEV1). The LHS represents one of the largest COPD cohorts worldwide (N=5,887). With additional support from the Canadian Institutes of Health Research, DNA is now available from over 4,600 of the LHS participants.

Ascertainment for ESP sequencing study: As part of the Lung-GO component of the NHLBI Exome Sequencing Project (ESP), exome sequencing was performed on 337 samples from European American participants who have lung function decline measures in at least 3 of 5 time points.

Prior GWAS: Genome-wide SNP genotyping was obtained in the European American LHS subjects using the HumanHap660W Quad Genotyping BeadChip.

Reference:
Connett JE, Kusek JW, Bailey WC, O'Hara P, Wu M. Design of the Lung Health Study: a randomized clinical trial of early intervention for chronic obstructive pulmonary disease. Control Clin Trials. 1993 Apr;14(2 Suppl):3S-19S. PubMed PMID: 8500311.

Website: http://www.biostat.umn.edu/lhs/

4. Pulmonary Arterial Hypertension (PAH):
Original cohort ascertainment: The Johns Hopkins SCCOR program entitled "Molecular Determinants of Pulmonary Arterial Hypertension" was funded by NIH in 2006 to utilize state-of-the-art physiological, molecular, genomic and proteomic approaches as well as novel phenotyping instrumentation that will provide the deepest understanding of the critical pathobiologic processes of pulmonary vascular (PV) and right ventricular remodeling, resulting RV-PV uncoupling, and their crucial impact on morbidity and mortality in PAH. This study is comprised of idiopathic PAH (IPAH) and PAH-scleroderma (SSc) cases and healthy controls. Rigorous definitions for primary phenotypes of interest have been used. Pulmonary hypertension is defined in IPAH or PAH-SSc patients as a mean pulmonary artery pressure greater than 25 mm Hg proven by right heart catheterization defined. For patients with scleroderma, the presence of disease is defined as systemic sclerosis with diffuse or limited scleroderma meeting the American College of Rheumatology criteria (LeRoy, 1988).

All patients underwent baseline routine clinical (e.g., 6 minute walk test) and echocardiographic evaluation, and hemodynamic assessment of pressures with vasodilator challenge. Repeat assessment was performed every 6 months with routine hemodynamic assessment performed at

one year, or at earlier time-points if clinically indicated. Data for the following are available for all participants: patient age, gender, race, severity and duration of illness (for scleroderma, IPAH and PAH-SSc), and other data related to clinical, hemodynamic (mean pulmonary artery pressure, pulmonary vascular resistance, cardiac index), and echocardiographic parameters (e.g., TAPSE measurement) related to PAH.

Ascertainment for ESP sequencing study: European American and African American participants with and without PAH. The study also focuses on patients with scleroderma, who are further stratified according to those who have or do not have PAH.

Prior GWAS: None

Reference: None

Website: none

5. Severe Asthma Research Program (SARP)

Original cohort ascertainment: SARP participants were recruited at the NHLBI SARP sites with an emphasis on recruiting severe asthmatics (Moore et al, Am J Respir Crit Care Med, 2010). Asthma status was based on both a physician's diagnosis and either bronchodilator reversibility or hyper-responsiveness to methacholine as well as less than 5 pack years of smoking. All subjects were carefully characterized using the standardized SARP protocol which included spirometry (medication withheld), maximum bronchodilator reversibility, hyper-responsiveness to methacholine (not performed in subjects with low baseline FEV1), skin-tests to common allergens, questionnaires on health care utilization and medication use and sputum, lung imaging and bronchoscopy in a subset.

Ascertainment for ESP sequencing study: The exome sequencing asthma project includes 191 African-Americans with asthma (82 severe, 109 non-severe).

Prior GWAS: Genotyping was performed on the Illumina 1Mv1 platform. GWAS results for asthma susceptibility were reported as part of the EVE consortium (Torgerson et al, Nat Genet 2011).

Reference:

Jarjour NN, Erzurum SC, Bleecker ER, Calhoun WJ, Castro M, Comhair SA, Chung KF, Curran-Everett D, Dweik RA, Fain SB, Fitzpatrick AM, Gaston BM, Israel E, Hastie A, Hoffman EA, Holguin F, Levy BD, Meyers DA, Moore WC, Peters SP, Sorkness RL, Teague WG, Wenzel SE, Busse WW; NHLBI Severe Asthma Research Program (SARP). Severe asthma: lessons learned from the National Heart, Lung, and Blood Institute Severe Asthma Research Program. Am J Respir Crit Care Med. 2012 Feb 15;185(4):356-62.

Website: www.severeasthma.org

6. Acute Lung Injury (ALI)

Original cohort ascertainment:

Acute Lung Injury (ALI) is a syndrome defined by the presence of acute hypoxemic respiratory failure (arterial oxygen: inspired oxygen ratio < 300), bilateral pulmonary infiltrates on chest

radiograph, a known clinical risk factor (e.g. sepsis, trauma, gastric fluid aspiration, massive transfusion), and the absence of physiologic or clinical evidence of congestive heart failure. ALI is a common occurrence in hospitalized patients in the United States with an estimated incidence of 78.9 cases/100,000 person-years and an associated mortality of 25-35 percent. This leads to the estimation that over 75,000 people die of ALI in the United States each year. This study was designed to use exome sequencing to identify coding variants associated with the extremes of ALI severity.

Original Cohort: We selected subjects from a cohort of approximately 800 patients with ALI enrolled from the Massachusetts General Hospital Intensive Care Unit (ICU) by Dr. David Christiani between 2000-2010 as part of the Molecular Epidemiology of acute respiratory distress syndrome (ARDS) study. These patients were followed through their intensive care unit (ICU) and hospital stay until death or discharge from the hospital. Dr. Christiani has used this cohort to identify several common genetic variants that alter susceptibility to ALI and related outcomes in the genes for *NFKBIA*, *IL10*, and *PBEF* [1-4].

Ascertainment for ESP sequencing study:

Phenotypes: The rate of resolution of ALI is highly variable. Examination of the composite variable of 'Ventilator-free days' (VFDs)[5] (all days between enrollment and day 28 during which the patient was both alive and breathing without mechanical ventilator support) in patients with ALI reveals a remarkably bi-modal distribution with approximately 17% of the patients dying or never liberating from mechanical ventilation over the first 28 days of observation while over 10% were free of mechanical ventilation for over 24 out of 28 days. We hypothesized that patients requiring prolonged mechanical ventilation (low VFDs) or who died with ALI relatively early after onset of the syndrome harbor functional genetic variants that predispose to more severe lung injury and impaired tissue injury repair compared with patients with rapidly liberating from dependence on mechanical ventilation (high VFDs). Notably, our subjects were all selected from this single-center cohort minimizing variation in processes of care such as the weaning from mechanical ventilation. We selected subjects with ALI at the two extremes of ventilator-free days, representing roughly the upper and lower 5th percentiles of this distribution, using the following criteria:

Exclusions: Individuals who are not European-Americans and early death unlikely to be attributable to ALI (Death within the 1st 48 hours after admission to intensive care unit), presence of DNAR (Do not attempt resuscitation/CPR) order indicative of an incomplete commitment to aggressive intensive care.

Inclusions: We restricted the cohort to those with an underlying diagnosis of sepsis (infection and a systemic inflammatory response) and who were found to have at least a moderate severity of illness (APACHE III acute physiology score ≥ 45) and severe hypoxemia ($PaO2/FiO2 < 200$).

Definitions of Extremes of VFD phenotype:

1. **Mild ALI :** VFDs ≥ 23 (n=46)
2. **Severe ALI :** VFDs = 0 (n=43)

Prior GWAS:
Subjects selected for this ESP study were also included in a GWAS designed to identify common genetic variants associated with risk for ALI (RC2 HL101779, 'Genetic determinants of ALI in

the iSPAAR consortium'). ALI cases (n=1200) and at-risk critically ill controls (n=1200) were genotyped using the Illumina 660W quad beadchip.

References:

1. Bajwa EK, Yu CL, Gong MN, Thompson BT, Christiani DC. Pre-B-cell colony-enhancing factor gene polymorphisms and risk of acute respiratory distress syndrome. *Crit Care Med.* May 2007;35(5):1290-1295.

Website: none

C. The HeartGO Consortium

HeartGO is a multiethnic consortium consisting of six NHLBI population-based cohorts of men and women with extensive baseline and follow-up data related to CVD outcomes and risk factors**.** The age range of participants in these six cohorts spans the spectrum from early adulthood to old age, providing a broad age representation. Each participating cohort in HeartGO has completed (a) genomewide SNP (GWAS) genotyping in most of its participants and (b) ascertainment of multiple high-resolution phenotypes, including all of the major CVD risk factors (blood pressure, lipids, diabetes status), biomarkers including measures of blood cell counts, subclinical disease imaging, and CVD and lung outcomes including myocardial infarction and stroke.

1. The Atherosclerosis Risk in Communities Study (ARIC):

Original Cohort Ascertainment: The ARIC study is a multi-center prospective investigation of atherosclerotic disease in a predominantly bi-racial population. Men and women aged 45-64 years at baseline were recruited from 4 communities: Forsyth County, North Carolina; Jackson, Mississippi; suburban areas of Minneapolis, Minnesota; and Washington County, Maryland. A total of 15,792 individuals participated in the baseline examination in 1987-1989, with follow-up examinations in approximate 3-year intervals, during 1990-1992, 1993-1995, and 1996-1998.

Ascertainment for ESP sequencing study: As part of the HeartGO component of the NHLBI Exome Sequencing Project (ESP), DNA from 1,235 participants were sent to both sequencing centers for exome sequencing. ARIC contributed samples and data to analysis of early-onset MI, blood pressure, LDL, ischemic stroke and the deeply phenotyped reference group. Of the 860 DNA samples that passed initial Q/C, finished sequence data was completed in 847 (98%) and deposited into the SRA; variant calls and phenotypic data were deposited into dbGaP.

Prior GWAS: ARIC Study samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0.

Reference:

The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* **129**, 687-702 (1989).

Website: http://www.cscc.unc.edu/aric/

2. The Cardiovascular Risk in Communities Study (CARDIA):

Original Cohort Ascertainment: The CARDIA study is a prospective, multi-center investigation of the natural history and etiology of cardiovascular disease in African Americans and European Americans 18-30 years of age at the time of initial examination [1] (http://www.cardia.dopm.uab.edu/index.htm). The CARDIA sample was recruited at random during

1985-86 primarily from geographically based populations in Birmingham AL, Chicago IL, and Minneapolis MN and, in Oakland, CA, from the membership of the Kaiser-Permanente Health Plan. The initial examination included 5,115 participants selectively recruited to represent proportionate racial, gender, age, and education groups from each of the four communities. Each participant's age, race, and sex were self-reported during the recruitment phase and verified during the baseline clinic visit. Details of the study design and procedures for data collection have been published. From the time of initiation of the study in 1985-1986 (baseline examination), six follow-up examinations have been conducted at years 2, 5, 7, 10, 15, 20, and 25.

Ascertainment for ESP sequencing study: As part of the HeartGO component of the NHLBI Exome Sequencing Project (ESP), DNA from 209 participants were sent to both sequencing centers for exome sequencing. CARDIA contributed samples and data to analysis of blood pressure, LDL and the deeply phenotyped reference group. Of the 209 DNA samples that passed initial Q/C, finished sequence data was completed in 207 (99%) and deposited into the SRA; variant calls and phenotypic data were deposited into dbGaP.

Prior GWAS: Prior genome-wide genotyping in CARDIA was performed separately for European Americans and African Americans. African American samples were genotyped at the Broad Institute using the Affymetrix Genome-Wide Human SNP Array 6.0, as part of the NHLBI Candidate Gene Association Resource (CARe) project. European American samples were genotyped using Affy6.0 at the Broad Institute of MIT and Harvard through the Gene Environment Association Studies initiative (GENEVA, http://www.genevastudy.org).

References:

1. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, et al (1988) CARDIA: study design, recruitment, and some characteristics of the examined subjects. J Clin Epidemiol 41:1105-1116.

Website: **http://www.cardia.dopm.uab.edu/**

3. The Cardiovascular Health Study (CHS):

Original Cohort Ascertainment: The CHS is a population-based cohort study of risk factors for CHD and stroke in adults ≥65 years conducted across four field centers in the United States. The original predominantly Caucasian cohort of 5201 persons was recruited in 1989-1990 from a random sample of people on Medicare eligibility lists and an additional 687 African-Americans were enrolled subsequently for a total sample of 5888.

Ascertainment for ESP sequencing study: As part of the HeartGO component of the NHLBI Exome Sequencing Project (ESP), DNA from 376 participants were sent to both sequencing centers for exome sequencing. CHS contributed samples and data to analysis of early-onset MI, blood pressure, LDL, ischemic stroke and the deeply phenotyped reference group. Of the 239 DNA samples that passed initial Q/C, finished sequence data was completed in 222 (93%) and deposited into the SRA; variant calls and phenotypic data were deposited into dbGaP.

Prior GWAS: In 2007-2008, genomewide genotyping was performed at the General Clinical Research Center's Phenotyping/Genotyping Laboratory at Cedars-Sinai using the Illumina 370CNV Duo® BeadChip system on the 3980 CHS participants who were free of CVD at baseline.

Reference:
Fried, L.P. *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-76 (1991).

Website: http://www.chs-nhlbi.org/default.htm

4. The Framingham Heart Study (FHS):
Original Cohort Ascertainment: The methods of recruitment and data collection have been described previously for the original Framingham Heart Study cohort (5,209 participants ascertained systematically from two-thirds of the households in the town of Framingham, MA, beginning in 1948), the Framingham Heart Study Offspring cohort (5,124 children of the original cohort, and spouses of those children, beginning in 1972) and the Third Generation cohort (4,095 children of the Offspring cohort, beginning in 2002).

Ascertainment for ESP sequencing study: As part of the HeartGO component of the NHLBI Exome Sequencing Project (ESP), DNA from 499 unrelated FHS Offspring participants were sent to both sequencing centers for exome sequencing. FHS contributed samples and data to analysis of early-onset MI, blood pressure, LDL, ischemic stroke and the deeply phenotyped reference group. Of the 493 DNA samples that passed initial Q/C, finished sequence data was completed in 475 (96%) and deposited into the SRA; variant calls and phenotypic data were deposited into dbGaP.

Prior GWAS: Prior genomewide genotyping was conducted for the SNP Health Association Resource (SHARe) project (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v10.p5) using the Affymetrix 500K mapping array (250K Nsp and 250K Sty arrays) and the Affymetrix 50K supplemental gene focused array.

References:
1. Kannel, W.B., Feinleib, M., McNamara, P.M., Garrison, R.J. & Castelli, W.P. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* **110**, 281-90 (1979).

Website: http://www.framinghamheartstudy.org/

5. The Jackson Heart Study (JHS):
Original Cohort Ascertainment: The Jackson Heart Study (JHS) is a prospective population-based study to seek the causes of the high prevalence of common complex diseases among African Americans in the Jackson, Mississippi metropolitan area. During the baseline examination period (2000-2004) 5,301 self-identified African Americans were recruited from four sources, including (1) randomly sampled households from a commercial listing; (2) ARIC participants; (3) a structured volunteer sample that was designed to mirror the eligible population; and (4) a nested family cohort. Unrelated participants were between 35 and 84 years old, and members of the family cohort were ≥ 21 years old when consent for genetic testing was obtained and blood was drawn for DNA extraction.

Ascertainment for ESP sequencing study: As part of the HeartGO component of the NHLBI Exome Sequencing Project (ESP), DNA from 535 participants were sent to both sequencing centers for exome sequencing. FHS contributed samples and data to analysis of early-onset MI, blood pressure, LDL, ischemic stroke, obesity/diabetes and the deeply phenotyped reference group. Of the 441 DNA

samples that passed initial Q/C, finished sequence data was completed in 424 (96%) and deposited into the SRA; variant calls and phenotypic data were deposited into dbGaP.

Prior GWAS: Prior genome wide association genotyping was performed at the Broad Institute of Harvard and MIT using the Affymetrix Genome-Wide Human SNP Array 6.0.

Reference:
Taylor HA, Jr., Wilson, JG, Jones DW, Sarpong, DF, Srinivasan A, et al (2005) Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. Ethn Dis 15: S6-4-17.

Website: http://www.jsums.edu/jhs/

6. The Multiethnic Study of Atherosclerosis (MESA):
Original Cohort Ascertainment: The Multi-Ethnic Study of Atherosclerosis (MESA) is a National Heart, Lung and Blood Institute-sponsored, population-based investigation of subclinical cardiovascular disease and its progression. A total of 6,814 individuals, aged 45 to 84 years, were recruited from six US communities (Baltimore City and County, MD; Chicago, IL; Forsyth County, NC; Los Angeles County, CA; New York, NY; and St. Paul, MN) between July 2000 and August 2002. Participants were excluded if they had physician-diagnosed cardiovascular disease prior to enrollment, including angina, myocardial infarction, heart failure, stroke or TIA, resuscitated cardiac arrest or a cardiovascular intervention (e.g., CABG, angioplasty, valve replacement, or pacemaker/defibrillator placement). Pre-specified recruitment plans identified four racial/ethnic groups (non-Hispanic European-American, African-American, Hispanic-American, and Chinese-American) for enrollment, with targeted oversampling of minority groups to enhance statistical power. Ethnicity was self-reported. The institutional review boards at each participating institution approved MESA and each individual participant provided informed written consent prior to enrollment.

Ascertainment for ESP sequencing study: As part of the HeartGO component of the NHLBI Exome Sequencing Project (ESP), DNA from 424 participants were sent to both sequencing centers for exome sequencing. MESA contributed samples and data to analysis of blood pressure, LDL, ischemic stroke, obesity/diabetes and the deeply phenotyped reference group. Of the 424 DNA samples that passed initial Q/C, finished sequence data was completed in 409 (96%) and deposited into the SRA; variant calls and phenotypic data were deposited into dbGaP.

Prior GWAS: Prior genomewide association genotyping was performed at the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) and at the Affymetrix Laboratory (Santa Clara, CA, USA) using the Affymetric Genome-Wide Human SNP Array 6.0.

Reference:
Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol 2002;156:871-81.

Website: http://www.mesa-nhlbi.org/

C. The Women's Health Initiative (WHI)

WHI is one of the largest (n=161,808) studies of women's health ever undertaken in the U.S. There are two major components of WHI: (1) a Clinical Trial (CT) that enrolled and randomized 68,132 women ages 50 – 79 into at least one of three placebo-control clinical trials (hormone therapy, dietary modification, and calcium/vitamin D); and (2) an Observational Study (OS) that enrolled 93,676 women of the same age range into a parallel prospective cohort study [1]. A diverse population including 26,045 (17%) women from minority groups were recruited from 1993-1998 at 40 clinical centers across the U.S. The design has been published [1]. For the CT and OS participants enrolled in WHI and who had consented to genetic research, DNA was extracted by the Specimen Processing Laboratory at the Fred Hutchinson Cancer Research Center (FHCRC) using specimens that were collected at the time of enrollment in to the study (between 1993 and 1998).

Prior GWAS: Genotyping was done at Affymetrix Laboratory on the Affymetrix 6.0 array.

References:

A series of papers describing methods for WHI, dealing with design [2], recruitment [3], postmenopausal hormone therapy trials [4], dietary modification trial [4], calcium and vitamin D supplement trial [6],observational study [7], and outcomes ascertainment [8] were published in the Annals of Epidemiology. The main trial results are published in [9-15].

1. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. Control Clin Trials. 1998 Feb;19(1):61-109.
2. Anderson GL, Manson J, Wallace R, Lund B, Hall D, Davis S, Shumaker S, Wang CY, Stein E, Prentice RL. Implementation of the Women's Health Initiative study design. Ann Epidemiol. 2003 Oct;13(9 Suppl):S5-17.
3. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, Rossouw JE. The Women's Health Initiative recruitment methods and results. Ann Epidemiol. 2003 Oct;13(9 Suppl):S18-77
4. Stefanick ML, Cochrane BB, Hsia J, Barad DH, Liu JH, Johnson SR. The Women's Health Initiative postmenopausal hormone trials: Overview and baseline characteristics of participants. Ann Epidemiol. 2003 Oct;13(9 Suppl):S78-86.
5. Ritenbaugh C, Patterson RE, Chlebowski RT, Caan B, Fels-Tinker L, Howard B, Ockene J. The Women's Health Initiative Dietary Modification Trial: Overview and baseline characteristics of participants. Ann Epidemiol. 2003 Oct;13(9 Suppl):S87-97.
6. Jackson RD, LaCroix AZ, Cauley JA, McGowan J. The Women's Health Initiative calcium-vitamin D trial: Overview and baseline characteristics of participants. Ann Epidemiol. 2003 Oct;13(9 Suppl):S98-106.
7. Langer RD, White E, Lewis CE, Kotchen JM, Hendrix SL, Trevisan M. The Women's Health Initiative Observational Study: Baseline characteristics of participants and reliability of baseline measures. Ann Epidemiol. 2003 Oct;13(9 Suppl):S107-21.
8. Curb JD, McTiernan A, Heckbert SR, Kooperberg C, Stanford J, Nevitt M, Johnson KC, Proulx-Burns L, Pastore L, Criqui M, Daugherty S, WHI Morbidity and Mortality Committee. Outcomes ascertainment and adjudication methods in the Women's Health Initiative. Ann Epidemiol. 2003 Oct;13(9 Suppl):S122-8.
9. The Writing Group for the WHI Investigators. Risks and benefits of estrogen plus progestin in healthy post-menopausal women: Principal results of the Women's Health Initiative randomized controlled trial. JAMA 2002;288(3):321-333.

10. The Women's Health Initiative Steering Committee. Effects of Conjugated Equine Estrogen in Postmenopausal Women With Hysterectomy. The Women's Health Initiative Randomized Controlled Trial. JAMA 2004; 291: 1701-1712.
11. Beresford S, Johnson K, Ritenbaugh C, Lasser N, Snetselaar L, Black H, Anderson G, Assaf A, Bassford T, Bowen D, Brunner R, Brzyski R, Caan B, Chlebowski R, et al. Low-Fat Dietary Pattern and Risk of Colorectal Cancer:The Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA 2006;295:643-654.
12. Howard B, Van Horn L, Hsia J, Manson J, Stefanick M, Wassertheil-Smoller S, Kuller L, LaCroix A, Langer R, Lasser N, Lewis C, Limacher M, Margolis K, Mysiw, et al.Low-Fat Dietary Pattern and Risk of Cardiovascular Disease:The Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA 2006;295:655-666.
13. Prentice R, Caan B, Chlebowski R, Patterson R, Kuller L, Ockene J, Margolis K, Limacher M, Manson J, Parker L, Paskett E, Phillips L, Robbins J, Rossouw J, et al.Low-Fat Dietary Pattern and Risk of Invasive Breast Cancer:The Women's Health Initiative Randomized Controlled Dietary Modification Trial. JAMA 2006;295:629-642.
14. Wactawski-Wende J, Kotchen J, Anderson G, Assaf A, Brunner R, O'Sullivan M, Margolis K, Ockene J, Phillips L, Pottern L, Prentice R, Robbins J, Rohan T, Sarto G, et al. Calcium plus Vitamin D Supplementaion and the Risk of Colorectal Cancer. NEJM 2006;354:(7):684-696.
15. Jackson R, LaCroix A, Gass M, Wallace R, Robbins J, Lewis C, Bassford T, Beresford S, Black H, Blanchette P, Bonds D, Brunner R, Bryzski R, Caan B, et al. Calcium plus Vitamin D Supplementaion and the Risk of Fractures. NEJM 2006;354:(7):669-683.

Website: http://www.whiscience.org.

D. Other EOMI Studies

1. Cleveland Clinic GeneBank (CCGB) was a single-center prospective cohort-based study that enrolled patients undergoing elective diagnostic coronary angiography between 2001 and 2006 [1]. Coronary artery disease (CAD) was defined as adjudicated diagnoses of stable or unstable angina, MI (adjudicated definition based on defined electrocardiographic changes or elevated cardiac enzymes), angiographic evidence of ≥ 50% stenosis of one or more major epicardial vessel, and/or a history of known CAD (documented MI, CAD, or history of revascularization). For the ESP EOMI study, 35 cases were ascertained from the CCGB study.

2. Heart Attack Risk in Puget Sound (HARPS) was a population-based case-control study that enrolled cases with incident MI presenting to a network of hospitals in Washington State between 1991 and 2002 [2]. In HARPS, eligible cases were men with MI at age less than 50 and women with MI at age less than 60. For the ESP EOMI study, 406 cases were selected from the HARPS study.

3. Massachusetts General Hospital - Premature Coronary Artery Disease (MGH-PCAD) was a hospital-based case-control study that enrolled cases hospitalized with MI at MGH between 1999 and 2004 [3]. In MGH-PCAD, eligible cases were men with MI at age less than 50 and women with MI at age less than 60. For the ESP EOMI study, 155 cases were drawn from the MGH-PCAD study.

4. Penn-CATH was a catheterization-lab based cohort study from the University of Pennsylvania Medical Center and enrolled subjects at the time of cardiac catheterization and coronary angiography between 1998 and 2003 [4]. Persons undergoing cardiac catheterization at either the Hospital of the University of Pennsylvania or Penn Presbyterian Medical Center consented for the PennCath study to

identify genetic and biochemical factors related coronary disease. For the ESP EOMI study, 36 cases were selected from the PennCATH study.

5. Translational Research Investigating Underlying Disparities in Acute Myocardial Infarction Patients' Health Status (TRIUMPH) was an observational, multi-center prospective registry that enrolled subjects presenting with MI at participating medical centers between 2005 and 2008 [5]. For the ESP EOMI study, 122 cases were selected from the TRIUMPH study.

Broad EOMI References:

1. Tang, W.H., et al., *Plasma myeloperoxidase predicts incident cardiovascular risks in stable patients undergoing medical management for coronary artery disease.* Clin Chem, 2011. **57**(1): p. 33-9.
2. Meiner, V., et al., *Cholesteryl ester transfer protein (CETP) genetic variation and early onset of non-fatal myocardial infarction.* Ann Hum Genet, 2008. **72**(Pt 6): p. 732-41.
3. Low, A.F., et al., *Aging syndrome genes and premature coronary artery disease.* BMC Med Genet, 2005. **6**: p. 38.
4. Helgadottir, A., et al., *A common variant on chromosome 9p21 affects the risk of myocardial infarction.* Science, 2007. **316**(5830): p. 1491-3.
5. Arnold, S.V., et al., *Translational Research Investigating Underlying Disparities in Acute Myocardial Infarction Patients' Health Status (TRIUMPH): design and rationale of a prospective multicenter registry.* Circ Cardiovasc Qual Outcomes, 2011. **4**(4): p. 467-76.

E. Other Stroke Studies

1. The Ischemic Stroke Genetics Study (ISGS)

The Ischemic Stroke Genetics Study (ISGS) was supported to perform a prospective genetic association study of ischemic stroke focusing on the hemostatic system. ISGS was a 5-center case-control study of first-ever ischemic stroke cases and concurrent controls individually matched for age, sex and recruitment site. This study utilized the NINDS Repository Cerebrovascular/Stroke Study, and neurologically normal controls from the sample population which are banked in the National Institute of Neurological Disorders and Stroke (NINDS Repository) collection for a first stage whole genome analysis. The number of study subjects that have individual level data available through Authorized Access is 485.

ISGS Study Case eligibility criteria:

Ischemic stroke was diagnosed according to World Health Organization definition by history and physical examination, as well as by and findings on brain imaging (either head computed tomography or magnetic resonance imaging). Subjects were eligible cases if they were over the age of 18 years and had a first-ever ischemic stroke with onset of symptoms within 30 days of enrollment. The study excluded cases with postoperative or post-procedural stroke (i.e., stroke related to cardiac catheterization, carotid stenting or conventional cerebral angiography); stroke related to recent subarachnoid hemorrhage or mechanical heart valve (aortic or mitral); patients with known inherited stroke syndromes (CADASIL, MELAS, Fabry disease, homocysteinemia, sickle cell anemia); and patients with biopsy-proven central nervous system vasculitis or stroke occurring in the setting of active bacterial endocarditis.

ISGS Study Control Eligibility criteria:

Controls were volunteers over the age of 18 years who have been verified stroke-free by structured interview. This study excluded volunteers who were considered unreliable historians, who had a blood relative enrolled as a case; and who were inpatients being treated for coronary or peripheral vascular disease. Controls needed to answer negatively on all the items of the Questionnaire for Verifying Stroke-Free Status (*Stroke* 2000 May; 31(5): 1076-80).

Prior GWAS:

References:

Meschia JF, Brott TG, Brown RD Jr, Crook RJ, Frankel M, Hardy J, Merino JG, Rich SS, Silliman S, Worrall BB; Ischemic Stroke Genetics Study. The Ischemic Stroke Genetics Study (ISGS) Protocol. BMC Neurol 2003 Jul 8;3:4. Epub 2003 Jul 8. PMID: 12848902

Matarín M, Brown WM, Scholz S, Simón-Sánchez J, Fung HC, Hernandez D, Gibbs JR, De Vrieze FW, Crews C, Britton A, Langefeld CD, Brott TG, Brown RD Jr, Worrall BB, Frankel M, Silliman S, Case LD, Singleton A, Hardy JA, Rich SS, Meschia JF. A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. Lancet Neurol 2007 May; 6(5):414-20. PMID: 17434096

C. Siblings With Ischemic Stroke (SWISS)

The Siblings with Ischemic Stroke Study (SWISS) was supported to perform an affected sibpair (ASP) linkage and family-based genetic association study of ischemic stroke focusing on the hemostatic system. The number of study subjects that have individual level data available through Authorized Access: 100 (100 phenotyped subjects)

Probands were recruited at 70 US and Canadian medical centers. Probands were adult (>18 years old) men and women presenting to a participating center with a study neurologist–confirmed ischemic stroke. Stroke was defined as rapidly developing signs of a focal or global disturbance of cerebral function, with symptoms lasting at least 24 hours or leading to death, with no apparent cause other than vascular origin (World Health Organization definition). Stroke was defined as ischemic when computed tomography or magnetic resonance imaging of the brain was performed within 7 days of onset of stroke symptoms and identified the symptomatic cerebral infarct or failed to identify an alternative cause of symptoms.

Probands were required to have reported at least 1 living full sibling with a history of stroke. No probands were enrolled with iatrogenic vasospastic or vasculitic stroke or if the stroke occurred in the setting of a mechanical heart valve or in the setting of untreated or actively treated bacterial endocarditis.

Probands were also excluded if they were known to have CADASIL, Fabry disease, homocysteinuria, MELAS, or sickle-cell anemia. Study neurologists at each center assigned to the qualifying ischemic stroke of each proband a Trial of Org 10172 in Acute Stroke Treatment (TOAST) subtype diagnosis.

Stroke-affected siblings of the proband (concordant siblings) were recruited by using proband-initiated contact. Telephone interviews were performed to obtain demographic and clinical information and to gain permission for obtaining medical records pertaining to treatment for stroke. Medical records were compiled and adjudicated by a central committee to verify the diagnosis of ischemic stroke and to

assign a TOAST subtype diagnosis. Unaffected siblings were ascertained by telephone contact and interview.

Prior GWAS: 223 probands, 248 stroke-affected siblings, and 84 stroke-unaffected siblings (total sample size, 555; DNA samples were genotyped using Genotyping was performed with an Illumina 610-quad array (probands) and an Illumina linkage V array (affected siblings)

References:

Meschia JF, Brown RD Jr, Brott TG, Chukwudelunzu FE, Hardy J, Rich SS; The Siblings With Ischemic Stroke Study (SWISS) Protocol. BMC Med Genet 2002; 3:1. Epub 2002 Feb 12. PMID: 11882254

Meschia JF, Nalls M, Matarin M, Brott TG, Brown RD Jr, Hardy J, Kissela B, Rich SS, Singleton A, Hernandez D, Ferrucci L, Pearce K, Keller M, Worrall BB; Siblings With Ischemic Stroke Study Investigators. Siblings with ischemic stroke study: results of a genome-wide scan for stroke loci. Stroke. 2011 Oct;42(10):2726-32. doi: 10.1161/STROKEAHA.111.620484. Epub 2011 Sep 22. PMID: 21940970

Section 2: ESP phenotype definitions and sample selection criteria

1. Early-Onset Myocardial Infarction (EOMI)

EOMI cases and controls were selected from ten studies, including ARIC, CCGB, FHS, HARPS, MGH-PCAD, PennCATH, TRIUMPH, WHI, CHS, and JHS. We ascertained 1,090 cases with MI at an early age. Early onset myocardial infarction (EOMI) cases were defined as individuals who had experienced an incident MI at age ≤ 60 years in women or ≤ 50 years in men. As a comparison group, we selected 979 participants from prospective cohort studies who were free of MI despite advanced age. Controls were selected as individuals with no history of MI at baseline or during follow-up to at least age 60 for men and 70 for women. Controls were also selected as having the highest baseline calculated Framingham risk scores (selected in descending order). Approximately two-thirds of the EOMI sample was of European ancestry and one-third of African-American ancestry. By design, the cases were, on average, more than two decades younger than the controls. Thus, male cases suffered an MI on average at 44 years old whereas the average age for male controls was 73 years old.

For the ESP EOMI study, 129 cases were selected from the primary ESP population-based cohorts ARIC, FHS and MESA studies and 642 controls were drawn from the ARIC, CHS, FHS, JHS and MESA studies. From WHI, 21 AA and 138 EA female MI cases age 60 or younger were selected. From the WHI study, 146 female cases and 305 female controls were selected. Among the prospective cohort studies, incident cases were defined by MI, coronary revascularization, hospitalized angina or CHD death, as adjudicated from medical record data by committee using standardized criteria. Additional cases meeting the EOMI criteria were selected from: HARPS (406 cases), a population-based case-control study that enrolled cases with incident MI in Washington State; MGH-PCAD (155 cases), a hospital-based MI case-control study; TRIUMPH (122 cases), an observational, multi-center prospective MI registry; PennCATH (36 cases), a catheterization-lab based coronary angiography cohort study from the University of Pennsylvania Medical Center; CCGB (35 cases), a single-center prospective cohort of patients undergoing diagnostic coronary angiography. Among the angiography-based studies,

coronary artery disease (CAD) was defined as adjudicated diagnoses of stable or unstable angina, MI (adjudicated definition based on defined electrocardiographic changes or elevated cardiac enzymes), angiographic evidence of ≥ 50% stenosis of one or more major epicardial vessel, and/or a history of known CAD (documented MI, CAD, or history of revascularization).

2. Ischemic Stroke

HeartGO stroke cases were defined as participants who had experienced an incident ischemic stroke that was subcategorized as either large vessel (atherosclerotic) or small vessel (lacunar) and falling into one of the following categories: stroke occurring by age 65 years and a positive family history of stroke, stroke occurring by age 65 years and no positive family history of stroke, or stroke occurring after age 65 years and a positive family history of stroke. Participants were excluded from selection if they had previously been selected for the EOMI or LDL studies. HeartGO stroke cases were selected from ARIC, CHS, FHS and MESA. A total of 55 samples passed initial quality control and 53 samples (45 EA cases and 8 AA cases) generated finished sequence. From WHI, women who had experienced a large or small vessel ischemic stroke were considered for selection. Women were selected based on the following priorities, age <65 and a positive family history of stroke, age < 65 and no positive family history of stroke, or age >=65 and a positive family history of stroke. The following table summarizes the women selected for sequencing in the Stroke study:

	African Americans		European Americans	
	Large Vess	Small Vessel	Large Vessel	Small Vessel
>=65 positive family history	5	10	49	143
<65 positive family history	1	11	17	47
<65 no positive family history	2	11	10	31

There were 98 affected sibpair members (SWISS) and 94 stroke cases from ISGS that passed QC and 94 SWISS and 89 ISGS samples had completed exome sequence data. For the ischemic stroke cases, control exome sequence data were chosen from the deeply-phenotyped reference group and other (non-stroke associated) phenotypes.

3. LDL-cholesterol

To enrich for individuals with rare large-effect size variants, we initially ascertained 412 individuals with extreme (high or low) LDL-C levels from ~25,000 population-based samples. The LDL-C extreme samples were selected initially from four population-based cohorts: ARIC, CHS, FHS and JHS. Samples previously selected as EOMI cases or controls were excluded from selection. In each cohort, first visit LDL-C was calculated using the Friedewald formula, based on HDL, triglyceride and total cholesterol measurements obtained in fasting subjects. For individuals on lipid lowering medication, pre-treatment LDL-C values were estimated by dividing treated LDL-C values by 0.75 to model a 25% reduction in LDL-C on therapy. Estimated pre-treatment LDL-C levels (or actual LDL-C levels for those not on lipid-lowering therapies) were then regressed on sex, age, and age-squared within cohort and within ethnicity strata (European-American and African-American). Residuals were then combined across studies, within ethnicity strata. The N=120 residuals in each ethnic stratum associated with the largest adjusted LDL-C values were selected. A corresponding number of individuals with the smallest adjusted LDL-C residuals were selected so that the number of extreme low LDL-C samples matched the number of extreme high LDL-C samples for each cohort. A second set of extreme LDL samples was selected from the CARDIA and MESA cohorts, using the same criteria that were used for the first set of LDL samples. The 412 selected samples roughly represent the 1st and 99th percentiles for adjusted LDL in European-Americans and the 2nd and 98th percentiles in African-Americans. A total of 147 EA LDL High, 142 AA LDL High, 145 EA LDL Low, and 131

AA LDL Low samples were sequenced and passed initial quality control. An additional N=26/ N=27 respectively) were selected for sequencing. African American women whose residuals were in the 2nd high/low EA samples and N=23/N=23 high/low AA samples were selected from WHI? Mean LDL-C in high subjects was xx mg/dl and in low subjects was xx mg/dl.

The extreme LDL-C samples were augmented with data from additional samples sequenced as part of ESP for other phenotypes from those with LDL-C measured and lipid-lowering medication status available (N=1593). The primary phenotypes for sample selection were EOMI and controls (individuals with no baseline or incident MI and with high estimated Framingham risk scores), ischemic stroke cases (large or small vessel ischemic stroke before age 65 or with positive family history), blood pressure extremes (1st and 99th sex- and decade-specific percentile tails), body mass index (high and low BMI), and a set of randomly selected samples among participants with near-complete phenotype data across a range of traits.

The total of 2005 ESP samples were selected from seven population based cohorts: Atherosclerosis Risk in Communities (ARIC), the Coronary Artery Risk Development in Young Adults (CARDIA), the Cardiovascular Health Study (CHS), the Framingham Heart Study (FHS), the Jackson Heart Study (JHS), the Multi-Ethnic Study of Atherosclerosis (MESA), and the Women's Health Initiative (WHI). Of the 2005 sequenced individuals, a total of 854 (43%) individuals were African-American and the remainder (N=1151, 57%) were European-American.

4. Blood Pressure

Samples were selected for the blood pressure study based on blood pressure measurements from all available visits. Participants were excluded from selection if they had previously been selected for another ESP phenotype, had a history of MI or heart failure at baseline, or age<20 or >70 years. Blood pressure measurements from visits with concurrent or a prior visit report of incident MI, CHF or BMI measurement > 4 standard deviations from the mean were excluded from selection eligibility. At each eligible visit, systolic blood pressure (SBP) and diastolic blood pressure (DBP) were adjusted for self-report of hypertension medication use. Both SBP and DBP were then regressed on age, age^2, BMI, sex, and race, within cohort. The sum of the standardized residuals (summing across SBP and DBP residuals) was used for selection. Individuals with sum of standardized residuals in the 1st and 99th race-, sex-, and age category-specific (based on decade) percentiles at any visit were selected for sequencing. HeartGO blood pressure samples were selected from the ARIC, CARDIA, CHS, FHS, JHS and MESA cohorts. A total of 184 EA High Blood Pressure, 82 AA High Blood Pressure, 175 EA Low Blood Pressure, and 74 AA Low Blood Pressure samples were sequenced and passed initial quality control. From WHI, individuals with sum of standardized residuals in the 1st (66 AAs, 99 EAs) and 99th percentiles (70 AAs, 102 EAs) were selected for sequencing.

5. Body Mass Index (BMI)

For the body mass index (BMI) study, 217 lean (BMI between 18 and 25) non-diabetic women were selected along with 163 morbidly obese (BMI over 40) diabetic women and 163 morbidly obese non-diabetic women were selected from WHI. An extra 70 morbidly obese women were added regardless of diabetes status. All women in the BMI study were of African American (AA) ethnicity. High body mass index (BMI) samples were additionally selected from the CARDIA, JHS and MESA cohorts. Samples were restricted to female AA participants who had not previously been selected for another ESP phenotype, and with first visit BMI measurement of at least 45, regardless of type 2 diabetes status. A total of 144 AA HeartGO BMI samples were sequenced and passed initial quality control.

6. Deeply Phenotyped Reference (DPR):

For the deeply phenotype reference (DPR) group, individuals were chosen based on availability of multiple heart, lung, and blood phenotypes. Eligibility for sample selection was determined using the following criteria: participants had to have non-missing data for the majority of a range of baseline cardiovascular-related phenotypes measured in each cohort (>95% of each sample) and had not been selected previously for another ESP phenotype. Samples were selected from the ARIC, CARDIA, CHS, FHS, JHS and MESA cohorts so that the race-specific (EA and AA) sample sizes for each were proportional to the overall cohort sizes. Samples were then randomly selected from the pool of eligible participants within each cohort. A total of 431 EA samples and 124 AA samples from HeartGO cohorts were sequenced and passed initial quality control. An additional 95 AA and 311 EA women from WHI were selected for sequencing in the DPR group.

Table S2. Genes and phenotypes assessed

Category	Gene	Phenotype
Newborn screening (NBS: 39)	*ACADM*	Medium chain acyl CoA dehydrogenase deficiency
	ACADVL	Very long chain acyl CoA dehydrogenase deficiency
	ACAT1	Mitochondrial acetoacetyl-CoA thiolase defiency (beta-kethothiolase deficiency)
	ASL	Argininosuccinic aciduria
	ASS1	Citrullinemia, type I
	BCKDHA	Maple syrup urine disease
	BCKDHB	Maple syrup urine disease
	BTD	Biotinidase deficiency
	CBS	Homocystinuria
	CFTR	Cystic fibrosis
	CYP21A2	Congenital adrenal hyperplasia
	DBT	Maple syrup urine disease
	FAH	Tyrosinemia type 1
	GALT	Classic galactosemia
	GCDH	Glutaric acidemia type 1
	GCH1	Hyperphenylalaninemia, BHF4-deficient, B
	GJB2	Hearing loss
	HADHA	Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency; Trifunctional protein deficiency
	HADHB	Trifunctional protein deficiency
	HBA1	Alpha thalassemia
	HBA2	Alpha thalassemia
	HBB	Sickle cell anemia; Sickle cell disease; Beta thalassemia
	HLCS	Multiple carboxylase deficiency (holocarboxylase synthetase deficiency)
	HMGCL	3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) lyase deficiency
	IVD	Isovaleric acdemia
	MCCC1	Biotin-unresponsive 3-methylcrotonyl-CoA carboxylase deficiency
	MCCC2	Biotin-unresponsive 3-methylcrotonyl-CoA carboxylase deficiency
	MMAA	Methylmalonic acidemia cblA type
	MMAB	Methylmalonic acidemia cblB type
	MMACHC	Methylmalonic acidemia and homocystinuria, cblC type
	MMADHC	Methylmalonic acidemia and homocystinuria, cblD type
	MUT	Methylmalonic acidemia
	PAH	Phenylketonuria; Hyperphenylalaninaemia
	PCBD1	Hyperphenylalaniemia, BHF4-deficient, D

	PCCA	Propionic acidemia
	PCCB	Propionic acidemia
	PTS	Hyperphenylalaninemia, BHF4-deficient, A
	QDPR	Hyperphenylalaninemia, BHF4-deficient, C
	SLC22A5	Carnitine uptake deficiency
Age related macular degeneration (ARMD: 17)	*ABCA4*	Age related macular degeneration
	APOE	Age related macular degeneration
	ARMS2	Age related macular degeneration
	C2	Age related macular degeneration
	C3	Age related macular degeneration
	CFB	Age related macular degeneration
	CFH	Age related macular degeneration
	CFHR1	Age related macular degeneration
	CFHR3	Age related macular degeneration
	CST3	Age related macular degeneration
	CX3CR1	Age related macular degeneration
	ERCC6	Age related macular degeneration
	FBLN5	Age related macular degeneration
	HMCN1	Age related macular degeneration
	HTRA1	Age related macular degeneration
	RAX2	Age related macular degeneration
	TLR4	Age related macular degeneration
Drug response (PGx: 14)	*ABCB1*	Digoxin sensitivity; nevirapine hepatotoxicity; Simvastatin response
	ABCC1	Methotrexate response
	ABCC2	Tenofovir response
	ABCG2	Rosuvastatin response
	ADRB1	Atenolol efficacy
	COMT	Nicotine replacement therapy response
	CYP2C19	Clopidogrel sensitivity
	CYP2C9	Warfarin sensitivity
	DPYD	Fluoropyrimidine response
	DRD2	Clozapine and Olanzapine response
	EPHX1	Carbamazepine dosage
	SLCO1B1	Simvastatin response
	TPMT	Thiopurine response
	UGT1A1	Irinotecan response

Table S3. Number of Variants Removed in Each Step of curation.

Number of variants by exclusion step and category (A); Number of variants in EA (B) and AA (C) by exclusion step and category; Total number of variants by exclusion step in EA, AA, and combined (D)

A.

Variants in ESP6500	NBS 3806		ARMD 3235		PGx 2078	
Exclusion Steps	Excluded	Remaining	Excluded	Remaining	Excluded	Remaining
1	3353	453	3049	186	1971	107
2	79	374	127	59	50	57
3	33	341	14	45	7	50
4	13	328	20	25	4	46

B.

Variants in ESP6500	NBS 2336		ARMD 2013		PGx 1375	
Exclusion Steps	Excluded	Remaining	Excluded	Remaining	Excluded	Remaining
1	1992	344	1865	148	1287	88
2	65	279	98	50	37	51
3	22	257	9	41	7	44
4	10	247	19	22	4	40

C.

Variants in ESP6500	NBS 2167		ARMD 1844		PGx 1144	
Exclusion Steps	Excluded	Remaining	Excluded	Remaining	Excluded	Remaining
1	1950	217	1729	115	1062	82
2	45	172	79	36	37	45
3	17	155	13	23	4	41
4	11	144	8	15	3	38

D.

Variants in ESP6500	EA (n=4313) 5724		AA (n=2203) 5155		EA+AA (n=6516) 9199	
Exclusion Steps	Excluded	Remaining	Excluded	Remaining	Excluded	Remaining
1	5144	580	4741	414	8373	746
2	200	380	161	253	256	490
3	38	342	34	219	54	436
4	33	309	22	197	37	399

Exclusion Step 1: Not in OMIM or HGMD

Exclusion Step 2: Variants with other phenotypes; polymorphism in LSDB; not in PharmGKB

Exclusion Step 3: Variants <90% call rate in either EA or AA

Exclusion Step 4: NBS: Review of DM? and homozygotes in HGMD; ARMD: Not offered for clinical testing; PGx: Lowest level of evidence in PharmGKB (level 4)

Table S4. Locus specific databases queried

Gene	Locus Specific Database URL
ACADM	https://research.cchmc.org/LOVD2/home.php?select_db=ACADM
ACADVL	https://research.cchmc.org/LOVD2/home.php?select_db=ACADVL
ACAT1	N/A
ASL	http://chromium.liacs.nl/LOVD2/home.php?select_db=ASL
ASS1	http://chromium.liacs.nl/LOVD2/home.php?select_db=ASS1
BCKDHA	http://databases.lovd.nl/shared/variants/BCKDHA
BCKDHB	N/A
BTD	http://www.arup.utah.edu/database/BTD/BTD_display.php
CBS	http://cbs.lf1.cuni.cz/mutations.php
CFTR	http://www.genet.sickkids.on.ca/ http://www.cftr2.org/
CYP21A2	http://www.cypalleles.ki.se/cyp21.htm
DBT	N/A
FAH	http://databases.lovd.nl/shared/variants/FAH
GALT	http://arup.utah.edu/database/GALT/GALT_welcome.php
GCDH	http://databases.lovd.nl/shared/variants/GCDH
GCH1	http://www.biopku.org/BIOMDB/BIOMDB_Results.asp
GJB2	http://davinci.crg.es/deafness/
HADHA	N/A
HADHB	N/A
HBA1	N/A
HBA2	http://globin.cse.psu.edu/globin/hbvar/menu.html
HBB	http://globin.cse.psu.edu/
HLCS	N/A
HMGCL	N/A
IVD	http://databases.lovd.nl/shared/variants/IVD
MCCC1	N/A
MCCC2	N/A
MMAA	http://www.genomed.org/LOVD/mma/home.php?select_db=MMAA
MMAB	http://www.genomed.org/LOVD/mma/home.php?select_db=MMAB
MMACHC	http://www.genomed.org/lovd/mma/home.php?select_db=MMACHC
MMADHC	N/A
MUT	http://www.genomed.org/lovd/mma/variants.php?action=view_unique&select_db=MUT
PAH	http://www.pahdb.mcgill.ca
PCBD1	N/A
PCCA	https://grenada.lumc.nl/LOVD2/mendelian_genes/home.php?select_db=PCCA
PCCB	https://grenada.lumc.nl/LOVD2/mendelian_genes/home.php?select_db=PCCB
PTS	http://www.biopku.org/BIOMDB/BIOMDB_Results.asp
QDPR	http://www.biopku.org/BIOMDB/BIOMDB_Results.asp
SLC22A5	http://arup.utah.edu/database/OCTN2/OCTN2_display.php

N/A: not available

Table S5. Allele frequency comparisons of included variants in newborn screening (NBS), age related macular degeneration (ARMD), and drug response (PGx) to 1000Genomes Project data and data from NHLBI Exome Variant Server

Category	Gene	transcript	chr.pos. (hg19)	rsID	ESP6500 AA Disease AF%	1000Genomes (AFR) AF%	ESP6500 EA Disease AF%	1000Genomes (EUR) AF%	dbSNP MAF%/MinorAlleleCount	EVS* Disease AF% (EA/AA/All)
Newborn screening (NBS)	*ACADM*	NM_000016.4	chr1:76198337	rs147559466	0.05%	A=0%	0.37%	A=1%	A=0.003/6	0.3488/0.0454/0.2461
	ACADM	NM_000016.4	chr1:76198409	rs121434280	0.05%	NA	0.11%	NA	NA	0.1047/0.0454/0.0846
	ACADM	NM_000016.4	chr1:76205779	rs121434278	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
	ACADM	NM_000016.4	chr1:76211507	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACADM	NM_000016.4	chr1:76211508	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACADM	NM_000016.4	chr1:76215194	rs121434274	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACADM	NM_000016.4	chr1:76226846	rs77931234	0.14%	G=0%	0.74%	G=1%	G=0.002/5	0.7442/0.1362/0.5382
	ACADM	NM_000016.4	chr1:76226906	rs148207467	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ACADM	NM_000016.4	chr1:76215192	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACADVL	NM_000018.2	chr17:7124982	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ACADVL	NM_000018.2	chr17:7126063	rs149467828	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ACADVL	NM_000018.2	chr17:7124899	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACADVL	NM_000018.2	chr17:7125285	rs140629318	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ACADVL	NM_000018.2	chr17:7125522	rs113994168	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACADVL	NM_000018.2	chr17:7125591	rs113994167	0.02%	NA	0.13%	NA	NA	0.1279/0.0227/0.0923
	ACADVL	NM_000018.2	chr17:7125608	0	0.02%	NA	0.03%	NA	NA	0.0349/0.0227/0.0308
	ACADVL	NM_000018.2	chr17:7126179	rs146589640	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACADVL	NM_000018.2	chr17:7127312	rs138058572	0.05%	NA	0.00%	NA	NA	0.0/0.0454/0.0154
	ACADVL	NM_000018.2	chr17:7127359	rs113994170	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACADVL	NM_000018.2	chr17:7127698	rs146379816	0.02%	T=0%	0.07%	T=0%	T=0.000/1	0.0698/0.0227/0.0538
	ACADVL	NM_000018.2	chr17:7128292	rs148584617	0.09%	A=0%	0.34%	A=0%	A=0.001/2	0.3372/0.0908/0.2537
	ACAT1	NM_000019.3	chr11:108009661	rs148639841	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACAT1	NM_000019.3	chr11:108010835	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ACAT1	NM_000019.3	chr11:108016927	rs145229472	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ASL	NM_000048.3	chr7:65552367	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ASL	NM_000048.3	chr7:65546812	rs145138923	0.05%	NA	0.30%	NA	NA	0.314/0.0454/0.223
	ASL	NM_000048.3	chr7:65547430	rs28940585	0.00%	T=60%	0.01%	T=21%	NA	0.0116/0.0/0.0077
	ASL	NM_000048.3	chr7:65547906	rs138310841	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ASL	NM_000048.3	chr7:65548107	rs143793815	0.00%	T=0%	0.08%	T=0%	T=0.001/2	0.0814/0.0/0.0538
	ASL	NM_000048.3	chr7:65548162	rs142637046	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ASL	NM_000048.3	chr7:65557065	rs28940287	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ASS1	NM_000050.4	chr9:133327612	rs138350285	0.14%	T=0%	0.02%	T=0%	T=0.001/2	0.0233/0.1362/0.0615
	ASS1	NM_000050.4	chr9:133333869	rs121908644	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ASS1	NM_000050.4	chr9:133333936	rs35269064	1.52%	T=1%	0.16%	T=0%	T=0.005/10	0.1628/1.5433/0.6305
	ASS1	NM_000050.4	chr9:133342161	rs121908637	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ASS1	NM_000050.4	chr9:133346260	rs121908646	0.00%	C=0%	0.02%	C=0%	C=0.000/1	0.0233/0.0/0.0154
	ASS1	NM_000050.4	chr9:133355791	rs148918985	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ASS1	NM_000050.4	chr9:133355803	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ASS1	NM_000050.4	chr9:133355833	rs121908645	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ASS1	NM_000050.4	chr9:133355834	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	ASS1	NM_000050.4	chr9:133364810	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	ASS1	NM_000050.4	chr9:133364800	rs183276875	0.02%	T=0%	0.00%	T=0%	T=0.000/1	0.0/0.0227/0.0077
	ASS1	NM_000050.4	chr9:133374932	rs121908641	0.02%	NA	0.02%	NA	NA	0.0233/0.0227/0.0231
	BCKDHA	NM_000709.3	chr19:41916560	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	BCKDHA	NM_000709.3	chr19:41916570	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	BCKDHA	NM_000709.3	chr19:41920030	rs34442879	0.16%	T=0%	1.06%	T=1%	T=0.006/12	1.0581/0.1589/0.7535
	BCKDHA	NM_000709.3	chr19:41928081	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	BCKDHA	NM_000709.3	chr19:41928183	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	BCKDHA	NM_000709.3	chr19:41925055	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
	BCKDHA	NM_000709.3	chr19:41928569	rs145901144	0.05%	NA	0.01%	NA	NA	0.0116/0.0454/0.0231
	BCKDHA	NM_000709.3	chr19:41928570	0	0.02%	NA	0.01%	NA	NA	0.0116/0.0227/0.0154
	BCKDHA	NM_000709.3	chr19:41930487	rs137852870	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
	BCKDHB	NM_000056.3	chr6:80838934	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	BCKDHB	NM_000056.3	chr6:80878661	rs149766077	0.00%	T=0%	0.02%	T=0%	T=0.000/1	0.0233/0.0/0.0154
	BCKDHB	NM_000056.3	chr6:80878662	rs79761867	0.00%	NA	0.05%	NA	NA	0.0465/0.0/0.0308
	BCKDHB	NM_000056.3	chr6:80910740	rs150084361	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	BTD	NM_000060.2	chr3:15686178	rs148031701	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
	BTD	NM_000060.2	chr3:15677019	rs34885143	0.18%	A=0%	1.61%	A=1%	A=0.004/8	1.593/0.1816/1.1149
	BTD	NM_000060.2	chr3:15683446	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077

BTD	NM_000060.2	chr3:15683548	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
BTD	NM_000060.2	chr3:15685833	rs146015592	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
BTD	NM_000060.2	chr3:15685994	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
BTD	NM_000060.2	chr3:15686568	0	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
BTD	NM_000060.2	chr3:15686298	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
BTD	NM_000060.2	chr3:15686693	rs13078881	0.77%	C=0%	4.15%	C=4%	C=0.019/41	4.1512/0.7717/3.0063
BTD	NM_000060.2	chr3:15686731	rs80338685	0.00%	C=0%	0.15%	C=0%	C=0.000/1	0.1512/0.0/0.1
BTD	NM_000060.2	chr3:15686732	rs146600671	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
BTD	NM_000060.2	chr3:15686852	rs138818907	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
CBS	NM_000071.2	chr21:44480585	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
CBS	NM_000071.2	chr21:44480591	rs117687681	0.12%	A=0%	0.39%	A=0%	A=0.001/2	0.3953/0.1135/0.2999
CBS	NM_000071.2	chr21:44483098	rs121964962	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
CBS	NM_000071.2	chr21:44484053	rs149119723	0.02%	NA	0.01%	NA	NA	0.0116/0.0227/0.0154
CBS	NM_000071.2	chr21:44484063	rs143124288	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
CBS	NM_000071.2	chr21:44483184	rs5742905	0.36%	NA	0.28%	NA	NA	0.2791/0.3631/0.3076
CBS	NM_000071.2	chr21:44486389	rs121964965	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CBS	NM_000071.2	chr21:44486463	rs121964964	0.02%	NA	0.05%	NA	NA	0.0465/0.0227/0.0384
CBS	NM_000071.2	chr21:44486353	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CBS	NM_000071.2	chr21:44492158	rs148865119	0.02%	NA	0.05%	NA	NA	0.0465/0.0227/0.0384
CFTR	NM_000492.3	chr7:117144378	rs143456784	0.00%	NA	0.03%	NA	NA	0.0233/0.0/0.0154
CFTR	NM_000492.3	chr7:117149085	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117149101	rs77284892	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117149144	rs142540482	0.05%	A=0%	0.12%	A=0%	A=0.000/1	0.1047/0.0454/0.0846
CFTR	NM_000492.3	chr7:117170947	0	0.02%	NA	0.05%	NA	NA	0.0465/0.0227/0.0384
CFTR	NM_000492.3	chr7:117171007	rs113993958	0.02%	NA	0.00%	NA	NA	0.0116/0.0227/0.0154
CFTR	NM_000492.3	chr7:117171028	rs77834169	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117171169	rs78756941	0.00%	NA	0.01%	NA	NA	0.0349/0.0/0.0231
CFTR	NM_000492.3	chr7:117174375	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117175323	rs138338446	0.00%	A=0%	0.01%	A=0%	A=0.000/1	0.0233/0.0/0.0154
CFTR	NM_000492.3	chr7:117175339	rs121908752	0.00%	NA	0.01%	NA	NA	0.0233/0.0227/0.0231
CFTR	NM_000492.3	chr7:117175372	rs121909046	0.00%	G=0%	0.05%	G=1%	G=0.006/13	0.0465/0.0/0.0308
CFTR	NM_000492.3	chr7:117175437	rs147432698	0.00%	not found	0.01%	not found	NA	NA
CFTR	NM_000492.3	chr7:117176711	rs151073129	0.39%	T=0%	0.00%	T=0%	T=0.000/1	0.0/0.3858/0.1307
CFTR	NM_000492.3	chr7:117180174	rs143486492	0.07%	NA	0.18%	NA	NA	0.1628/0.0681/0.1307
CFTR	NM_000492.3	chr7:117180324	rs77932196	0.00%	NA	0.01%	NA	NA	0.0233/0.0/0.0154
CFTR	NM_000492.3	chr7:117180327	rs142920240	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117180330	rs121909021	0.00%	NA	0.05%	NA	NA	0.0465/0.0/0.0308
CFTR	NM_000492.3	chr7:117227905	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117199476	rs34906874	0.12%	A=0%	0.01%	A=0%	A=0.004/8	0.0/0.1135/0.0384
CFTR	NM_000492.3	chr7:117199525	rs139573311	0.02%	NA	0.00%	NA	NA	0.0233/0.0227/0.0231
CFTR	NM_000492.3	chr7:117199602	rs77101217	0.00%	NA	0.01%	NA	NA	0.0233/0.0/0.0154
CFTR	NM_000492.3	chr7:117199683	rs77646904	0.07%	A=0%	0.00%	A=0%	A=0.001/2	0.0/0.0681/0.0231
CFTR	NM_000492.3	chr7:117227792	rs76713772	0.00%	A=0%	0.05%	A=0%	A=0.000/1	0.1395/0.0/0.0923
CFTR	NM_000492.3	chr7:117227832	rs113993959	0.02%	NA	0.06%	NA	NA	0.186/0.0454/0.1384
CFTR	NM_000492.3	chr7:117227854	rs121908755	0.02%	NA	0.00%	NA	NA	0.0/0.0454/0.0154
CFTR	NM_000492.3	chr7:117227860	rs75527207	0.00%	NA	0.12%	NA	NA	0.2093/0.0/0.1384
CFTR	NM_000492.3	chr7:117227865	rs74597325	0.00%	T=0%	0.01%	T=0%	T=0.000/1	0.0581/0.0/0.0384
CFTR	NM_000492.3	chr7:117149123	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117232470	rs140455771	0.02%	T=0%	0.08%	T=0%	T=0.000/1	0.0698/0.0227/0.0538
CFTR	NM_000492.3	chr7:117232481	rs150157202	0.02%	A=0%	0.27%	A=0%	A=0.000/1	0.2442/0.0227/0.1692
CFTR	NM_000492.3	chr7:117232574	0	0.02%	NA	0.00%	NA	NA	0.0116/0.0227/0.0154
CFTR	NM_000492.3	chr7:117234999	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117235090	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117242854	0	0.00%	NA	0.14%	NA	NA	0.1395/0.0/0.0923
CFTR	NM_000492.3	chr7:117242922	rs80224560	0.00%	NA	0.04%	NA	NA	0.0465/0.0/0.0308
CFTR	NM_000492.3	chr7:117243651	0	0.00%	NA	0.03%	NA	NA	0.0233/0.0/0.0154
CFTR	NM_000492.3	chr7:117243667	rs149790377	0.02%	NA	0.00%	NA	NA	0.0/0.0454/0.0154
CFTR	NM_000492.3	chr7:117243686	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117171044	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117243783	rs142773283	0.00%	NA	0.05%	NA	NA	0.0465/0.0/0.0308
CFTR	NM_000492.3	chr7:117246808	rs75096551	0.05%	A=0%	0.00%	A=0%	A=0.000/1	0.0/0.0681/0.0231
CFTR	NM_000492.3	chr7:117250625	rs149279509	0.00%	G=0%	0.04%	G=0%	G=0.000/1	0.0349/0.0/0.0231
CFTR	NM_000492.3	chr7:117251649	rs150212784	0.00%	NA	0.17%	NA	NA	0.1512/0.0/0.1
CFTR	NM_000492.3	chr7:117251695	rs1800114	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117251700	0	0.00%	NA	0.08%	NA	NA	0.0814/0.0/0.0538
CFTR	NM_000492.3	chr7:117251769	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117254753	rs75541969	0.00%	NA	0.04%	NA	NA	0.0233/0.0227/0.0231

CFTR	NM_000492.3	chr7:117267610	rs150326506	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
CFTR	NM_000492.3	chr7:117251703	0	0.00%	NA	0.01%	NA	NA	0.0/0.0227/0.0077
CFTR	NM_000492.3	chr7:117267714	rs75647395	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
CFTR	NM_000492.3	chr7:117267869	rs145743767	0.02%	A=0%	0.00%	A=0%	A=0.001/2	0.0/0.0227/0.0077
CFTR	NM_000492.3	chr7:117282468	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117251704	rs78769542	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
CFTR	NM_000492.3	chr7:117282538	rs76649725	0.02%	T=0%	0.00%	T=0%	T=0.000/1	0.0/0.0227/0.0077
CFTR	NM_000492.3	chr7:117282620	rs77010898	0.00%	NA	0.06%	NA	NA	0.0698/0.0/0.0461
CFTR	NM_000492.3	chr7:117282649	rs146795445	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117292931	rs80034486	0.00%	G=0%	0.01%	G=0%	G=0.000/1	0.0349/0.0/0.0231
CFTR	NM_000492.3	chr7:117304781	rs145545286	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
CFTR	NM_000492.3	chr7:117304901	rs146947665	0.00%	NA	0.03%	NA	NA	0.0233/0.0/0.0154
CFTR	NM_000492.3	chr7:117306983	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117305628	rs138642693	0.46%	C=0%	0.00%	C=0%	C=0.000/1	0.0/0.4539/0.1538
CFTR	NM_000492.3	chr7:117305619	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CFTR	NM_000492.3	chr7:117307052	rs148783445	0.25%	A=0%	0.03%	A=0%	A=0.001/2	0.0233/0.2497/0.1
CFTR	NM_000492.3	chr7:117307145	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
CYP21A2	NM_000500.7	chr6:32007593	rs6476	1.83%	A=2%	0.00%	A=0%	NA	0.0/1.8157/0.6151
CYP21A2	NM_000500.7	chr6:32008343	0	0.02%	NA	0.03%	NA	NA	0.0233/0.0227/0.0231
DBT	NM_001918.2	chr1:100681586	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
DBT	NM_001918.2	chr1:100680411	0	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
DBT	NM_001918.2	chr1:100681641	rs74103423	0.05%	A=0%	0.00%	A=0%	A=0.001/2	0.0/0.0454/0.0154
DBT	NM_001918.2	chr1:100680485	rs121964999	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
FAH	NM_000137.2	chr15:80460394	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
FAH	NM_000137.2	chr15:80460605	rs80338895	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
FAH	NM_000137.2	chr15:80465355	rs149052294	0.07%	NA	0.00%	NA	NA	0.0/0.0681/0.0231
FAH	NM_000137.2	chr15:80465431	rs80338898	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
FAH	NM_000137.2	chr15:80467400	0	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
FAH	NM_000137.2	chr15:80472514	rs80338900	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
FAH	NM_000137.2	chr15:80472526	rs11555096	0.35%	T=0%	2.28%	T=2%	T=0.011/23	2.2907/0.3404/1.63
FAH	NM_000137.2	chr15:80472572	rs80338901	0.00%	A=0%	0.07%	A=0%	A=0.000/1	0.0698/0.0/0.0461
FAH	NM_000137.2	chr15:80473411	rs121965076	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
GALT	NM_000155.2	chr9:34647200	rs111033656	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
GALT	NM_000155.2	chr9:34647855	rs111033690	0.25%	T=1%	0.00%	T=0%	T=0.002/4	0.0/0.2497/0.0846
GALT	NM_000155.2	chr9:34647864	rs111033686	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GALT	NM_000155.2	chr9:34647879	rs111033697	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GALT	NM_000155.2	chr9:34648167	rs75391579	0.09%	G=0%	0.27%	G=0%	G=0.001/3	0.2674/0.0908/0.2076
GALT	NM_000155.2	chr9:34648373	rs111033736	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GALT	NM_000155.2	chr9:34648433	rs111033750	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GALT	NM_000155.2	chr9:34648843	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GALT	NM_000155.2	chr9:34648885	rs111033766	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
GALT	NM_000155.2	chr9:34649029	rs111033773	0.00%	T=0%	0.06%	T=0%	T=0.000/1	0.0581/0.0/0.0384
GALT	NM_000155.2	chr9:34649442	rs2070074	2.93%	G=2%	9.36%	G=10%	G=0.055/119	9.3372/2.8824/7.1505
GALT	NM_000155.2	chr9:34649484	rs144993986	0.02%	NA	0.01%	NA	NA	0.0116/0.0227/0.0154
GALT	NM_000155.2	chr9:34649499	rs111033800	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GALT	NM_000155.2	chr9:34649532	rs111033814	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GALT	NM_000155.2	chr9:34650438	rs111033819	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
GCDH	NM_000159.2	chr19:13002779	rs142967670	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
GCDH	NM_000159.2	chr19:13004378	rs139851890	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
GCDH	NM_000159.2	chr19:13006872	rs149120354	0.02%	NA	0.05%	NA	NA	0.0465/0.0227/0.0384
GCDH	NM_000159.2	chr19:13007063	rs121434373	0.02%	NA	0.02%	NA	NA	0.0233/0.0227/0.0231
GCDH	NM_000159.2	chr19:13007748	rs121434371	0.05%	NA	0.00%	NA	NA	0.0/0.0454/0.0154
GCDH	NM_000159.2	chr19:13007781	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
GCDH	NM_000159.2	chr19:13008527	rs121434370	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GCDH	NM_000159.2	chr19:13008581	rs150938052	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
GCDH	NM_000159.2	chr19:13008638	rs121434369	0.02%	NA	0.06%	NA	NA	0.0581/0.0227/0.0461
GCDH	NM_000159.2	chr19:13008647	rs141437721	0.09%	G=0%	0.00%	G=0%	G=0.000/1	0.0/0.0908/0.0308
GCDH	NM_000159.2	chr19:13008674	rs147611168	0.05%	NA	0.01%	NA	NA	0.0116/0.0454/0.0231
GCDH	NM_000159.2	chr19:13010299	rs151201155	0.16%	A=0%	0.00%	A=0%	A=0.001/2	0.0/0.1589/0.0538
GCH1	NM_000161.2	chr14:55369176	rs56127440	0.00%	NA	0.07%	NA	NA	0.0698/0.0/0.0461
GJB2	NM_004004.5	chr13:20763104	rs111033294	0.00%	C=0%	0.01%	C=0%	C=0.000/1	0.0116/0.0/0.0077
GJB2	NM_004004.5	chr13:20763210	0	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
GJB2	NM_004004.5	chr13:20763222	rs111033360	0.11%	T=0%	0.00%	T=0%	T=0.000/1	0.0/0.1135/0.0384
GJB2	NM_004004.5	chr13:20763246	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GJB2	NM_004004.5	chr13:20763294	rs80338948	0.07%	A=0%	0.00%	A=0%	A=0.000/1	0.0/0.0681/0.0231
GJB2	NM_004004.5	chr13:20763305	rs76434661	0.00%	T=0%	0.08%	T=0%	T=0.001/2	0.0814/0.0/0.0538
GJB2	NM_004004.5	chr13:20763366	rs150529554	0.00%	T=0%	0.03%	T=0%	T=0.000/1	0.0349/0.0/0.0231

GJB2	NM_004004.5	chr13:20763423	rs143343083	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
GJB2	NM_004004.5	chr13:20763452	rs80338945	0.02%	NA	0.06%	NA	NA	0.0581/0.0227/0.0461
GJB2	NM_004004.5	chr13:20763471	rs104894409	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
GJB2	NM_004004.5	chr13:20763480	rs145216882	0.05%	NA	0.01%	NA	NA	0.0116/0.0454/0.0231
GJB2	NM_004004.5	chr13:20763483	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
GJB2	NM_004004.5	chr13:20763490	rs80338944	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GJB2	NM_004004.5	chr13:20763552	rs111033297	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GJB2	NM_004004.5	chr13:20763395	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GJB2	NM_004004.5	chr13:20763611	rs141774369	0.05%	NA	0.00%	NA	NA	0.0/0.0454/0.0154
GJB2	NM_004004.5	chr13:20763612	rs72474224	0.02%	T=0%	0.19%	T=0%	T=0.012/26	0.186/0.0227/0.1307
GJB2	NM_004004.5	chr13:20763627	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
GJB2	NM_004004.5	chr13:20763633	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GJB2	NM_004004.5	chr13:20763710	rs111033222	0.21%	T=0%	0.00%	T=0%	T=0.000/1	0.0/0.2043/0.0692
GJB2	NM_004004.5	chr13:20763534	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
GJB2	NM_004004.5	chr13:20763602	rs111033296	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
HADHA	NM_000182.4	chr2:26414191	rs142120825	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
HADHA	NM_000182.4	chr2:26418053	rs137852769	0.02%	G=0%	0.05%	G=0%	G=0.000/0	0.0465/0.0227/0.0384
HADHA	NM_000182.4	chr2:26435497	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
HADHA	NM_000182.4	chr2:26437990	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
HADHB	NM_000183.2	chr2:26486320	rs121913132	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
HADHB	NM_000183.2	chr2:26496605	rs146328300	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
HADHB	NM_000183.2	chr2:26508339	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
HBB	NM_000518.4	chr11:5246908	rs33946267	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
HBB	NM_000518.4	chr11:5246959	rs33913413	0.02%	NA	0.01%	NA	NA	0.0116/0.0227/0.0154
HBB	NM_000518.4	chr11:5247806	rs33945777	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
HBB	NM_000518.4	chr11:5247859	rs33993568	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
HBB	NM_000518.4	chr11:5247860	rs35553496	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
HBB	NM_000518.4	chr11:5248004	rs76728603	0.00%	A=0%	0.01%	A=0%	A=0.000/1	0.0116/0.0/0.0077
HBB	NM_000518.4	chr11:5248050	rs35004220	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
HBB	NM_000518.4	chr11:5248052	rs111851677	1.57%	G=2%	0.00%	G=0%	G=0.005/11	0.0/1.5902/0.5385
HBB	NM_000518.4	chr11:5248159	rs33971440	0.00%	T=0%	0.01%	T=0%	T=0.000/1	0.0116/0.0/0.0077
HBB	NM_000518.4	chr11:5248170	rs35424040	0.02%	NA	0.00%	NA	NA	NA
HBB	NM_000518.4	chr11:5248177	rs75680770	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
HBB	NM_000518.4	chr11:5248232	rs334	4.00%	A=9%	0.02%	A=0%	A=0.022/49	0.0233/4.0209/1.3771
HBB	NM_000518.4	chr11:5248233	rs33930165	1.61%	T=1%	0.01%	T=0%	T=0.002/5	0.0116/1.6129/0.5539
HBB	NM_000518.4	chr11:5248282	rs63750628	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
HLCS	NM_000411.6	chr21:38309329	rs144572349	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
HLCS	NM_000411.6	chr21:38128859	rs146448211	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
HLCS	NM_000411.6	chr21:38128865	rs140951243	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
HLCS	NM_000411.6	chr21:38128952	rs149399432	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
HLCS	NM_000411.6	chr21:38132079	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
IVD	NM_001159508.1	chr15:40708531	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
IVD	NM_001159508.1	chr15:40702898	rs142761835	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
IVD	NM_001159508.1	chr15:40707154	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
IVD	NM_001159508.1	chr15:40707653	rs28940889	0.07%	T=0%	0.07%	T=0%	T=0.001/2	0.0698/0.0681/0.0692
IVD	NM_001159508.1	chr15:40710364	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
MCCC1	NM_020166.3	chr3:182763210	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
MCCC1	NM_020166.3	chr3:182763310	rs119103212	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
MCCC2	NM_022132.4	chr5:70898412	rs141030969	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
MCCC2	NM_022132.4	chr5:70936845	rs150591260	0.02%	NA	0.09%	NA	NA	0.093/0.0227/0.0692
MCCC2	NM_022132.4	chr5:70942096	rs142887940	0.00%	NA	0.06%	NA	NA	0.0582/0.0/0.0384
MCCC2	NM_022132.4	chr5:70945029	rs139852818	0.05%	NA	0.12%	NA	C=0.002/5	0.1163/0.0454/0.0923
MCCC2	NM_022132.4	chr5:70948566	rs150327768	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
MCCC2	NM_022132.4	chr5:70895499	rs119103219	0.02%	NA	0.02%	NA	C=0.001/2	0.0233/0.0227/0.0231
MMAA	NM_172250.2	chr4:146560724	rs104893851	0.00%	NA	0.05%	NA	NA	0.0465/0.0/0.0308
MMAB	NM_052845.3	chr12:109994886	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
MMAB	NM_052845.3	chr12:109998858	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
MMADHC	NM_015702.2	chr2:150432296	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
MUT	NM_000255.3	chr6:49399544	rs121918252	0.07%	NA	0.00%	NA	NA	0.0/0.0681/0.0231
MUT	NM_000255.3	chr6:49403194	rs140600746	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
MUT	NM_000255.3	chr6:49403260	rs147094927	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
MUT	NM_000255.3	chr6:49407986	rs143023066	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
MUT	NM_000255.3	chr6:49408008	rs121918254	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
MUT	NM_000255.3	chr6:49407995	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
MUT	NM_000255.3	chr6:49425601	rs148331800	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
MUT	NM_000255.3	chr6:49419403	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
MUT	NM_000255.3	chr6:49426975	rs115923556	0.14%	NA	0.38%	NA	C=0.001/2	0.3837/0.1362/0.2999

PAH	NM_000277.1	chr12:103234177	rs5030861	0.02%	NA	0.10%	NA	NA	0.1047/0.0227/0.0769
PAH	NM_000277.1	chr12:103234252	rs5030860	0.00%	NA	0.05%	NA	NA	0.0465/0.0/0.0308
PAH	NM_000277.1	chr12:103234271	rs5030858	0.00%	NA	0.17%	NA	A=0.001/2	0.1744/0.0/0.1153
PAH	NM_000277.1	chr12:103234285	rs5030857	0.00%	NA	0.06%	NA	A=0.001/3	0.0581/0.0/0.0384
PAH	NM_000277.1	chr12:103237439	rs62508736	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103237454	rs5030856	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103237461	rs62516101	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
PAH	NM_000277.1	chr12:103237466	rs62516141	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103237484	rs62642937	0.00%	NA	0.07%	NA	NA	0.0698/0.0/0.0461
PAH	NM_000277.1	chr12:103237557	rs62507320	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
PAH	NM_000277.1	chr12:103237568	rs5030855	0.00%	T=0%	0.06%	T=0%	T=0.001/3	0.0581/0.0/0.0384
PAH	NM_000277.1	chr12:103238137	rs62516092	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103240716	rs62642935	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103245464	rs62514956	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
PAH	NM_000277.1	chr12:103245479	rs5030853	0.00%	NA	0.05%	NA	NA	0.0465/0.0/0.0308
PAH	NM_000277.1	chr12:103245481	rs62642933	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
PAH	NM_000277.1	chr12:103246588	rs62516146	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103246593	rs5030851	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103246597	rs62508698	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
PAH	NM_000277.1	chr12:103246612	rs62508691	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103246615	rs142934616	1.63%	C=3%	0.00%	C=0%	NA	0.0/1.6341/0.5536
PAH	NM_000277.1	chr12:103246624	rs62517164	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103246653	rs5030849	0.05%	NA	0.06%	NA	NA	0.0581/0.0454/0.0538
PAH	NM_000277.1	chr12:103246681	rs5030847	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
PAH	NM_000277.1	chr12:103246701	rs76212747	0.02%	NA	0.07%	NA	NA	0.0698/0.0227/0.0538
PAH	NM_000277.1	chr12:103246713	rs62508730	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103248926	rs62507348	0.05%	NA	0.00%	NA	NA	0.0/0.0454/0.0154
PAH	NM_000277.1	chr12:103248932	rs62516152	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
PAH	NM_000277.1	chr12:103246698	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103249091	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103249093	rs74486803	0.02%	NA	0.01%	NA	NA	0.0116/0.0227/0.0154
PAH	NM_000277.1	chr12:103249099	rs138809906	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
PAH	NM_000277.1	chr12:103260383	rs77554925	1.66%	C=2%	0.00%	C=0%	C=0.005/11	0.0/1.6341/0.5536
PAH	NM_000277.1	chr12:103260410	rs5030843	0.00%	NA	0.03%	NA	NA	0.0349/0.0/0.0231
PAH	NM_000277.1	chr12:103260411	rs75166491	0.02%	NA	0.01%	NA	NA	0.0116/0.0227/0.0154
PAH	NM_000277.1	chr12:103260446	rs62514909	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103271235	rs62507321	0.00%	NA	0.01%	NA	NA	0.0233/0.0/0.0154
PAH	NM_000277.1	chr12:103271239	rs62517166	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103271247	rs140175796	0.02%	NA	0.01%	NA	NA	0.0116/0.0227/0.0154
PAH	NM_000277.1	chr12:103288566	rs148393887	0.05%	NA	0.05%	NA	NA	0.0465/0.0454/0.0461
PAH	NM_000277.1	chr12:103288576	rs142516271	0.20%	NA	0.00%	NA	NA	0.0/0.2043/0.0692
PAH	NM_000277.1	chr12:103288590	rs62514903	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103288671	rs75193786	0.00%	G=0%	0.06%	G=0%	G=0.000/1	0.0581/0.0/0.0384
PAH	NM_000277.1	chr12:103306579	rs118092776	0.00%	T=0%	0.09%	T=0%	T=0.006/13	0.093/0.0/0.0615
PAH	NM_000277.1	chr12:103306594	rs5030841	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PAH	NM_000277.1	chr12:103306620	rs62642926	0.05%	NA	0.01%	NA	NA	0.0116/0.0454/0.0231
PAH	NM_000277.1	chr12:103310879	rs1801145	0.00%	C=0%	0.07%	C=0%	C=0.000/1	0.0698/0.0/0.0461
PCBD1	NM_000281.2	chr10:72643730	rs121913015	0.00%	NA	0.02%	NA	NA	0.0233/0.0/0.0154
PCBD1	NM_000281.2	chr10:72643759	rs115117837	0.16%	T=1%	0.01%	T=0%	T=0.001/3	0.0116/0.1589/0.0615
PCCA	NM_000282.3	chr13:101167714	rs145428347	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
PCCA	NM_000282.3	chr13:100764140	rs141371306	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
PCCA	NM_000282.3	chr13:101020733	rs61749895	0.23%	T=0%	0.81%	T=1%	T=0.003/7	0.8023/0.227/0.6074
PCCB	NM_000532.4	chr3:136016902	rs77820367	0.16%	A=0%	0.29%	A=0%	A=0.001/2	0.2907/0.1589/0.246
PCCB	NM_000532.4	chr3:136012626	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PCCB	NM_000532.4	chr3:136046480	rs121964961	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077
PTS	NM_000317.2	chr11:112103939	rs145882709	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PTS	NM_000317.2	chr11:112101362	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
PTS	NM_000317.2	chr11:112104210	rs150726932	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
SLC22A5	NM_003060.3	chr5:131705698	rs139203363	0.00%	A=0%	0.17%	A=0%	A=0.001/2	0.1628/0.0/0.1076
SLC22A5	NM_003060.3	chr5:131719973	rs121908888	0.00%	NA	0.01%	NA	NA	NA
SLC22A5	NM_003060.3	chr5:131721062	rs114269482	0.00%	T=0%	0.03%	T=0%	T=0.001/2	0.0233/0.0/0.0154
SLC22A5	NM_003060.3	chr5:131721136	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
SLC22A5	NM_003060.3	chr5:131722736	rs121908886	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
SLC22A5	NM_003060.3	chr5:131722731	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077
SLC22A5	NM_003060.3	chr5:131726522	rs144547521	0.02%	NA	0.01%	NA	NA	0.0116/0.0227/0.0154
SLC22A5	NM_003060.3	chr5:131728202	rs11568514	0.36%	G=0%	0.00%	G=0%	G=0.000/1	0.0/0.3631/0.123
SLC22A5	NM_003060.3	chr5:131729379	0	0.00%	NA	0.01%	NA	NA	0.0116/0.0/0.0077

	SLC22A5	NM_003060.3	chr5:131729380	rs28383481	0.18%	A=0%	0.52%	A=0%	A=0.004/8	0.5233/0.1816/0.4075
Age related macular degeneration (ARMD)	*ABCA4*	NM_000350.2	chr1:94463425	rs61748521	0.00%	0.00%	C=0%	0.01%	C=0%	C=0.000/1
	ABCA4	NM_000350.2	chr1:94463617	rs1800555	0.20%	0.20%	T=0%	1.24%	T=2%	T=0.007/15
	ABCA4	NM_000350.2	chr1:94473287	rs28938473	0.20%	0.20%	A=0%	0.52%	A=1%	A=0.003/6
	ABCA4	NM_000350.2	chr1:94473807	rs1800553	0.11%	0.11%	T=0%	0.42%	T=0%	T=0.002/5
	ABCA4	NM_000350.2	chr1:94476377	rs1800552	0.05%	0.05%	T=0%	0.26%	T=0%	T=0.001/2
	ABCA4	NM_000350.2	chr1:94487443	rs1800551	0.00%	0.00%	NA	0.01%	NA	NA
	ABCA4	NM_000350.2	chr1:94487490	rs1762111	0.02%	0.02%	G=0%	0.20%	G=0%	G=0.001/2
	ABCA4	NM_000350.2	chr1:94496053	rs1800549	0.02%	0.02%	A=0%	0.01%	A=0%	A=0.010/22
	ABCA4	NM_000350.2	chr1:94496666	rs61750130	0.00%	0.00%	NA	0.02%	NA	NA
	ABCA4	NM_000350.2	chr1:94508969	rs61751374	0.07%	0.07%	A=0%	0.22%	A=0%	A=0.001/2
	ABCA4	NM_000350.2	chr1:94512565	rs1801581	1.43%	1.43%	T=0%	4.09%	T=4%	T=0.015/32
	ABCA4	NM_000350.2	chr1:94512574	rs144995371	0.00%	0.00%	NA	0.06%	NA	NA
	ABCA4	NM_000350.2	chr1:94514466	rs139655975	0.02%	0.02%	C=0%	0.26%	C=0%	C=0.001/2
	ABCA4	NM_000350.2	chr1:94543389	rs1800548	0.00%	0.00%	NA	0.13%	NA	NA
	ABCA4	NM_000350.2	chr1:94564483	rs6657239	7.06%	7.06%	T=7%	3.46%	T=5%	T=0.051/112
	ABCA4	NM_000350.2	chr1:94461676	0	0.00%	0.00%	NA	0.01%	NA	NA
	ABCA4	NM_000350.2	chr1:94466426	rs140142529	0.00%	0.00%	NA	0.01%	NA	NA
	ABCA4	NM_000350.2	chr1:94508343	0	0.00%	0.00%	NA	0.01%	NA	NA
	ABCA4	NM_000350.2	chr1:94544906	rs150686179	0.00%	0.00%	NA	0.01%	NA	NA
	C3	NM_000064.2	chr19:6718387	rs2230199	4.81%	4.81%	C=4%	20.90%	C=20%	C=0.098/213
	CFH	NM_000186.3	chr1:196642233	rs800292	70.71%	70.71%	A=78%	22.21%	A=26%	A=0.435/947
	CFH	NM_000186.3	chr1:196659237	rs1061170	36.32%	36.32%	C=38%	38.21%	C=37%	C=0.278/605
	CFH	NM_000186.3	chr1:196682947	rs2274700	46.86%	46.86%	A=45%	40.53%	A=41%	A=0.437/951
	CFH	NM_000186.3	chr1:196709774	rs1065489	6.62%	6.62%	T=4%	17.47%	T=18%	T=0.233/508
	CFH	NM_000186.3	chr1:196716375	rs121913059	0.00%	0.00%	NA	0.02%	NA	NA
Drug response (PGx)	*ABCB1*	NM_000927.4	chr7:87138645	rs1045642	22.71%	77.29%	G=85%	47.75%	G=47%	A=0.397/864
	ABCB1	NM_000927.4	chr7:87160561	rs2032583	17.11%	17.11%	G=22%	13.25%	G=13%	G=0.132/287
	ABCB1	NM_000927.4	chr7:87160618	rs2032582	10.75%	10.75%	A=3%	43.13%	A=43%	A=0.340/741
	ABCB1	NM_000927.4	chr7:87179601	rs1128503	21.55%	21.55%	A=14%	42.94%	A=43%	A=0.422/919
	ABCB1	NM_000927.4	chr7:87199564	rs2235015	34.35%	34.35%	A=39%	19.69%	A=20%	A=0.202/440
	ABCC2	NM_000392.3	chr10:101563815	rs2273697	18.69%	18.69%	A=20%	19.43%	A=20%	A=0.174/378
	ABCC2	NM_000392.3	chr10:101595996	rs17222723	5.97%	5.97%	A=7%	6.17%	A=6%	A=0.043/93
	ABCC2	NM_000392.3	chr10:101604207	rs3740066	26.02%	26.02%	T=21%	37.16%	T=36%	T=0.304/663
	ABCC2	NM_000392.3	chr10:101611294	rs8187710	15.72%	15.72%	A=17%	6.19%	A=6%	A=0.070/152
	ABCC2	NM_000392.3	chr10:101594183	0	0.00%	0.00%	NA	0.01%	NA	NA
	ABCC2	NM_000392.3	chr10:101591385	rs17222547	0.02%	0.02%	NA	0.00%	NA	NA
	ABCC2	NM_000392.3	chr10:101610372	rs145715632	0.02%	0.02%	NA	0.00%	NA	NA
	ABCG2	NM_004827.2	chr4:89052323	rs2231142	3.22%	3.22%	T=2%	11.14%	T=10%	T=0.139/303
	COMT	NM_000754.3	chr22:19951271	rs4680	68.36%	68.36%	G=69%	47.79%	G=48%	A=0.390/850
	COMT	NM_000754.3	chr22:19956180	0	0.00%	0.00%	NA	0.01%	NA	NA
	CYP2C19	NM_000769.1	chr10:96540410	rs4986893	0.05%	0.05%	A=0%	0.02%	A=0%	A=0.014/31
	CYP2C19	NM_000769.1	chr10:96535278	0	0.00%	0.00%	NA	0.01%	NA	NA
	CYP2C9	NM_000771.3	chr10:96702047	rs1799853	2.68%	2.68%	T=2%	13.03%	T=12%	T=0.068/149
	CYP2C9	NM_000771.3	chr10:96702066	rs7900194	5.81%	5.81%	A=5%	0.05%	A=0%	A=0.012/27
	CYP2C9	NM_000771.3	chr10:96740981	rs28371685	1.91%	1.91%	A=3%	0.19%	A=0%	T=0.006/14
	CYP2C9	NM_000771.3	chr10:96741053	rs1057910	1.43%	1.43%	C=1%	6.61%	C=6%	C=0.042/92
	CYP2C9	NM_000771.3	chr10:96741058	rs28371686	1.07%	1.07%	G=2%	0.01%	G=0%	G=0.005/11
	DPYD	NM_000110.3	chr1:97915614	rs3918290	0.09%	0.09%	T=0%	0.58%	T=1%	T=0.003/6
	DPYD	NM_000110.3	chr1:97981395	rs1801159	15.51%	15.51%	C=20%	19.77%	C=17%	C=0.205/446
	DPYD	NM_000110.3	chr1:97981421	rs1801158	0.45%	0.45%	T=1%	2.02%	T=3%	T=0.014/31
	DPYD	NM_000110.3	chr1:98165091	rs2297595	4.10%	4.10%	C=5%	9.90%	C=12%	C=0.066/144
	DPYD	NM_000110.3	chr1:98348885	rs1801265	39.97%	39.97%	G=45%	22.50%	G=22%	G=0.230/502
	DPYD	NM_000110.3	chr1:97770839	0	0.02%	0.02%	NA	0.00%	NA	NA
	DPYD	NM_000110.3	chr1:98164926	rs146170505	0.00%	0.00%	NA	0.01%	NA	NA
	DPYD	NM_000110.3	chr1:98293695	rs141597515	0.00%	0.00%	NA	0.01%	NA	NA
	DRD2	NM_000795.3	chr11:113283459	rs6277	85.69%	85.69%	G=96%	45.40%	G=46%	A=0.273/595
	EPHX1	NM_000120.3	chr1:226019633	rs1051740	17.63%	17.63%	C=16%	29.96%	C=30%	C=0.316/689
	EPHX1	NM_000120.3	chr1:226019653	rs2292566	14.59%	14.59%	A=13%	14.43%	A=15%	A=0.179/389
	EPHX1	NM_000120.3	chr1:226026406	rs2234922	33.69%	33.69%	G=34%	20.41%	G=17%	G=0.185/402
	EPHX1	NM_000120.3	chr1:226026525	0	0.00%	0.00%	NA	0.01%	NA	NA
	EPHX1	NM_000120.3	chr1:226033029	0	0.02%	0.02%	NA	0.00%	NA	NA
	SLCO1B1	NM_006446.4	chr12:21329738	rs2306283	23.38%	76.62%	G=80%	40.27%	G=40%	A=0.405/882
	SLCO1B1	NM_006446.4	chr12:21329813	rs11045819	8.62%	8.62%	A=5%	16.16%	A=14%	A=0.079/171
	SLCO1B1	NM_006446.4	chr12:21331549	rs4149056	3.63%	3.63%	C=3%	15.48%	C=17%	C=0.123/268
	SLCO1B1	NM_006446.4	chr12:21294563	0	0.02%	0.02%	NA	0.00%	NA	NA
	TPMT	NM_000367.2	chr6:18130918	rs1142345	5.26%	5.26%	C=10%	4.24%	C=3%	C=0.046/100

	TPMT	NM_000367.2	chr6:18131012	rs1800584	0.00%	0.00%	NA	0.01%	NA	NA
	TPMT	NM_000367.2	chr6:18139228	rs1800460	1.00%	1.00%	T=1%	3.80%	T=3%	T=0.017/37
	TPMT	NM_000367.2	chr6:18143955	rs1800462	0.00%	0.00%	G=0%	0.23%	G=0%	G=0.004/8
	UGT1A10	NM_000463.2	chr2:234669144	rs4148323	0.14%	A=0%	0.13%	A=1%	A=0.052/114	0.1279/0.1362/0.1307
	UGT1A10	NM_000463.2	chr2:234675792	0	0.02%	NA	0.00%	NA	NA	0.0/0.0227/0.0077

NA: not available

*EVS version: v.0.0.26

Table S6. Known and inferred prevalence of phenotypes associated with pathogenic NBS variant set

Condition	Estimated carrier frequency*	Revised carrier frequency**	Published carrier frequency^
3-methylcrotonyl-CoA carboxylase deficiency	1/186	1/186	1/137
Argininosuccinic aciduria	1/163	1/163	1/132
B-ketothiolase deficiency	1/2172	1/2172	1/354
Beta thalassemia (AA)	1/31	1/735	1/121
Biotinidase deficiency	1/12	1/260**	1/124 profound biotinidase deficiency
Carnitine uptake deficiency	170	1/168**	1/132
Citrullinemia type 1	1/67	1/310**	1/224
Congenital adrenal hyperplasia	1/80	0	1/61
Cystic fibrosis (AA)	1/28	1/74** (non-delF508)	1/116 (non-delF508)
Cystic fibrosis (EA)	1/27	1/68** (non-delF508)	1/89 (non-delF508)
Galactosemia	1/7	1/116**	1/87 classic galactosemia
Glutaric acidemia type 1	1/181	1/181	1/177
Hemoglobin C disease (AA)	1/31	1/31	1/35
Holocarboxylase deficiency	1/1086	1/1086	1/250
Homocystinuria	1/69	1/119**	1/298
Isovaleric acidemia	1/501	1/501	1/144
Long chain 3-hydroxyacyl-CoA dehydrogenase deficiency	1/1086	1/1086	1/250
Maple syrup urine disease	1/53	1/233**	1/215
Medium chain acyl-coenzyme A dehydrogenase deficiency	1/54	1/60**	1/63
Methylmalonic acidemia	1/148	1/148	1/158
Phenylketonuria	1/23	1/61**	1/48
Propionic acidemia	1/56	1/181**	1/158
Sickle cell anemia (AA)	1/13	1/13	1/12
Trifunctional protein deficiency	1/543	1/724**	1/354
Tyrosinemia type 1	1/29	1/343**	1/150
Very long chain acyl-coenzyme A dehydrogenase deficiency	1/100	1/100	1/87

*Calculated by including all variants listed in Table S5 that are associated with the NBS condition
**Calculated after removal of variants (see Table S7)

^Published carrier frequencies were extracted from Feuchtbaum, L., Carter, J., Dowray, S., Currier, R.J., and Lorey, F. (2012). Birth prevalence of disorders detectable through newborn screening by race/ethnicity. Genetics in medicine : official journal of the American College of Medical Genetics 14, 937-945, and GeneReviews.

Table S7. Variants requiring additional manual review when calculating revised NBS carrier frequency estimates

Gene	Condition	Variants	ESP6500 AA MAF	ESP6500 EA MAF	HGMD DM?	Present as homozygote in ESP6500 EA or AA	>0.5% MAF in ESP6500 EA or AA	Included	Reason for inclusion*	Excluded	Reason for Exclusion**
ACADM	MCAD deficiency	c.199T>C	0.05%	0.11%	X					X	2
ACADM	MCAD deficiency	c.617G>A	0.00%	0.01%	X					X	2
ACADM	MCAD deficiency	c.985A>G	0.18%	0.52%			X	X	Known pathogenic variant***		
ASS1	Citrullinemia type 1	c.323G>T	1.52%	0.16%	X		X			X	5
BCKDHA	Maple syrup urine disease	c.452C>T	0.16%	1.06%		X	X			X	5
BTD	Biotinidase deficiency	c.133G>A	0.18%	1.61%		X	X			X	5
BTD	Biotinidase deficiency	c.1330G>C	0.77%	4.15%		X	X			X	2
CBS	Homocystinuria	c.833A>G	0.36%	0.28%		X				X	2
CFTR	Cystic fibrosis	c.3200C>T	0.00%	0.01%						X	1
CFTR	Cystic fibrosis	c.2620-26A>G	0.00%	0.14%						X	2
CFTR	Cystic fibrosis	c.3154T>G	0.00%	0.17%						X	1
CFTR	Cystic fibrosis	c.3205G>A	0.00%	0.08%						X	1
CFTR	Cystic fibrosis	c.3208C>T	0.00%	0.01%						X	4
CFTR	Cystic fibrosis	c.2260G>A	0.02%	0.27%						X	2
CFTR	Cystic fibrosis	c.715G>A	0.00%	0.01%						X	2, 6
CFTR	Cystic fibrosis	c.853A>T	0.39%	0.00%						X	2, 6
CFTR	Cystic fibrosis	c.1046C>T	0.00%	0.05%						X	1
CFTR	Cystic fibrosis	c.1558G>A	0.07%	0.00%						X	2, 6
CFTR	Cystic fibrosis	c.2506G>T	0.00%	0.01%						X	2, 6
CFTR	Cystic fibrosis	c.4003C>T	0.02%	0.00%						X	2
CFTR	Cystic fibrosis	c.4333G>A	0.25%	0.03%						X	6
CFTR	Cystic fibrosis	c.221G>A	0.05%	0.12%						X	4
CFTR	Cystic fibrosis	c.328G>T	0.02%	0.00%						X	4
CFTR	Cystic fibrosis	c.601G>A	0.00%	0.01%						X	4
CFTR	Cystic fibrosis	c.650A>G	0.00%	0.05%						X	4, 6
CFTR	Cystic fibrosis	c.2597G>A	0.00%	0.01%						X	4
CFTR	Cystic fibrosis	c.2723C>A	0.00%	0.03%						X	4
CFTR	Cystic fibrosis	c.3274T>C	0.00%	0.01%						X	2, 6
CFTR	Cystic fibrosis	c.3503A>G	0.02%	0.00%						X	2, 6
CFTR	Cystic fibrosis	c.4123C>A	0.00%	0.03%						X	4
CFTR	Cystic fibrosis	c.4426C>T	0.00%	0.01%						X	4
CFTR	Cystic fibrosis	c.274-6T>C	0.02%	0.05%						X	3
CFTR	Cystic fibrosis	c.890G>A	0.07%	0.18%						X	3
CFTR	Cystic fibrosis	c.1043T>A	0.00%	0.01%						X	2, 6
CFTR	Cystic fibrosis	c.3764C>A	0.02%	0.00%						X	5
CFTR	Cystic fibrosis	c.3849+45G>A	0.02%	0.00%						X	1
CFTR	Cystic fibrosis	c.3718-24G>A	0.00%	0.01%						X	1
CFTR	Cystic fibrosis	c.1679+18G>A	0.00%	0.01%	X					X	2, 6
CFTR	Cystic fibrosis	c.1393-42G>A	0.12%	0.01%	X					X	6
CFTR	Cystic fibrosis	c.3454G>C	0.00%	0.04%	X					X	4
CFTR	Cystic fibrosis	c.3209G>A	0.02%	0.00%				X	A		
CFTR	Cystic fibrosis	c.2855T>C	0.00%	0.05%				X	A		
CFTR	Cystic fibrosis	c.349C>T	0.00%	0.01%	X			X	B		
CFTR	Cystic fibrosis	c.4242+10T>C	0.46%	0.00%	X			X	A		
CFTR	Cystic fibrosis	c.125C>T	0.00%	0.03%				X	A		
CFTR	Cystic fibrosis	c.165-3C>T	0.00%	0.01%				X	A		
CFTR	Cystic fibrosis	c.178G>T	0.00%	0.01%				X	A, B		
CFTR	Cystic fibrosis	c.489+1G>T	0.00%	0.01%				X	B		
CFTR	Cystic fibrosis	c.535C>A	0.00%	0.01%				X	A		
CFTR	Cystic fibrosis	c.617T>G	0.00%	0.01%				X	A. B		
CFTR	Cystic fibrosis	c.1040G>C	0.00%	0.01%				X	B		

CFTR	Cystic fibrosis	c.1400T>C	0.02%	0.00%				X	B			
CFTR	Cystic fibrosis	c.1477C>T	0.00%	0.01%				X	B			
CFTR	Cystic fibrosis	c.1585-1G>A	0.00%	0.05%				X	A, B			
CFTR	Cystic fibrosis	c.1624G>T	0.02%	0.06%				X	B			
CFTR	Cystic fibrosis	c.1646G>A	0.02%	0.00%				X	B			
CFTR	Cystic fibrosis	c.1652G>A	0.00%	0.12%				X	B			
CFTR	Cystic fibrosis	c.1657C>T	0.00%	0.01%				X	B			
CFTR	Cystic fibrosis	c.200C>T	0.00%	0.01%				X	A, B			
CFTR	Cystic fibrosis	c.2249C>T	0.02%	0.08%				X	A			
CFTR	Cystic fibrosis	c.2353C>T	0.02%	0.00%				X	A			
CFTR	Cystic fibrosis	c.2657+5G>A	0.00%	0.04%				X	B			
CFTR	Cystic fibrosis	c.2739T>A	0.02%	0.00%				X	A			
CFTR	Cystic fibrosis	c.2758G>A	0.00%	0.01%				X	A			
CFTR	Cystic fibrosis	c.365A>G	0.00%	0.01%				X	A			
CFTR	Cystic fibrosis	c.2988+1G>A	0.05%	0.00%				X	A, B			
CFTR	Cystic fibrosis	c.3041A>G	0.00%	0.04%				X	A			
CFTR	Cystic fibrosis	c.3607A>G	0.02%	0.00%				X	A			
CFTR	Cystic fibrosis	c.3846G>A	0.00%	0.06%				X	A, B			
CFTR	Cystic fibrosis	c.3873+2T>C	0.00%	0.01%				X	A			
CFTR	Cystic fibrosis	c.3909C>G	0.00%	0.01%				X	A, B			
CFTR	Cystic fibrosis	c.4264C>T	0.00%	0.01%				X	A			
CFTR	Cystic fibrosis	c.4242+1G>T	0.00%	0.01%				X	A			
CYP21A2	Congenital adrenal hyperplasia	c.719T>A	1.83%	0.00%			X			X	2, 5	
CYP21A2	Congenital adrenal hyperplasia	c.1100G>A	0.02%	0.03%	X					X	2, 5	
FAH	Tyrosinemia type 1	c.1021C>T	0.35%	2.28%		X	X			X	2	
GALT	Galactosemia	c.940A>G	2.93%	9.36%		X	X			X	2	
HBB	Sickle cell disease	c.20T>A	4.00%	0.02%		X	X	X	Known pathogenic variant***			
HBB	Sickle cell disease	c.19C>T	1.61%	0.01%				X	Known pathogenic variant***			
B C	Beta thalassemia	c.93-23T>C	1.57%	0.00%		X	X			X	1	
PAH	Phenylketonuria	c.1241T>C	0.00%	0.05%						X	2	
PAH	Phenylketonuria	c.1208G>A	0.00%	0.06%						X	2	
PAH	Phenylketonuria	c.1169T>C	0.00%	0.01%						X	2	
PAH	Phenylketonuria	c.1139G>A	0.00%	0.07%						X	2	
PAH	Phenylketonuria	c.898C>A	0.00%	0.05%						X	2	
PAH	Phenylketonuria	c.823G>A	0.00%	0.01%						X	2	
PAH	Phenylketonuria	c.734A>G	0.02%	0.07%						X	2	
PAH	Phenylketonuria	c.688C>T	0.00%	0.03%						X	2	
PAH	Phenylketonuria	c.527C>A	0.02%	0.01%						X	2	
PAH	Phenylketonuria	c.442-5G>C	0.00%	0.01%						X	2	
PAH	Phenylketonuria	c.289T>G	0.20%	0.00%						X	2	
PAH	Phenylketonuria	c.275G>A	0.00%	0.01%						X	2	
PAH	Phenylketonuria	c.500T>C	1.66%	0.00%			X			X	2	
PAH	Phenylketonuria	c.820T>C	1.63%	0.00%			X			X	5	
HADHA	Trifunctional protein deficiency	c.919-2A>G	0.00%	0.01%	X					X	2	
HADHA	Trifunctional protein deficiency	c.731G>A	0.00%	0.01%	X					X	2	
HADHB	Trifunctional protein deficiency	c.341A>G	0.00%	0.01%	X					X	2	
PCCA	Propionic acidemia	c.1651G>T	0.23%	0.81%			X			X	5	
SLC22A5	Carnitine uptake deficiency	c.1463G>A	0.18%	0.52%			X			X	5	

Methods:

We performed an additional step of manual review of variants listed in HGMD as DM? (n=12) or present as homozygotes in either EA or AA, and/or variants with a MAF > 0.5% in either EA or AA (n=15) (Table S8), as well as all variants in CFTR. (n=60)). We conducted an in-depth literature review and review of detailed data in any available LSDBs. We excluded variants based on any of the following criteria 1) no clinical information available in a LSDB, 2) partial phenotype/non-classical phenotype, 3) no phenotype/polymorphism, 4) other phenotype, 5) found in *cis* with another known pathogenic variant, or 6) found heterozygous, second variant not identified. For *CFTR* [MIM 602421] we included variants listed as "CF" in the LSDB (CFTR2@Johns Hopkins), or cases with at least one case of CF listed in either LSDB (CFTR2@johnshopkins or the Cystic Fibrosis Mutation Database, n=33). Three non-CFTR variants, either present as homozygotes or with MAF > 0.5%, were pathogenic variants listed in GeneReviews and were also included. All 99 reviewed variants and

reason for inclusion or exclusion are listed in Table S8. For *PAH* [MIM 261600], we excluded all variants listed in OMIM or HGMD with the phenotype hyperphenylalanemia [MIM 261600] (n=13). While hyperphenylalaninemia is identified by NBS, it is not associated with the phenotype of phenylketonuria [MIM 261600]. Overall, 63 variants were excluded in this additional review. Following this review, carrier frequencies for 14 of 26 NBS conditions changed (Table S6, Figure 2). Revised carrier frequencies for 4 conditions were slightly higher than corresponding published estimates, suggesting that our criteria may still not be excluding non-pathogenic variants for those conditions or there may be variable penetrance among some variants.

*Reason for Exclusion: 1: No clinical information in locus specific database, 2: Partial phenotype/non-classic phenotype, 3: No phenotype/polymorphism, 4: Other phenotype, 5: Found in cis with another known pathogenic variant, 6: Heterozygous for variant, second variant not identified.
LSDB: Cystic Fibrosis Mutation Database
LSDB2: CFTR2@Johns Hopkins
**Reasons for Inclusion for CFTR variants: A: At least one case of CF with the variant in either LSDB or LSDB2, B: Listed as CF in LDSB2.
***Cited in GeneReviews as known pathogenic variants

Table S8. Identified risk variants in Newborn screening (NBS), Age related macular degeneration (ARMD), and Drug response (PGx) genes

Category	Gene	transcript	chr.pos. (hg19)	rsID	amino-acid	cDNA_pos	protein_pos	Ref.allele	Alt.allele	ESP6500 AA Disease AF%	ESP6500 EA Disease AF%
Newborn screening (NBS)	*ACADM*	NM_000016.4	chr1:76198337	rs147559466	GLU,LYS	127	43/422	G	A	0.05%	0.37%
	ACADM	NM_000016.4	chr1:76198409	rs121434280	TYR,HIS	199	67/422	T	C	0.05%	0.11%
	ACADM	NM_000016.4	chr1:76205779	rs121434278	GLY,ARG	583	195/422	G	A	0.00%	0.02%
	ACADM	NM_000016.4	chr1:76211507	0	ARG,CYS	616	206/422	C	T	0.00%	0.01%
	ACADM	NM_000016.4	chr1:76211508	0	ARG,HIS	617	206/422	G	A	0.00%	0.01%
	ACADM	NM_000016.4	chr1:76215194	rs121434274	GLY,ARG	799	267/422	G	A	0.00%	0.01%
	ACADM	NM_000016.4	chr1:76226846	rs77931234	LYS,GLU	985	329/422	A	G	0.14%	0.74%
	ACADM	NM_000016.4	chr1:76226906	rs148207467	ARG,stop	1045	349/422	C	T	0.02%	0.00%
	ACADM	NM_000016.4	chr1:76215192	0	ASP,GLY	797	266/422	A	G	0.00%	0.01%
	ACADVL	NM_000018.2	chr17:7124982	0	TYR,stop	603	201/656	C	G	0.02%	0.00%
	ACADVL	NM_000018.2	chr17:7126063	rs149467828	SER,stop	956	319/656	C	A	0.02%	0.00%
	ACADVL	NM_000018.2	chr17:7124899	0	VAL,MET	520	174/656	G	A	0.00%	0.01%
	ACADVL	NM_000018.2	chr17:7125285	rs140629318	ALA,PRO	637	213/656	G	C	0.02%	0.00%
	ACADVL	NM_000018.2	chr17:7125522	rs113994168	THR,MET	779	260/656	C	T	0.00%	0.01%
	ACADVL	NM_000018.2	chr17:7125591	rs113994167	VAL,ALA	848	283/656	T	C	0.02%	0.13%
	ACADVL	NM_000018.2	chr17:7125608	0	GLY,ARG	865	289/656	G	A	0.02%	0.03%
	ACADVL	NM_000018.2	chr17:7126179	rs146589640	LYS,GLU	1072	358/656	A	G	0.00%	0.01%
	ACADVL	NM_000018.2	chr17:7127312	rs138058572	ARG,GLN	1358	453/656	G	A	0.05%	0.00%
	ACADVL	NM_000018.2	chr17:7127359	rs113994170	ARG,TRP	1405	469/656	C	T	0.00%	0.01%
	ACADVL	NM_000018.2	chr17:7127698	rs146379816	ARG,TRP	1591	531/656	C	T	0.02%	0.07%
	ACADVL	NM_000018.2	chr17:7128292	rs148584617	ARG,GLN	1844	615/656	G	A	0.09%	0.34%
	ACAT1	NM_000019.3	chr11:108009661	rs148639841	ASN,ASP	472	158/428	A	G	0.00%	0.01%
	ACAT1	NM_000019.3	chr11:108010835	0	ARG,GLN	623	208/428	G	A	0.00%	0.01%
	ACAT1	NM_000019.3	chr11:108016927	rs145229472	none	NA	NA	A	C	0.02%	0.00%
	ASL	NM_000048.3	chr7:65552367	0	ARG,stop	649	217/465	C	T	0.02%	0.00%
	ASL	NM_000048.3	chr7:65546812	rs145138923	ARG,GLN	35	12/465	G	A	0.05%	0.30%
	ASL	NM_000048.3	chr7:65547430	rs28940585	ARG,CYS	283	95/465	C	T	0.00%	0.01%
	ASL	NM_000048.3	chr7:65547906	rs138310841	ARG,TRP	331	111/465	C	T	0.00%	0.01%
	ASL	NM_000048.3	chr7:65548107	rs143793815	THR,MET	392	131/465	C	T	0.00%	0.08%
	ASL	NM_000048.3	chr7:65548162	rs142637046	none	NA	NA	G	A	0.00%	0.01%
	ASL	NM_000048.3	chr7:65557065	rs28940287	ARG,CYS	1135	379/465	C	T	0.02%	0.00%
	ASS1	NM_000050.4	chr9:133327612	rs138350285	none	NA	NA	C	T	0.14%	0.02%
	ASS1	NM_000050.4	chr9:133333869	rs121908644	ARG,CYS	256	86/413	C	T	0.00%	0.01%
	ASS1	NM_000050.4	chr9:133333936	rs35269064	ARG,LEU	323	108/413	G	T	1.52%	0.16%
	ASS1	NM_000050.4	chr9:133342161	rs121908637	ARG,HIS	470	157/413	G	A	0.00%	0.01%
	ASS1	NM_000050.4	chr9:133346260	rs121908646	TRP,ARG	535	179/413	T	C	0.00%	0.02%
	ASS1	NM_000050.4	chr9:133355791	rs148918985	ARG,CYS	793	265/413	C	T	0.02%	0.00%
	ASS1	NM_000050.4	chr9:133355803	0	VAL,MET	805	269/413	G	A	0.02%	0.00%
	ASS1	NM_000050.4	chr9:133355833	rs121908645	ARG,stop	835	279/413	C	T	0.00%	0.01%
	ASS1	NM_000050.4	chr9:133355834	0	ARG,GLN	836	279/413	G	A	0.02%	0.00%
	ASS1	NM_000050.4	chr9:133364810	0	LYS,ARG	929	310/413	A	G	0.00%	0.01%
	ASS1	NM_000050.4	chr9:133364800	rs183276875	ARG,CYS	919	307/413	C	T	0.02%	0.00%
	ASS1	NM_000050.4	chr9:133374932	rs121908641	GLY,ARG	1168	390/413	G	A	0.02%	0.02%
	BCKDHA	NM_000709.3	chr19:41916560	0	GLN,stop	127	43/446	C	T	0.02%	0.00%
	BCKDHA	NM_000709.3	chr19:41916570	0	SER,stop	137	46/446	C	A	0.00%	0.01%
	BCKDHA	NM_000709.3	chr19:41920030	rs34442879	THR,MET	452	151/446	C	T	0.16%	1.06%
	BCKDHA	NM_000709.3	chr19:41928081	0	ALA,VAL	659	220/446	C	T	0.00%	0.01%
	BCKDHA	NM_000709.3	chr19:41928183	0	ALA,ASP	761	254/446	C	A	0.00%	0.01%
	BCKDHA	NM_000709.3	chr19:41925055	0	ARG,GLN	500	167/446	G	A	0.02%	0.00%
	BCKDHA	NM_000709.3	chr19:41928569	rs145901144	ARG,CYS	889	297/446	C	T	0.05%	0.01%
	BCKDHA	NM_000709.3	chr19:41928570	0	ARG,HIS	890	297/446	G	A	0.02%	0.01%
	BCKDHA	NM_000709.3	chr19:41930487	rs137852870	TYR,ASN	1312	438/446	T	A	0.00%	0.02%
	BCKDHB	NM_000056.3	chr6:80838934	0	ARG,stop	331	111/393	C	T	0.00%	0.01%
	BCKDHB	NM_000056.3	chr6:80878661	rs149766077	ARG,TRP	547	183/393	C	T	0.00%	0.02%
	BCKDHB	NM_000056.3	chr6:80878662	rs79761867	ARG,PRO	548	183/393	G	C	0.00%	0.05%
	BCKDHB	NM_000056.3	chr6:80910740	rs150084361	GLY,SER	832	278/393	G	A	0.00%	0.01%
	BTD	NM_000060.2	chr3:15686178	rs148031701	TRP,stop	815	272/544	G	A	0.00%	0.01%
	BTD	NM_000060.2	chr3:15677019	rs34885143	GLY,ARG	133	45/544	G	A	0.18%	1.61%
	BTD	NM_000060.2	chr3:15683446	0	GLY,VAL	341	114/544	G	T	0.00%	0.01%
	BTD	NM_000060.2	chr3:15683548	0	ARG,HIS	443	148/544	G	A	0.02%	0.00%
	BTD	NM_000060.2	chr3:15685833	rs146015592	ARG,HIS	470	157/544	G	A	0.00%	0.03%
	BTD	NM_000060.2	chr3:15685994	0	ARG,CYS	631	211/544	C	T	0.02%	0.00%
	BTD	NM_000060.2	chr3:15686568	0	ASN,SER	1205	402/544	A	G	0.00%	0.02%
	BTD	NM_000060.2	chr3:15686298	0	GLY,ASP	935	312/544	G	A	0.00%	0.01%
	BTD	NM_000060.2	chr3:15686693	rs13078881	ASP,HIS	1330	444/544	G	C	0.77%	4.15%
	BTD	NM_000060.2	chr3:15686731	rs80338685	GLN,HIS	1368	456/544	A	C	0.00%	0.15%
	BTD	NM_000060.2	chr3:15686732	rs146600671	VAL,MET	1369	457/544	G	A	0.00%	0.01%
	BTD	NM_000060.2	chr3:15686852	rs138818907	PRO,SER	1489	497/544	C	T	0.02%	0.00%

CBS	NM_000071.2	chr21:44480585	0	VAL,MET	1111	371/552	C	T	0.02%	0.00%
CBS	NM_000071.2	chr21:44480591	rs117687681	ARG,CYS	1105	369/552	G	A	0.12%	0.39%
CBS	NM_000071.2	chr21:44483098	rs121964962	GLY,SER	919	307/552	C	T	0.02%	0.00%
CBS	NM_000071.2	chr21:44484053	rs149119723	THR,MET	785	262/552	G	A	0.02%	0.01%
CBS	NM_000071.2	chr21:44484063	rs143124288	GLY,SER	775	259/552	C	T	0.02%	0.00%
CBS	NM_000071.2	chr21:44483184	rs5742905	ILE,THR	833	278/552	A	G	0.36%	0.28%
CBS	NM_000071.2	chr21:44486389	rs121964965	GLY,ARG	415	139/552	C	T	0.00%	0.01%
CBS	NM_000071.2	chr21:44486463	rs121964964	ALA,VAL	341	114/552	G	A	0.02%	0.05%
CBS	NM_000071.2	chr21:44486353	0	GLY,ARG	451	151/552	C	T	0.00%	0.01%
CBS	NM_000071.2	chr21:44492158	rs148865119	PRO,LEU	146	49/552	G	A	0.02%	0.05%
CFTR	NM_000492.3	chr7:117144378	rs143456784	SER,PHE	125	42/1481	C	T	0.00%	0.03%
CFTR	NM_000492.3	chr7:117149085	0	none	NA	NA	C	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117149101	rs77284892	GLU,stop	178	60/1481	G	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117149144	rs142540482	ARG,GLN	221	74/1481	G	A	0.05%	0.12%
CFTR	NM_000492.3	chr7:117170947	0	none	NA	NA	T	C	0.02%	0.05%
CFTR	NM_000492.3	chr7:117171007	rs113993958	ASP,TYR	328	110/1481	G	T	0.02%	0.00%
CFTR	NM_000492.3	chr7:117171028	rs77834169	ARG,CYS	349	117/1481	C	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117171169	rs78756941	none	NA	NA	G	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117174375	0	GLN,LYS	535	179/1481	C	A	0.00%	0.01%
CFTR	NM_000492.3	chr7:117175323	rs138338446	VAL,MET	601	201/1481	G	A	0.00%	0.01%
CFTR	NM_000492.3	chr7:117175339	rs121908752	LEU,TRP	617	206/1481	T	G	0.00%	0.01%
CFTR	NM_000492.3	chr7:117175372	rs121909046	GLU,GLY	650	217/1481	A	G	0.00%	0.05%
CFTR	NM_000492.3	chr7:117175437	rs147432698	GLY,ARG	715	239/1481	G	A	0.00%	0.01%
CFTR	NM_000492.3	chr7:117176711	rs151073129	ILE,PHE	853	285/1481	A	T	0.39%	0.00%
CFTR	NM_000492.3	chr7:117180174	rs143486492	ARG,GLN	890	297/1481	G	A	0.07%	0.18%
CFTR	NM_000492.3	chr7:117180324	rs77932196	ARG,PRO	1040	347/1481	G	C	0.00%	0.01%
CFTR	NM_000492.3	chr7:117180327	rs142920240	MET,LYS	1043	348/1481	T	A	0.00%	0.01%
CFTR	NM_000492.3	chr7:117180330	rs121909021	ALA,VAL	1046	349/1481	C	T	0.00%	0.05%
CFTR	NM_000492.3	chr7:117227905	0	none	NA	NA	G	A	0.00%	0.01%
CFTR	NM_000492.3	chr7:117199476	rs34906874	none	NA	NA	G	A	0.12%	0.01%
CFTR	NM_000492.3	chr7:117199525	rs139573311	LEU,PRO	1400	467/1481	T	C	0.02%	0.00%
CFTR	NM_000492.3	chr7:117199602	rs77101217	GLN,stop	1477	493/1481	C	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117199683	rs77646904	VAL,ILE	1558	520/1481	G	A	0.07%	0.00%
CFTR	NM_000492.3	chr7:117227792	rs76713772	none	NA	NA	G	A	0.00%	0.05%
CFTR	NM_000492.3	chr7:117227832	rs113993959	GLY,stop	1624	542/1481	G	T	0.02%	0.06%
CFTR	NM_000492.3	chr7:117227854	rs121908755	SER,ASN	1646	549/1481	G	A	0.02%	0.00%
CFTR	NM_000492.3	chr7:117227860	rs75527207	GLY,ASP	1652	551/1481	G	A	0.00%	0.12%
CFTR	NM_000492.3	chr7:117227865	rs74597325	ARG,stop	1657	553/1481	C	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117149123	0	PRO,LEU	200	67/1481	C	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117232470	rs140455771	PRO,LEU	2249	750/1481	C	T	0.02%	0.08%
CFTR	NM_000492.3	chr7:117232481	rs150157202	VAL,MET	2260	754/1481	G	A	0.02%	0.27%
CFTR	NM_000492.3	chr7:117232574	0	ARG,stop	2353	785/1481	C	T	0.02%	0.00%
CFTR	NM_000492.3	chr7:117234999	0	ASP,TYR	2506	836/1481	G	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117235090	0	CYS,TYR	2597	866/1481	G	A	0.00%	0.01%
CFTR	NM_000492.3	chr7:117242854	0	none	NA	NA	A	G	0.00%	0.14%
CFTR	NM_000492.3	chr7:117242922	rs80224560	none	NA	NA	G	A	0.00%	0.04%
CFTR	NM_000492.3	chr7:117243651	0	THR,ASN	2723	908/1481	C	A	0.00%	0.03%
CFTR	NM_000492.3	chr7:117243667	rs149790377	TYR,stop	2739	913/1481	T	A	0.02%	0.00%
CFTR	NM_000492.3	chr7:117243686	0	VAL,MET	2758	920/1481	G	A	0.00%	0.01%
CFTR	NM_000492.3	chr7:117171044	0	TYR,CYS	365	122/1481	A	G	0.00%	0.01%
CFTR	NM_000492.3	chr7:117243783	rs142773283	MET,THR	2855	952/1481	T	C	0.00%	0.05%
CFTR	NM_000492.3	chr7:117246808	rs75096551	none	NA	NA	G	A	0.05%	0.00%
CFTR	NM_000492.3	chr7:117250625	rs149279509	TYR,CYS	3041	1014/1481	A	G	0.00%	0.04%
CFTR	NM_000492.3	chr7:117251649	rs150212784	PHE,VAL	3154	1052/1481	T	G	0.00%	0.17%
CFTR	NM_000492.3	chr7:117251695	rs1800114	ALA,VAL	3200	1067/1481	C	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117251700	0	GLY,ARG	3205	1069/1481	G	A	0.00%	0.08%
CFTR	NM_000492.3	chr7:117251769	0	TYR,HIS	3274	1092/1481	T	C	0.00%	0.01%
CFTR	NM_000492.3	chr7:117254753	rs75541969	ASP,HIS	3454	1152/1481	G	C	0.00%	0.04%
CFTR	NM_000492.3	chr7:117267610	rs150326506	ASP,GLY	3503	1168/1481	A	G	0.02%	0.00%
CFTR	NM_000492.3	chr7:117251703	0	ARG,TRP	3208	1070/1481	C	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117267714	rs75647395	ILE,VAL	3607	1203/1481	A	G	0.02%	0.00%
CFTR	NM_000492.3	chr7:117267869	rs145743767	none	NA	NA	G	A	0.02%	0.00%
CFTR	NM_000492.3	chr7:117282468	0	none	NA	NA	G	A	0.00%	0.01%
CFTR	NM_000492.3	chr7:117251704	rs78769542	ARG,GLN	3209	1070/1481	G	A	0.02%	0.00%
CFTR	NM_000492.3	chr7:117282538	rs76649725	SER,stop	3764	1255/1481	C	A	0.02%	0.00%
CFTR	NM_000492.3	chr7:117282620	rs77010898	TRP,stop	3846	1282/1481	G	A	0.00%	0.06%
CFTR	NM_000492.3	chr7:117282649	rs146795445	none	NA	NA	T	C	0.00%	0.01%
CFTR	NM_000492.3	chr7:117292931	rs80034486	ASN,LYS	3909	1303/1481	C	G	0.00%	0.01%
CFTR	NM_000492.3	chr7:117304781	rs145545286	LEU,PHE	4003	1335/1481	C	T	0.02%	0.00%
CFTR	NM_000492.3	chr7:117304901	rs146947665	HIS,ASN	4123	1375/1481	C	A	0.00%	0.03%
CFTR	NM_000492.3	chr7:117306983	0	ARG,TRP	4264	1422/1481	C	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117305628	rs138642693	none	NA	NA	T	C	0.46%	0.00%
CFTR	NM_000492.3	chr7:117305619	0	none	NA	NA	G	T	0.00%	0.01%
CFTR	NM_000492.3	chr7:117307052	rs148783445	ASP,ASN	4333	1445/1481	G	A	0.25%	0.03%
CFTR	NM_000492.3	chr7:117307145	0	GLN,stop	4426	1476/1481	C	T	0.00%	0.01%

CYP21A2	NM_000500.7	chr6:32007593	rs6476	MET,LYS	719	240/496	T	A	1.83%	0.00%
CYP21A2	NM_000500.7	chr6:32008343	0	ARG,HIS	1100	367/496	G	A	0.02%	0.03%
DBT	NM_001918.2	chr1:100681586	0	SER,stop	725	242/483	G	C	0.00%	0.01%
DBT	NM_001918.2	chr1:100680411	0	ARG,CYS	901	301/483	G	A	0.00%	0.02%
DBT	NM_001918.2	chr1:100681641	rs74103423	GLU,stop	670	224/483	C	A	0.05%	0.00%
DBT	NM_001918.2	chr1:100680485	rs121964999	PHE,CYS	827	276/483	A	C	0.00%	0.03%
FAH	NM_000137.2	chr15:80460394	0	TRP,stop	456	152/420	G	A	0.00%	0.01%
FAH	NM_000137.2	chr15:80460605	rs80338895	none	NA	NA	G	T	0.00%	0.03%
FAH	NM_000137.2	chr15:80465355	rs149052294	none	NA	NA	G	A	0.07%	0.00%
FAH	NM_000137.2	chr15:80465431	rs80338898	PRO,LEU	782	261/420	C	T	0.00%	0.01%
FAH	NM_000137.2	chr15:80467400	0	THR,PRO	880	294/420	A	C	0.00%	0.03%
FAH	NM_000137.2	chr15:80472514	rs80338900	GLY,SER	1009	337/420	G	A	0.02%	0.00%
FAH	NM_000137.2	chr15:80472526	rs11555096	ARG,TRP	1021	341/420	C	T	0.35%	2.28%
FAH	NM_000137.2	chr15:80472572	rs80338901	none	NA	NA	G	A	0.00%	0.07%
FAH	NM_000137.2	chr15:80473411	rs121965076	GLU,stop	1090	364/420	G	T	0.02%	0.00%
GALT	NM_000155.2	chr9:34647200	rs111033656	PRO,LEU	197	66/380	C	T	0.00%	0.03%
GALT	NM_000155.2	chr9:34647855	rs111033690	SER,LEU	404	135/380	C	T	0.25%	0.00%
GALT	NM_000155.2	chr9:34647864	rs111033686	THR,MET	413	138/380	C	T	0.00%	0.01%
GALT	NM_000155.2	chr9:34647879	rs111033697	SER,LEU	428	143/380	C	T	0.00%	0.01%
GALT	NM_000155.2	chr9:34648167	rs75391579	GLN,ARG	563	188/380	A	G	0.09%	0.27%
GALT	NM_000155.2	chr9:34648373	rs111033736	GLU,LYS	607	203/380	G	A	0.00%	0.01%
GALT	NM_000155.2	chr9:34648433	rs111033750	ARG,SER	667	223/380	C	A	0.00%	0.01%
GALT	NM_000155.2	chr9:34648843	0	ARG,CYS	772	258/380	C	T	0.00%	0.01%
GALT	NM_000155.2	chr9:34648885	rs111033766	ARG,CYS	814	272/380	C	T	0.02%	0.00%
GALT	NM_000155.2	chr9:34649029	rs111033773	LYS,ASN	855	285/380	G	T	0.00%	0.06%
GALT	NM_000155.2	chr9:34649442	rs2070074	ASN,ASP	940	314/380	A	G	2.93%	9.36%
GALT	NM_000155.2	chr9:34649484	rs144993986	ARG,CYS	982	328/380	C	T	0.02%	0.01%
GALT	NM_000155.2	chr9:34649499	rs111033800	ARG,TRP	997	333/380	C	T	0.00%	0.01%
GALT	NM_000155.2	chr9:34649532	rs111033814	GLN,LYS	1030	344/380	C	A	0.00%	0.01%
GALT	NM_000155.2	chr9:34650438	rs111033819	ILE,VAL	1132	378/380	A	G	0.00%	0.03%
GCDH	NM_000159.2	chr19:13002779	rs142967670	ARG,CYS	262	88/439	C	T	0.00%	0.02%
GCDH	NM_000159.2	chr19:13004378	rs139851890	SER,LEU	416	139/439	C	T	0.02%	0.00%
GCDH	NM_000159.2	chr19:13006872	rs149120354	MET,THR	572	191/439	T	C	0.02%	0.05%
GCDH	NM_000159.2	chr19:13007063	rs121434373	ARG,PRO	680	227/439	G	C	0.02%	0.02%
GCDH	NM_000159.2	chr19:13007748	rs121434371	ALA,THR	877	293/439	G	A	0.05%	0.00%
GCDH	NM_000159.2	chr19:13007781	0	ALA,THR	910	304/439	G	A	0.02%	0.00%
GCDH	NM_000159.2	chr19:13008527	rs121434370	GLU,LYS	1093	365/439	G	A	0.00%	0.01%
GCDH	NM_000159.2	chr19:13008581	rs150938052	ARG,CYS	1147	383/439	C	T	0.02%	0.00%
GCDH	NM_000159.2	chr19:13008638	rs121434369	ARG,TRP	1204	402/439	C	T	0.02%	0.06%
GCDH	NM_000159.2	chr19:13008647	rs141437721	MET,VAL	1213	405/439	A	G	0.09%	0.00%
GCDH	NM_000159.2	chr19:13008674	rs147611168	GLU,LYS	1240	414/439	G	A	0.05%	0.01%
GCDH	NM_000159.2	chr19:13010299	rs151201155	ALA,THR	1261	421/439	G	A	0.16%	0.00%
GCH1	NM_000161.2	chr14:55369176	rs56127440	PRO,LEU	206	69/251	G	A	0.00%	0.07%
GJB2	NM_004004.5	chr13:20763104	rs111033294	ASN,SER	617	206/227	T	C	0.00%	0.01%
GJB2	NM_004004.5	chr13:20763210	0	ALA,THR	511	171/227	C	T	0.00%	0.02%
GJB2	NM_004004.5	chr13:20763222	rs111033360	VAL,MET	499	167/227	C	T	0.11%	0.00%
GJB2	NM_004004.5	chr13:20763246	0	ASP,ASN	475	159/227	C	T	0.00%	0.01%
GJB2	NM_004004.5	chr13:20763294	rs80338948	ARG,TRP	427	143/227	G	A	0.07%	0.00%
GJB2	NM_004004.5	chr13:20763305	rs76434661	SER,ASN	416	139/227	C	T	0.00%	0.08%
GJB2	NM_004004.5	chr13:20763366	rs150529554	GLU,LYS	355	119/227	C	T	0.00%	0.03%
GJB2	NM_004004.5	chr13:20763423	rs143343083	HIS,TYR	298	100/227	G	A	0.02%	0.00%
GJB2	NM_004004.5	chr13:20763452	rs80338945	LEU,PRO	269	90/227	A	G	0.02%	0.06%
GJB2	NM_004004.5	chr13:20763471	rs104894409	VAL,LEU	250	84/227	C	G	0.02%	0.00%
GJB2	NM_004004.5	chr13:20763480	rs145216882	LEU,VAL	241	81/227	G	C	0.05%	0.01%
GJB2	NM_004004.5	chr13:20763483	0	GLN,stop	238	80/227	G	A	0.02%	0.00%
GJB2	NM_004004.5	chr13:20763490	rs80338944	TRP,stop	231	77/227	C	T	0.00%	0.01%
GJB2	NM_004004.5	chr13:20763552	rs111033297	GLN,stop	169	57/227	G	A	0.00%	0.01%
GJB2	NM_004004.5	chr13:20763395	0	GLY,GLU	326	109/227	C	T	0.00%	0.01%
GJB2	NM_004004.5	chr13:20763611	rs141774369	VAL,ALA	110	37/227	A	G	0.05%	0.00%
GJB2	NM_004004.5	chr13:20763612	rs72474224	VAL,ILE	109	37/227	C	T	0.02%	0.19%
GJB2	NM_004004.5	chr13:20763627	0	ARG,CYS	94	32/227	G	A	0.02%	0.00%
GJB2	NM_004004.5	chr13:20763633	0	ILE,VAL	88	30/227	T	C	0.00%	0.01%
GJB2	NM_004004.5	chr13:20763710	rs111033222	GLY,ASP	11	4/227	C	T	0.21%	0.00%
GJB2	NM_004004.5	chr13:20763534	0	VAL,MET	187	63/227	C	T	0.00%	0.01%
GJB2	NM_004004.5	chr13:20763602	rs111033296	ALA,GLU	119	40/227	G	T	0.00%	0.01%
HADHA	NM_000182.4	chr2:26414191	rs142120825	TYR,stop	2220	740/764	A	T	0.02%	0.00%
HADHA	NM_000182.4	chr2:26418053	rs137852769	GLU,GLN	1528	510/764	C	G	0.02%	0.05%
HADHA	NM_000182.4	chr2:26435497	0	none	NA	NA	T	C	0.00%	0.01%
HADHA	NM_000182.4	chr2:26437990	0	ALA,VAL	731	244/764	G	A	0.00%	0.01%
HADHB	NM_000183.2	chr2:26486320	rs121913132	ARG,HIS	182	61/475	G	A	0.00%	0.02%
HADHB	NM_000183.2	chr2:26496605	rs146328300	ASN,SER	341	114/475	A	G	0.00%	0.01%
HADHB	NM_000183.2	chr2:26508339	0	PHE,SER	1289	430/475	T	C	0.00%	0.01%
HBB	NM_000518.4	chr11:5246908	rs33946267	GLU,GLN	364	122/148	C	G	0.00%	0.02%
HBB	NM_000518.4	chr11:5246959	rs33913413	none	NA	NA	G	T	0.02%	0.01%
HBB	NM_000518.4	chr11:5247806	rs33945777	none	NA	NA	C	T	0.00%	0.01%

HBB	NM_000518.4	chr11:5247859	rs33993568	THR,ILE	263	88/148	G	A	0.00%	0.02%
HBB	NM_000518.4	chr11:5247860	rs35553496	THR,PRO	262	88/148	T	G	0.00%	0.01%
HBB	NM_000518.4	chr11:5248004	rs76728603	GLN,stop	118	40/148	G	A	0.00%	0.01%
HBB	NM_000518.4	chr11:5248050	rs35004220	none	NA	NA	C	T	0.00%	0.02%
HBB	NM_000518.4	chr11:5248052	rs111851677	none	NA	NA	A	G	1.57%	0.00%
HBB	NM_000518.4	chr11:5248159	rs33971440	none	NA	NA	C	T	0.00%	0.01%
HBB	NM_000518.4	chr11:5248170	rs35424040	ALA,SER	82	28/148	C	A	0.02%	0.00%
HBB	NM_000518.4	chr11:5248177	rs75680770	none	75	25/148	A	T	0.02%	0.00%
HBB	NM_000518.4	chr11:5248232	rs334	GLU,VAL	20	7/148	T	A	4.00%	0.02%
HBB	NM_000518.4	chr11:5248233	rs33930165	GLU,LYS	19	7/148	C	T	1.61%	0.01%
HBB	NM_000518.4	chr11:5248282	rs63750628	none	NA	NA	G	A	0.00%	0.01%
HLCS	NM_000411.6	chr21:38309329	rs144572349	LEU,stop	416	139/727	A	T	0.00%	0.01%
HLCS	NM_000411.6	chr21:38128859	rs146448211	ARG,stop	1993	665/727	G	A	0.00%	0.02%
HLCS	NM_000411.6	chr21:38128865	rs140951243	TYR,HIS	1987	663/727	A	G	0.02%	0.00%
HLCS	NM_000411.6	chr21:38128952	rs149399432	ASP,ASN	1900	634/727	C	T	0.02%	0.00%
HLCS	NM_000411.6	chr21:38132079	0	GLY,ARG	1744	582/727	C	T	0.00%	0.01%
IVD	NM_001159508.1	chr15:40708531	0	GLN,stop	1027	343/397	C	T	0.02%	0.00%
IVD	NM_001159508.1	chr15:40702898	rs142761835	GLY,ARG	277	93/397	G	A	0.02%	0.00%
IVD	NM_001159508.1	chr15:40707154	0	ARG,GLN	770	257/397	G	A	0.02%	0.00%
IVD	NM_001159508.1	chr15:40707653	rs28940889	ALA,VAL	851	284/397	C	T	0.07%	0.07%
IVD	NM_001159508.1	chr15:40710364	0	ARG,CYS	1093	365/397	C	T	0.02%	0.00%
MCCC1	NM_020166.3	chr3:182763210	0	TRP,stop	1074	358/726	C	T	0.00%	0.01%
MCCC1	NM_020166.3	chr3:182763310	rs119103212	MET,ARG	974	325/726	A	C	0.00%	0.02%
MCCC2	NM_022132.4	chr5:70898412	rs141030969	ARG,TRP	463	155/564	C	T	0.00%	0.02%
MCCC2	NM_022132.4	chr5:70936845	rs150591260	VAL,MET	1015	339/564	G	A	0.02%	0.09%
MCCC2	NM_022132.4	chr5:70942096	rs142887940	ASN,SER	1208	403/564	A	G	0.00%	0.06%
MCCC2	NM_022132.4	chr5:70945029	rs139852818	ILE,THR	1322	441/564	T	C	0.05%	0.12%
MCCC2	NM_022132.4	chr5:70948566	rs150327768	TYR,SER	1559	520/564	A	C	0.00%	0.01%
MCCC2	NM_022132.4	chr5:70895499	rs119103219	GLU,GLN	295	99/564	G	C	0.02%	0.02%
MMAA	NM_172250.2	chr4:146560724	rs104893851	ARG,stop	433	145/419	C	T	0.00%	0.05%
MMAB	NM_052845.3	chr12:109994886	0	GLN,stop	700	234/251	G	A	0.00%	0.01%
MMAB	NM_052845.3	chr12:109998858	0	ARG,TRP	571	191/251	G	A	0.00%	0.01%
MMADHC	NM_015702.2	chr2:150432296	0	GLN,stop	538	180/297	G	A	0.02%	0.00%
MUT	NM_000255.3	chr6:49399544	rs121918252	GLY,VAL	2150	717/751	C	A	0.07%	0.00%
MUT	NM_000255.3	chr6:49403194	rs140600746	MET,LYS	2099	700/751	A	T	0.00%	0.01%
MUT	NM_000255.3	chr6:49403260	rs147094927	HIS,ARG	2033	678/751	T	C	0.02%	0.00%
MUT	NM_000255.3	chr6:49407986	rs143023066	GLY,GLU	1889	630/751	C	T	0.00%	0.01%
MUT	NM_000255.3	chr6:49408008	rs121918254	GLY,ARG	1867	623/751	C	T	0.02%	0.00%
MUT	NM_000255.3	chr6:49407995	0	HIS,ARG	1880	627/751	T	C	0.00%	0.01%
MUT	NM_000255.3	chr6:49425601	rs148331800	MET,VAL	556	186/751	T	C	0.00%	0.01%
MUT	NM_000255.3	chr6:49419403	0	THR,PRO	1108	370/751	T	G	0.02%	0.00%
MUT	NM_000255.3	chr6:49426975	rs115923556	ILE,VAL	205	69/751	T	C	0.14%	0.38%
PAH	NM_000277.1	chr12:103234177	rs5030861	none	NA	NA	C	T	0.02%	0.10%
PAH	NM_000277.1	chr12:103234252	rs5030860	TYR,CYS	1241	414/453	T	C	0.00%	0.05%
PAH	NM_000277.1	chr12:103234271	rs5030858	ARG,TRP	1222	408/453	G	A	0.00%	0.17%
PAH	NM_000277.1	chr12:103234285	rs5030857	ALA,VAL	1208	403/453	G	A	0.00%	0.06%
PAH	NM_000277.1	chr12:103237439	rs62508736	ALA,GLY	1184	395/453	G	C	0.00%	0.01%
PAH	NM_000277.1	chr12:103237454	rs5030856	GLU,GLY	1169	390/453	T	C	0.00%	0.01%
PAH	NM_000277.1	chr12:103237461	rs62516101	VAL,MET	1162	388/453	C	T	0.02%	0.00%
PAH	NM_000277.1	chr12:103237466	rs62516141	TYR,CYS	1157	386/453	T	C	0.00%	0.01%
PAH	NM_000277.1	chr12:103237484	rs62642937	THR,MET	1139	380/453	G	A	0.00%	0.07%
PAH	NM_000277.1	chr12:103237557	rs62507320	TYR,HIS	1066	356/453	A	G	0.00%	0.03%
PAH	NM_000277.1	chr12:103237568	rs5030855	none	NA	NA	C	T	0.00%	0.06%
PAH	NM_000277.1	chr12:103238137	rs62516092	LEU,VAL	1042	348/453	G	C	0.00%	0.01%
PAH	NM_000277.1	chr12:103240716	rs62642935	ALA,VAL	926	309/453	G	A	0.00%	0.01%
PAH	NM_000277.1	chr12:103245464	rs62514956	none	NA	NA	C	T	0.00%	0.02%
PAH	NM_000277.1	chr12:103245479	rs5030853	ALA,SER	898	300/453	C	A	0.00%	0.05%
PAH	NM_000277.1	chr12:103245481	rs62642933	PHE,CYS	896	299/453	A	C	0.00%	0.03%
PAH	NM_000277.1	chr12:103246588	rs62516146	none	NA	NA	C	T	0.00%	0.01%
PAH	NM_000277.1	chr12:103246593	rs5030851	PRO,LEU	842	281/453	G	A	0.00%	0.01%
PAH	NM_000277.1	chr12:103246597	rs62508698	GLU,LYS	838	280/453	C	T	0.00%	0.02%
PAH	NM_000277.1	chr12:103246612	rs62508691	PRO,SER	823	275/453	G	A	0.00%	0.01%
PAH	NM_000277.1	chr12:103246615	rs142934616	LYS,GLU	820	274/453	T	C	1.63%	0.00%
PAH	NM_000277.1	chr12:103246624	rs62517164	HIS,TYR	811	271/453	G	A	0.00%	0.01%
PAH	NM_000277.1	chr12:103246653	rs5030849	ARG,GLN	782	261/453	C	T	0.05%	0.06%
PAH	NM_000277.1	chr12:103246681	rs5030847	ARG,TRP	754	252/453	G	A	0.00%	0.02%
PAH	NM_000277.1	chr12:103246701	rs76212747	VAL,ALA	734	245/453	A	G	0.02%	0.07%
PAH	NM_000277.1	chr12:103246713	rs62508730	ARG,HIS	722	241/453	C	T	0.00%	0.01%
PAH	NM_000277.1	chr12:103248926	rs62507348	GLN,stop	694	232/453	G	A	0.05%	0.00%
PAH	NM_000277.1	chr12:103248932	rs62516152	VAL,ILE	688	230/453	C	T	0.00%	0.03%
PAH	NM_000277.1	chr12:103246698	0	ALA,VAL	737	246/453	G	A	0.00%	0.01%
PAH	NM_000277.1	chr12:103249091	0	VAL,LEU	529	177/453	C	G	0.00%	0.01%
PAH	NM_000277.1	chr12:103249093	rs74486803	ARG,LEU	527	176/453	C	A	0.02%	0.01%
PAH	NM_000277.1	chr12:103249099	rs138809906	ILE,ASN	521	174/453	A	T	0.02%	0.00%
PAH	NM_000277.1	chr12:103260383	rs77554925	ASN,SER	500	167/453	T	C	1.66%	0.00%

	PAH	NM_000277.1	chr12:103260410	rs5030843	ARG,GLN	473	158/453	C	T	0.00%	0.03%
	PAH	NM_000277.1	chr12:103260411	rs75166491	ARG,TRP	472	158/453	G	A	0.02%	0.01%
	PAH	NM_000277.1	chr12:103260446	rs62514909	none	NA	NA	G	C	0.00%	0.01%
	PAH	NM_000277.1	chr12:103271235	rs62507321	none	NA	NA	C	A	0.00%	0.01%
	PAH	NM_000277.1	chr12:103271239	rs62517166	none	NA	NA	C	T	0.00%	0.01%
	PAH	NM_000277.1	chr12:103271247	rs140175796	ASP,VAL	434	145/453	T	A	0.02%	0.01%
	PAH	NM_000277.1	chr12:103288566	rs148393887	HIS,ARG	299	100/453	T	C	0.05%	0.05%
	PAH	NM_000277.1	chr12:103288576	rs142516271	ILE,LEU	289	97/453	T	G	0.20%	0.00%
	PAH	NM_000277.1	chr12:103288590	rs62514903	THR,ILE	275	92/453	G	A	0.00%	0.01%
	PAH	NM_000277.1	chr12:103288671	rs75193786	ILE,THR	194	65/453	A	G	0.00%	0.06%
	PAH	NM_000277.1	chr12:103306579	rs118092776	ARG,HIS	158	53/453	C	T	0.00%	0.09%
	PAH	NM_000277.1	chr12:103306594	rs5030841	LEU,SER	143	48/453	A	G	0.00%	0.01%
	PAH	NM_000277.1	chr12:103306620	rs62642926	PHE,LEU	117	39/453	G	C	0.05%	0.01%
	PAH	NM_000277.1	chr12:103310879	rs1801145	none	30	10/453	G	C	0.00%	0.07%
	PCBD1	NM_000281.2	chr10:72643730	rs121913015	GLN,stop	292	98/105	G	A	0.00%	0.02%
	PCBD1	NM_000281.2	chr10:72643759	rs115117837	ARG,GLN	263	88/105	C	T	0.16%	0.01%
	PCCA	NM_000282.3	chr13:101167714	rs145428347	GLU,stop	1933	645/729	G	T	0.02%	0.00%
	PCCA	NM_000282.3	chr13:100764140	rs141371306	ARG,TRP	229	77/729	C	T	0.02%	0.00%
	PCCA	NM_000282.3	chr13:101020733	rs61749895	VAL,PHE	1651	551/729	G	T	0.23%	0.81%
	PCCB	NM_000532.4	chr3:136016902	rs77820367	CYS,TYR	872	291/540	G	A	0.16%	0.29%
	PCCB	NM_000532.4	chr3:136012626	0	PRO,LEU	683	228/540	C	T	0.00%	0.01%
	PCCB	NM_000532.4	chr3:136046480	rs121964961	TYR,CYS	1304	435/540	A	G	0.02%	0.00%
	PTS	NM_000317.2	chr11:112103939	rs145882709	TYR,stop	297	99/146	C	A	0.00%	0.01%
	PTS	NM_000317.2	chr11:112101362	0	THR,MET	200	67/146	C	T	0.00%	0.01%
	PTS	NM_000317.2	chr11:112104210	rs150726932	VAL,LEU	370	124/146	G	T	0.00%	0.01%
	SLC22A5	NM_003060.3	chr5:131705698	rs139203363	GLY,SER	34	12/558	G	A	0.00%	0.17%
	SLC22A5	NM_003060.3	chr5:131719973	rs121908888	TYR,CYS	632	211/558	A	G	0.00%	0.01%
	SLC22A5	NM_003060.3	chr5:131721062	rs114269482	THR,MET	695	232/558	C	T	0.00%	0.03%
	SLC22A5	NM_003060.3	chr5:131721136	0	ARG,TRP	769	257/558	C	T	0.00%	0.01%
	SLC22A5	NM_003060.3	chr5:131722736	rs121908886	ARG,stop	844	282/558	C	T	0.00%	0.01%
	SLC22A5	NM_003060.3	chr5:131722731	0	SER,PHE	839	280/558	C	T	0.00%	0.01%
	SLC22A5	NM_003060.3	chr5:131726522	rs144547521	PRO,LEU	1193	398/558	C	T	0.02%	0.01%
	SLC22A5	NM_003060.3	chr5:131728202	rs11568514	TYR,ASP	1345	449/558	T	G	0.36%	0.00%
	SLC22A5	NM_003060.3	chr5:131729379	0	ARG,CYS	1462	488/558	C	T	0.00%	0.01%
	SLC22A5	NM_003060.3	chr5:131729380	rs28383481	ARG,HIS	1463	488/558	G	A	0.18%	0.52%
Age related	*ABCA4*	NM_000350.2	chr1:94463425	rs61748521	LEU,VAL	6721	2241/2274	G	C	0.00%	0.37%
macular	*ABCA4*	NM_000350.2	chr1:94463617	rs1800555	ASP,ASN	6529	2177/2274	C	T	0.20%	0.11%
degeneration	*ABCA4*	NM_000350.2	chr1:94473287	rs28938473	LEU,PHE	5908	1970/2274	G	A	0.20%	0.02%
(ARMD)	*ABCA4*	NM_000350.2	chr1:94473807	rs1800553	GLY,GLU	5882	1961/2274	C	T	0.11%	0.01%
	ABCA4	NM_000350.2	chr1:94476377	rs1800552	ARG,HIS	5693	1898/2274	C	T	0.05%	0.01%
	ABCA4	NM_000350.2	chr1:94487443	rs1800551	GLY,ARG	4732	1578/2274	C	T	0.00%	0.01%
	ABCA4	NM_000350.2	chr1:94487490	rs1762111	ILE,THR	4685	1562/2274	A	G	0.02%	0.74%
	ABCA4	NM_000350.2	chr1:94496053	rs1800549	THR,MET	4283	1428/2274	G	A	0.02%	0.00%
	ABCA4	NM_000350.2	chr1:94496666	rs61750130	PRO,LEU	4139	1380/2274	G	A	0.00%	0.01%
	ABCA4	NM_000350.2	chr1:94508969	rs61751374	ALA,VAL	3113	1038/2274	G	A	0.07%	0.00%
	ABCA4	NM_000350.2	chr1:94512565	rs1801581	ARG,GLN	2828	943/2274	C	T	1.43%	0.00%
	ABCA4	NM_000350.2	chr1:94512574	rs144995371	PRO,ARG	2819	940/2274	G	C	0.00%	0.01%
	ABCA4	NM_000350.2	chr1:94514466	rs139655975	THR,ALA	2701	901/2274	T	C	0.02%	0.00%
	ABCA4	NM_000350.2	chr1:94543389	rs1800548	GLU,LYS	1411	471/2274	C	T	0.00%	0.01%
	ABCA4	NM_000350.2	chr1:94564483	rs6657239	ARG,HIS	635	212/2274	C	T	7.06%	0.13%
	ABCA4	NM_000350.2	chr1:94461676	0	ARG,stop	6805	2269/2274	G	A	0.00%	0.03%
	ABCA4	NM_000350.2	chr1:94466426	rs140142529	ARG,stop	6445	2149/2274	G	A	0.00%	0.01%
	ABCA4	NM_000350.2	chr1:94508343	0	TRP,stop	3302	1101/2274	C	T	0.00%	0.00%
	ABCA4	NM_000350.2	chr1:94544906	rs150686179	SER,stop	1211	404/2274	G	C	0.00%	0.01%
	C3	NM_000064.2	chr19:6718387	rs2230199	ARG,GLY	304	102/1664	G	C	4.81%	0.07%
	CFH	NM_000186.3	chr1:196642233	rs800292	VAL,ILE	184	62/1232	G	A	70.71%	0.34%
	CFH	NM_000186.3	chr1:196659237	rs1061170	HIS,TYR	1204	402/1232	**C**	**C**	36.32%	0.01%
	CFH	NM_000186.3	chr1:196682947	rs2274700	none	1419	473/1232	G	A	46.86%	0.01%
	CFH	NM_000186.3	chr1:196709774	rs1065489	GLU,ASP	2808	936/1232	G	T	6.62%	0.00%
	CFH	NM_000186.3	chr1:196716375	rs121913059	ARG,CYS	3628	1210/1232	C	T	0.00%	0.00%
Drug	*ABCB1*	NM_000927.4	chr7:87138645	rs1045642	none	3435	1145/1281	A	G	22.71%	47.75%
response	*ABCB1*	NM_000927.4	chr7:87160561	rs2032583	none	NA	NA	A	G	17.11%	13.25%
(PGx)	*ABCB1*	NM_000927.4	chr7:87160618	rs2032582	SER,ALA	2677	893/1281	A	C	10.75%	43.13%
	ABCB1	NM_000927.4	chr7:87179601	rs1128503	none	1236	412/1281	A	G	21.55%	42.94%
	ABCB1	NM_000927.4	chr7:87199564	rs2235015	none	NA	NA	C	A	34.35%	19.69%
	ABCC2	NM_000392.3	chr10:101563815	rs2273697	VAL,ILE	1249	417/1546	G	A	18.69%	19.43%
	ABCC2	NM_000392.3	chr10:101595996	rs17222723	VAL,GLU	3563	1188/1546	T	A	5.97%	6.17%
	ABCC2	NM_000392.3	chr10:101604207	rs3740066	none	3972	1324/1546	C	T	26.02%	37.16%
	ABCC2	NM_000392.3	chr10:101611294	rs8187710	CYS,TYR	4544	1515/1546	G	A	15.72%	6.19%
	ABCC2	NM_000392.3	chr10:101594183	0	TRP,stop	3305	1102/1546	G	A	0.00%	0.01%
	ABCC2	NM_000392.3	chr10:101591385	rs17222547	TYR,stop	2901	967/1546	C	A	0.02%	0.00%
	ABCC2	NM_000392.3	chr10:101610372	rs145715632	GLN,stop	4327	1443/1546	C	T	0.02%	0.00%
	ABCG2	NM_004827.2	chr4:89052323	rs2231142	GLN,LYS	421	141/656	G	T	3.22%	11.14%
	COMT	NM_000754.3	chr22:19951271	rs4680	VAL,MET	472	158/272	G	A	68.36%	47.79%
	COMT	NM_000754.3	chr22:19956180	0	SER,stop	737	246/272	C	A	0.00%	0.01%

CYP2C19	NM_000769.1	chr10:96540410	rs4986893	TRP,stop	636	212/491	G	A	0.05%	0.02%
CYP2C19	NM_000769.1	chr10:96535278	0	GLU,stop	463	155/491	G	T	0.00%	0.01%
CYP2C9	NM_000771.3	chr10:96702047	rs1799853	ARG,CYS	430	144/491	C	T	2.68%	13.03%
CYP2C9	NM_000771.3	chr10:96702066	rs7900194	ARG,HIS	449	150/491	G	A	5.81%	0.05%
CYP2C9	NM_000771.3	chr10:96740981	rs28371685	ARG,TRP	1003	335/491	C	T	1.91%	0.19%
CYP2C9	NM_000771.3	chr10:96741053	rs1057910	ILE,LEU	1075	359/491	A	C	1.43%	6.61%
CYP2C9	NM_000771.3	chr10:96741058	rs28371686	ASP,GLU	1080	360/491	C	G	1.07%	0.01%
DPYD	NM_000110.3	chr1:97915614	rs3918290	none	NA	NA	C	T	0.09%	0.58%
DPYD	NM_000110.3	chr1:97981395	rs1801159	ILE,VAL	1627	543/1026	T	C	15.51%	19.77%
DPYD	NM_000110.3	chr1:97981421	rs1801158	SER,ASN	1601	534/1026	C	T	0.45%	2.02%
DPYD	NM_000110.3	chr1:98165091	rs2297595	MET,VAL	496	166/1026	T	C	4.10%	9.90%
DPYD	NM_000110.3	chr1:98348885	rs1801265	ARG,CYS	85	29/1026	G	A	39.97%	22.50%
DPYD	NM_000110.3	chr1:97770839	0	ARG,stop	2275	759/1026	G	A	0.02%	0.00%
DPYD	NM_000110.3	chr1:98164926	rs146170505	GLU,stop	661	221/1026	C	A	0.00%	0.01%
DPYD	NM_000110.3	chr1:98293695	rs141597515	ARG,stop	208	70/1026	G	A	0.00%	0.01%
DRD2	NM_000795.3	chr11:113283459	rs6277	none	957	319/444	G	A	85.69%	45.40%
EPHX1	NM_000120.3	chr1:226019633	rs1051740	TYR,HIS	337	113/456	T	C	17.63%	29.96%
EPHX1	NM_000120.3	chr1:226019653	rs2292566	none	357	119/456	G	A	14.59%	14.43%
EPHX1	NM_000120.3	chr1:226026406	rs2234922	HIS,ARG	416	139/456	A	G	33.69%	20.41%
EPHX1	NM_000120.3	chr1:226026525	0	GLU,stop	535	179/456	G	T	0.00%	0.01%
EPHX1	NM_000120.3	chr1:226033029	0	SER,stop	1349	450/456	C	A	0.02%	0.00%
SLCO1B1	NM_006446.4	chr12:21329738	rs2306283	ASN,ASP	388	130/692	A	G	23.38%	40.27%
SLCO1B1	NM_006446.4	chr12:21329813	rs11045819	PRO,THR	463	155/692	C	A	8.62%	16.16%
SLCO1B1	NM_006446.4	chr12:21331549	rs4149056	VAL,ALA	521	174/692	T	C	3.63%	15.48%
SLCO1B1	NM_006446.4	chr12:21294563	0	LYS,stop	55	19/692	A	T	0.02%	0.00%
TPMT	NM_000367.2	chr6:18130918	rs1142345	TYR,CYS	719	240/246	T	C	5.26%	4.24%
TPMT	NM_000367.2	chr6:18131012	rs1800584	none	NA	NA	C	T	0.00%	0.01%
TPMT	NM_000367.2	chr6:18139228	rs1800460	ALA,THR	460	154/246	C	T	1.00%	3.80%
TPMT	NM_000367.2	chr6:18143955	rs1800462	ALA,PRO	238	80/246	C	G	0.00%	0.23%
UGT1A10	NM_000463.2	chr2:234669144	rs4148323	GLY,ARG	211	71/534	G	A	0.14%	0.13%
UGT1A10	NM_000463.2	chr2:234675792	0	LEU,stop	977	326/534	T	A	0.02%	0.00%

Supplemental Acknowledgments

HeartGO:

Atherosclerosis Risk in Communities (ARIC): NHLBI (N01 HC-55015, N01 HC-55016, N01HC-55017, N01 HC-55018, N01 HC-55019, N01 HC-55020, N01 HC-55021); **Cardiovascular Health Study (CHS)**: NHLBI (HHSN268201200036C, HHSN268200800007C, N01-HC-85239, N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, and grant HL080295), with additional support from NINDS and from NIA (AG-023629, AG-15928, AG-20098, and AG-027058); **Coronary Artery Risk Development in Young Adults (CARDIA)**: NHLBI (N01-HC95095 & N01-HC48047, N01-HC48048, N01-HC48049, and N01-HC48050); **Framingham Heart Study (FHS)**: NHLBI (N01-HC-25195 and grant R01 NS17950) with additional support from NIA (AG08122 and AG033193); **Jackson Heart Study (JHS)**: NHLBI and the National Institute on Minority Health and Health Disparities (N01 HC-95170, N01 HC-95171 and N01 HC-95172); **Multi-Ethnic Study of Atherosclerosis (MESA):** NHLBI (N01-HC-95159 through N01-HC-95169 and RR-024156).

Lung GO:

Cystic Fibrosis (CF): Cystic Fibrosis Foundation (GIBSON07K0, KNOWLE00A0, OBSERV04K0, RDP R026), the NHLBI (R01 HL-068890, R02 HL-095396), NIH National Center for Research Resources (UL1 RR-025014), and the National Human Genome Research Institute (NHGRI) (5R00 HG-004316). **Chronic Obstructive Pulmonary Disease (COPDGene):** NHLBI (U01 HL-089897, U01 HL-089856), and the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer, and Sunovian. The COPDGene clinical centers and investigators are available at www.copdgene.org. **Acute Lung Injury (ALI)**: NHLBI (RC2 HL-101779). **Lung Health Study (LHS)**: NHLBI (RC2 HL-066583), the NHGRI (HG-004738), and the NHLBI Division of Lung Diseases (HR-46002). **Pulmonary Arterial Hypertension (PAH):** NIH (P50 HL-084946, K23 AR-52742), and the NHLBI (F32 HL-083714). **Asthma:** NHLBI (RC2 HL-101651), and the NIH (HL-077916, HL-69197, HL-76285, M01 RR-07122).

SWISS and ISGS:

Siblings with Ischemic Stroke Study (SWISS): National Institute of Neurological Disorders and Stroke (NINDS) (R01 NS039987); Ischemic Stroke Genetics Study (ISGS): NINDS (R01 NS042733)

WHISP:

Women's Health Initiative (WHI): The WHI Sequencing Project is funded by the National Heart, Lung, and Blood Institute (HL-102924) as well as the National Institutes of Health (NIH), U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: *https://cleo.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf*

NHLBI GO Exome Sequencing Project

BroadGO

Stacey B. Gabriel (Broad Institute)[4, 5, 11, 16, 17], David M. Altshuler (Broad Institute, Harvard Medical School, Massachusetts General Hospital)[1, 5, 7, 17], Gonçalo R. Abecasis (University of Michigan)[3, 5, 9, 13, 15, 17], Hooman Allayee (University of Southern California)[5], Sharon Cresci (Washington University School of Medicine)[5], Mark J. Daly (Broad Institute, Massachusetts General Hospital), Paul I. W. de Bakker (Broad Institute, Harvard Medical School, University Medical Center Utrecht)[3, 15], Mark A. DePristo (Broad Institute)[4, 13, 15, 16], Ron Do (Broad Institute)[5, 9, 13, 15], Peter Donnelly (University of Oxford)[5], Deborah N. Farlow (Broad Institute)[3, 4, 5, 12, 14, 16, 17], Tim Fennell (Broad Institute), Kiran Garimella (University of Oxford)[4, 16], Stanley L. Hazen (Cleveland Clinic)[5], Youna Hu (University of Michigan)[3, 9, 15], Daniel M. Jordan (Harvard Medical School, Harvard University)[13], Goo Jun (University of Michigan)[13], Sekar Kathiresan (Broad Institute, Harvard Medical School, Massachusetts General Hospital)[5, 8, 9, 12, 14, 15, 17, 20], Hyun Min Kang (University of Michigan)[9, 13, 16], Adam Kiezun (Broad Institute)[5, 13, 15], Guillaume Lettre (Broad Institute, Montreal Heart Institute, Université de Montréal)[1, 2, 13, 15], Bingshan Li (University of Michigan)[3], Mingyao Li (University of Pennsylvania)[5], Christopher H. Newton-Cheh (Broad Institute, Massachusetts General Hospital, Harvard Medical School)[3, 8, 15], Sandosh Padmanabhan (University of Glasgow School of Medicine)[3, 12, 15], Gina Peloso (Broad Institute, Harvard Medical School, Massachusetts General Hospital)[5], Sara Pulit (Broad Institute)[3, 15], Daniel J. Rader (University of Pennsylvania)[5], David Reich (Broad Institute, Harvard Medical School)[15], Muredach P. Reilly (University of Pennsylvania)[5], Manuel A. Rivas (Broad Institute, Massachusetts General Hospital)[5], Steve Schwartz (Fred Hutchinson Cancer Research Center)[5, 12], Laura Scott (University of Michigan)[1], David S. Siscovick (University of Washington)[5, 1, 25], John A. Spertus (University of Missouri Kansas City)[5], Nathan O. Stitziel (Brigham and Women's Hospital)[5, 15], Nina Stoletzki (Brigham and Women's Hospital, Broad Institute, Harvard Medical School)[13], Shamil R. Sunyaev (Brigham and Women's Hospital, Broad Institute, Harvard Medical School)[1, 3, 5, 13, 15], Benjamin F. Voight (Broad Institute, Massachusetts General Hospital), Cristen J. Willer (University of Michigan)[1, 9, 13, 15]

HeartGO

Stephen S. Rich (University of Virginia)[2, 4, 7, 8, 9, 11,14, 15, 17, 18, 31], Ermeg Akylbekova (Jackson State University, University of Mississippi Medical Center)[29], Larry D. Atwood* (Boston University)[1, 11, 28], Christie M. Ballantyne (Baylor College of Medicine, Methodist DeBakey Heart Center)[9, 22], Maja Barbalic (University of Texas Health Science Center Houston)[9, 14, 15, 17, 22], R. Graham Barr (Columbia University Medical Center)[10, 31], Emelia J. Benjamin (Boston University)[14, 20, 28], Joshua Bis (University of Washington)[15, 23], Eric Boerwinkle (University of Texas Health Science Center Houston)[3, 5, 9, 13, 15, 17, 22], Donald W. Bowden (Wake Forest University)[1, 31], Jennifer Brody (University of Washington)[3, 5, 15, 23], Matthew Budoff (Harbor-UCLA Medical Center)[31], Greg Burke (Wake Forest University)[5, 31], Sarah Buxbaum (Jackson State University)[3, 13, 15, 29], Jeff Carr (Wake Forest University)[25, 29, 31], Donna T. Chen (University of Virginia)[6, 11], Ida Y. Chen (Cedars-Sinai Medical Center)[1, 31], Wei-Min Chen (University of Virginia)[13, 15, 18], Pat Concannon (University of Virginia)[11], Jacy Crosby (University of Texas Health Science Center Houston)[22], L. Adrienne Cupples (Boston University)[1, 3, 5, 9, 13, 15, 18, 28], Ralph D'Agostino (Boston University)[28], Anita L. DeStefano (Boston University)[13, 18, 28], Albert Dreisbach (University of Mississippi Medical Center)[3, 29], Josée Dupuis (Boston University)[1, 28], J. Peter Durda (University of Vermont)[15, 23], Jaclyn Ellis (University of North Carolina Chapel Hill)[1], Aaron R. Folsom (University of Minnesota)[5, 22], Myriam Fornage (University of Texas Health Science Center Houston)[3, 18, 25], Caroline S. Fox (National Heart, Lung, and Blood Institute)[1, 28], Ervin Fox (University of Mississippi Medical Center)[3, 9, 29], Vincent Funari (Cedars-Sinai Medical Center)[1, 11, 31], Santhi K. Ganesh (University of Michigan)[2, 22],

Julius Gardin (Hackensack University Medical Center)[25], David Goff (Wake Forest University)[25], Ora Gordon (Cedars-Sinai Medical Center)[11, 31], Wayne Grody (University of California Los Angeles)[11, 31], Myron Gross (University of Minnesota)[1, 5, 14, 25], Xiuqing Guo (Cedars-Sinai Medical Center)[3, 15, 31], Ira M. Hall (University of Virginia), Nancy L. Heard-Costa (Boston University)[1, 11, 28], Susan R. Heckbert (University of Washington)[10, 14, 20, 23], Nicholas Heintz (University of Vermont), David M. Herrington (Wake Forest University)[5, 31], DeMarc Hickson (Jackson State University, University of Mississippi Medical Center)[29], Jie Huang (National Heart, Lung, and Blood Institute)[5, 28], Shih-Jen Hwang (Boston University, National Heart, Lung, and Blood Institute)[3, 28], David R. Jacobs (University of Minnesota)[25], Nancy S. Jenny (University of Vermont)[1, 2, 23], Andrew D. Johnson (National Heart, Lung, and Blood Institute)[2, 5, 11, 28], Craig W. Johnson (University of Washington)[15, 31], Steven Kawut (University of Pennsylvania)[10,31], Richard Kronmal (University of Washington)[31], Raluca Kurz (Cedars-Sinai Medical Center)[11, 31], Ethan M. Lange (University of North Carolina Chapel Hill)[3, 5, 9, 13, 34], Leslie A. Lange (University of North Carolina Chapel Hill)[1, 2, 3, 5, 9, 12, 13, 15, 17, 18, 20, 25, 34], Martin G. Larson (Boston University)[3, 15, 28], Mark Lawson (University of Virginia), Cora E. Lewis (University of Alabama at Birmingham)[25,34], Daniel Levy (National Heart, Lung, and Blood Institute)[3, 15, 17, 28], Dalin Li (Cedars-Sinai Medical Center)[11, 15, 31], Honghuang Lin (Boston University)[20, 28], Chunyu Liu (National Heart, Lung, and Blood Institute)[3, 28], Jiankang Liu (University of Mississippi Medical Center)[1, 29], Kiang Liu (Northwestern University)[25], Xiaoming Liu (University of Texas Health Science Center Houston)[15, 22], Yongmei Liu (Wake Forest University)[2, 5, 31], William T. Longstreth (University of Washington)[18, 23], Cay Loria (National Heart, Lung, and Blood Institute)[25], Thomas Lumley (University of Auckland)[9, 23], Kathryn Lunetta (Boston University)[28], Aaron J. Mackey (University of Virginia)[16, 18], Rachel Mackey (University of Pittsburgh)[1, 23, 31], Ani Manichaikul (University of Virginia)[8, 15, 18, 31], Taylor Maxwell (University of Texas Health Science Center Houston)[22], Barbara McKnight (University of Washington)[15, 23], James B. Meigs (Brigham and Women's Hospital, Harvard Medical School, Massachusetts General Hospital)[1, 28], Alanna C. Morrison (University of Texas Health Science Center Houston)[3, 15, 17], Solomon K. Musani (University of Mississippi Medical Center)[3, 29], Josyf C. Mychaleckyj (University of Virginia)[13, 15, 31], Jennifer A. Nettleton (University of Texas Health Science Center Houston)[9, 22], Kari North (University of North Carolina Chapel Hill)[1, 3, 9, 10, 13, 15, 17, 34], Christopher J. O'Donnell (Massachusetts General Hospital, National Heart, Lung, and Blood Institute)[2, 5, 9, 11, 12, 14, 15, 17, 20, 28], Daniel O'Leary (Tufts University School of Medicine)[25, 31], Frank S. Ong (Cedars-Sinai Medical Center)[3, 11, 31], Walter Palmas (Columbia University)[3, 15, 31], James S. Pankow (University of Minnesota)[1, 22], Nathan D. Pankratz (Indiana University School of Medicine)[15, 25], Shom Paul (University of Virginia), Marco Perez (Stanford University School of Medicine), Sharina D. Person (University of Alabama at Birmingham, University of Alabama at Tuscaloosa)[25], Joseph Polak (Tufts University School of Medicine)[31], Wendy S. Post (Johns Hopkins University)[3, 9, 11, 14, 20, 31], Bruce M. Psaty (Group Health Research Institute, University of Washington)[3, 5, 9, 11, 14, 15, 23], Aaron R. Quinlan (University of Virginia)[18, 19], Leslie J. Raffel (Cedars-Sinai Medical Center)[6, 11, 31], Vasan S. Ramachandran (Boston University)[3, 28], Alexander P. Reiner (Fred Hutchinson Cancer Research Center, University of Washington)[1, 2, 3, 5, 9, 11, 12, 13, 14, 15, 20, 25, 34], Kenneth Rice (University of Washington)[15, 23], Jerome I. Rotter (Cedars-Sinai Medical Center)[1, 3, 6, 8, 11, 15, 31], Jill P. Sanders (University of Vermont)[23], Pamela Schreiner (University of Minnesota)[25], Sudha Seshadri (Boston University)[18, 28], Steve Shea (Brigham and Women's Hospital, Harvard University)[28], Stephen Sidney (Kaiser Permanente Division of Research, Oakland, CA)[25], Kevin Silverstein (University of Minnesota)[25], David S. Siscovick (University of Washington)[5, 1, 25], Nicholas L. Smith (University of Washington)[2, 15, 20, 23], Nona Sotoodehnia (University of Washington)[3, 15, 23], Asoke Srinivasan (Tougaloo College)[29], Herman A. Taylor (Jackson State University, Tougaloo College, University of Mississippi Medical Center)[5, 29], Kent Taylor (Cedars-Sinai Medical Center)[31], Fridtjof Thomas (University of Texas Health Science Center Houston)[3, 22], Russell P. Tracy (University of Vermont)[5, 9, 11, 12, 14, 15, 17, 20, 23], Michael Y. Tsai (University of Minnesota)[9, 31], Kelly A. Volcik (University of Texas Health Science Center Houston)[22], Chrstina L Wassel (University of

California San Diego)[9, 15, 31], Karol Watson (University of California Los Angeles)[31], Gina Wei (National Heart, Lung, and Blood Institute)[25], Wendy White (Tougaloo College)[29], Kerri L. Wiggins (University of Vermont)[23], Jemma B. Wilk (Boston University)[10, 28], O. Dale Williams (Florida International University)[25], Gregory Wilson (Jackson State University)[29], James G. Wilson (University of Mississippi Medical Center)[1, 2, 5, 8, 9, 11, 12, 14, 17, 20, 29], Phillip Wolf (Boston University)[28], Neil A. Zakai (University of Vermont)[2, 23]

ISGS and SWISS

John Hardy (Reta Lila Weston Research Laboratories, Institute of Neurology, University College London)[18], James F. Meschia (Mayo Clinic)[18], Michael Nalls (National Institute on Aging)[2, 18], Stephen S. Rich (University of Virginia)[2, 4, 7, 8, 9, 11, 14, 15, 17, 18, 31], Andrew Singleton (National Institute on Aging)[18], Brad Worrall (University of Virginia)[18]

LungGO

Michael J. Bamshad (Seattle Children's Hospital, University of Washington)[4, 6, 7, 8, 10, 11, 13, 15, 17, 27], Kathleen C. Barnes (Johns Hopkins University)[2, 10, 12, 14, 15, 17, 20, 24, 30, 32], Ibrahim Abdulhamid (Children's Hospital of Michigan)[27], Frank Accurso (University of Colorado)[27], Ran Anbar (Upstate Medical University)[27], Terri Beaty (Johns Hopkins University)[24, 30], Abigail Bigham (University of Washington)[13, 15, 27], Phillip Black (Children's Mercy Hospital)[27], Eugene Bleecker (Wake Forest University)[33], Kati Buckingham (University of Washington)[27], Anne Marie Cairns (Maine Medical Center)[27], Wei-Min Chen (University of Virginia)[13, 15, 18], Daniel Caplan (Emory University)[27], Barbara Chatfield (University of Utah)[27], Aaron Chidekel (A.I. Dupont Institute Medical Center)[27], Michael Cho (Brigham and Women's Hospital, Harvard Medical School)[13, 15, 24], David C. Christiani (Massachusetts General Hospital)[21], James D. Crapo (National Jewish Health)[24, 30], Julia Crouch (Seattle Children's Hospital)6, Denise Daley (University of British Columbia)[30], Anthony Dang (University of North Carolina Chapel Hill)[26], Hong Dang (University of North Carolina Chapel Hill)[26], Alicia De Paula (Ochsner Health System)[27], Joan DeCelie-Germana (Schneider Children's Hospital)[27], Allen Dozor (New York Medical College, Westchester Medical Center)[27,] Mitch Drumm (University of North Carolina Chapel Hill)[26], Maynard Dyson (Cook Children's Med. Center)[27], Julia Emerson (Seattle Children's Hospital, University of Washington)[27], Mary J. Emond (University of Washington)[10, 13, 15, 17, 27], Thomas Ferkol (St. Louis Children's Hospital, Washington University School of Medicine)[27], Robert Fink (Children's Medical Center of Dayton)[27], Cassandra Foster (Johns Hopkins University)[30], Deborah Froh (University of Virginia)[27], Li Gao (Johns Hopkins University)[24, 30, 32], William Gershan (Children's Hospital of Wisconsin)[27], Ronald L. Gibson (Seattle Children's Hospital, University of Washington)[10, 27], Elizabeth Godwin (University of North Carolina Chapel Hill)[26], Magdalen Gondor (All Children's Hospital Cystic Fibrosis Center)[27], Hector Gutierrez (University of Alabama at Birmingham)[27], Nadia N. Hansel (Johns Hopkins University, Johns Hopkins University School of Public Health)[10, 15, 30], Paul M. Hassoun (Johns Hopkins University)[10, 14, 32], Peter Hiatt (Texas Children's Hospital)[27], John E. Hokanson (University of Colorado)[24], Michelle Howenstine (Indiana University, Riley Hospital for Children)[27], Laura K. Hummer (Johns Hopkins University)[32], Seema M. Jamal (University of Washington)[11], Jamshed Kanga (University of Kentucky)[27], Yoonhee Kim (National Human Genome Research Institute)[24, 32], Michael R. Knowles (University of North Carolina Chapel Hill)[10, 26], Michael Konstan (Rainbow Babies & Children's Hospital)[27], Thomas Lahiri (Vermont Children's Hospital at Fletcher Allen Health Care)[27], Nan Laird (Harvard School of Public Health)[24], Christoph Lange (Harvard School of Public Health)[24], Lin Lin (Harvard Medical School)[21], Xihong Lin (Harvard School of Public Health)[21], Tin L. Louie (University of Washington)[13, 15, 27], David Lynch (National Jewish Health)[24], Barry Make (National Jewish Health)[24], Thomas R. Martin (University of Washington, VA Puget Sound Medical Center)[10, 21], Steve C. Mathai (Johns Hopkins University)[32], Rasika A. Mathias (Johns Hopkins University)[10, 13, 15, 30, 32], John McNamara (Children's Hospitals and Clinics of Minnesota)[27], Sharon

McNamara (Seattle Children's Hospital)[27], Deborah Meyers (Wake Forest University)[33], Susan Millard (DeVos Children's Butterworth Hospital, Spectrum Health Systems)[27], Peter Mogayzel (Johns Hopkins University)[27], Richard Moss (Stanford University)[27], Tanda Murray (Johns Hopkins University)[30], Dennis Nielson (University of California at San Francisco)[27], Blakeslee Noyes (Cardinal Glennon Children's Hospital)[27], Wanda O'Neal (University of North Carolina Chapel Hill)[26], David Orenstein (Children's Hospital of Pittsburgh)[27], Brian O'Sullivan (University of Massachusetts Memorial Health Care)[27], Rhonda Pace (University of North Carolina Chapel Hill)[26], Peter Pare (St. Paul's Hospital)[30], H. Worth Parker (Dartmouth-Hitchcock Medical Center, New Hampshire Cystic Fibrosis Center)[27], Mary Ann Passero (Rhode Island Hospital)[27], Elizabeth Perkett (Vanderbilt University)[27,] Adrienne Prestridge (Children's Memorial Hospital)[27], Nicholas M. Rafaels (Johns Hopkins University)[30], Bonnie Ramsey (Seattle Children's Hospital, University of Washington)[27], Elizabeth Regan (National Jewish Health)[24], Clement Ren (University of Rochester)[27], George Retsch-Bogart (University of North Carolina Chapel Hill)[27], Michael Rock (University of Wisconsin Hospital and Clinics)[27], Antony Rosen (Johns Hopkins University)[32], Margaret Rosenfeld (Seattle Children's Hospital, University of Washington)[27], Ingo Ruczinski (Johns Hopkins University School of Public Health)[13, 15, 30], Andrew Sanford (University of British Columbia)[30], David Schaeffer (Nemours Children's Clinic)[27], Cindy Sell (University of North Carolina Chapel Hill)[26], Daniel Sheehan (Children's Hospital of Buffalo)[27], Edwin K. Silverman (Brigham and Women's Hospital, Harvard Medical School)[24, 30], Don Sin (Children's Medical Center of Dayton)[30], Terry Spencer (Elliot Health System)[27], Jackie Stonebraker (University of North Carolina Chapel Hill)[26], Holly K. Tabor (Seattle Children's Hospital, University of Washington)[6, 10, 11, 17, 27], Laurie Varlotta (St. Christopher's Hospital for Children)[27], Candelaria I. Vergara (Johns Hopkins University)[30], Robert Weiss [30], Fred Wigley (Johns Hopkins University)[32], Robert A. Wise (Johns Hopkins University)[30], Fred A. Wright (University of North Carolina Chapel Hill)[26], Mark M. Wurfel (University of Washington)[10, 14, 21], Robert Zanni (Monmouth Medical Center)[27], Fei Zou (University of North Carolina Chapel Hill)[26]

SeattleGO

Deborah A. Nickerson (University of Washington)[3, 4, 5, 7, 8, 9, 11, 15, 17, 18, 19], Mark J. Rieder (University of Washington)[4, 11, 13, 15, 16, 17, 19], Phil Green (University of Washington), Jay Shendure (University of Washington)[1, 8, 14, 16, 17], Joshua M. Akey (University of Washington)[13, 14, 15], Michael J. Bamshad (Seattle Children's Hospital, University of Washington)[4, 6, 7, 8, 10, 11, 13, 15, 17, 27], Kristine L. Bucasas (Baylor College of Medicine)[15], Carlos D. Bustamante (Stanford University School of Medicine)[3, 13, 15], David R. Crosslin (University of Washington)[2, 9], Evan E. Eichler (University of Washington)[19], P. Keolu Fox[2], Wenqing Fu (University of Washington)[13], Adam Gordon (University of Washington)[11], Simon Gravel (Stanford University School of Medicine)[13, 15], Gail P. Jarvik (University of Washington)[9, 15], Jill M. Johnsen (Puget Sound Blood Center, University of Washington)[2], Mengyuan Kan (Baylor College of Medicine)[13], Eimear E. Kenny (Stanford University School of Medicine)[3, 13, 15], Jeffrey M. Kidd (Stanford University School of Medicine)[13, 15], Fremiet Lara-Garduno (Baylor College of Medicine)[15], Suzanne M. Leal (Baylor College of Medicine)[1, 13, 15, 16, 17, 19, 20], Dajiang J. Liu (Baylor College of Medicine)[13, 15], Sean McGee (University of Washington)[13, 15, 19], Timothy D. O'Connor (University of Washington)[13], Bryan Paeper (University of Washington)[16], Peggy D. Robertson (University of Washington)[4], Joshua D. Smith (University of Washington)[4, 16, 19], Jeffrey C. Staples (University of Washington), Jacob A. Tennessen (University of Washington)[13], Emily H. Turner (University of Washington)[4, 16], Gao Wang (Baylor College of Medicine)[1,13,20], Qian Yi (University of Washington)[4]

WHISP

Rebecca Jackson (Ohio State University)[1, 2, 4, 5, 8, 12, 14, 15, 17, 18, 20, 34], Kari North (University of North Carolina Chapel Hill)[1, 3, 9, 10, 13, 15, 17, 34,] Ulrike Peters (Fred Hutchinson Cancer Research Center)[1, 3, 11, 12, 13, 15, 17, 18, 34], Christopher S. Carlson (Fred Hutchinson Cancer Research Center, University of Washington)[1, 2, 3, 5, 12, 13, 14, 15, 16, 17, 18, 19, 34], Garnet Anderson (Fred Hutchinson Cancer Research Center)[34], Hoda Anton-Culver (University of California at Irvine)[34], Themistocles L. Assimes (Stanford University School of Medicine)[5, 9, 11, 34], Paul L. Auer (Fred Hutchinson Cancer Research Center)[1, 2, 3, 5, 11, 12, 13, 15, 16, 18, 34], Shirley Beresford (Fred Hutchinson Cancer Research Center)[34], Chris Bizon (University of North Carolina Chapel Hill)[3, 9, 13, 15, 34], Henry Black (Rush Medical Center)[34], Robert Brunner (University of Nevada)[34], Robert Brzyski (University of Texas Health Science Center San Antonio)[34], Dale Burwen (National Heart, Lung, and Blood Institute WHI Project Office)[34], Bette Caan (Kaiser Permanente Division of Research, Oakland, CA)[34], Cara L. Carty (Fred Hutchinson Cancer Research Center)[18, 34], Rowan Chlebowski (Los Angeles Biomedical Research Institute)[34], Steven Cummings (University of California at San Francisco)[34], J. David Curb* (University of Hawaii)[9, 18, 34], Charles B. Eaton (Brown University, Memorial Hospital of Rhode Island)[12, 34], Leslie Ford (National Heart, Lung, and Blood Institute, National Heart, Lung, and Blood Institute WHI Project Office)[34], Nora Franceschini (University of North Carolina Chapel Hill)[2, 3, 9, 10, 15, 34], Stephanie M. Fullerton (University of Washington)[6, 11, 34], Margery Gass (University of Cincinnati)[34], Nancy Geller (National Heart, Lung, and Blood Institute WHI Project Office)[34], Gerardo Heiss (University of North Carolina Chapel Hill)[5, 34], Barbara V. Howard (Howard University, MedStar Research Institute)34, Li Hsu (Fred Hutchinson Cancer Research Center)[1, 13, 15, 18, 34], Carolyn M. Hutter (Fred Hutchinson Cancer Research Center)[13, 15, 18, 34], John Ioannidis (Stanford University School of Medicine)[11, 34], Shuo Jiao (Fred Hutchinson Cancer Research Center)[34], Karen C. Johnson (University of Tennessee Health Science Center)[3, 34], Charles Kooperberg (Fred Hutchinson Cancer Research Center)[1, 5, 9, 13, 14, 15, 17, 18, 34], Lewis Kuller (University of Pittsburgh)[34], Andrea LaCroix (Fred Hutchinson Cancer Research Center)[34], Kamakshi Lakshminarayan (University of Minnesota)[18, 34], Dorothy Lane (State University of New York at Stony Brook)[34], Ethan M. Lange (University of North Carolina Chapel Hill)[3, 5, 9, 13, 34], Leslie A. Lange (University of North Carolina Chapel Hill)[1, 2, 3, 5, 9, 12, 13, 15, 17, 18, 20, 25, 34], Norman Lasser (University of Medicine and Dentistry of New Jersey)[34], Erin LeBlanc (Kaiser Permanente Center for Health Research, Portland, OR)[34], Cora E. Lewis (University of Alabama at Birmingham)[25,34], Kuo-Ping Li (University of North Carolina Chapel Hill)[9, 34], Marian Limacher (University of Florida)[34], Dan-Yu Lin (University of North Carolina Chapel Hill)[1, 3, 9, 13, 15, 34], Benjamin A. Logsdon (Fred Hutchinson Cancer Research Center)[2, 34], Shari Ludlam (National Heart, Lung, and Blood Institute WHI Project Office)[34], JoAnn E. Manson (Brigham and Women's Hospital, Harvard School of Public Health)[34], Karen Margolis (University of Minnesota)[34], Lisa Martin (George Washington University Medical Center)[9, 34], Joan McGowan (National Heart, Lung, and Blood Institute WHI Project Office)[34], Keri L. Monda (Amgen, Inc.)[1, 15, 34], Jane Morley Kotchen (Medical College of Wisconsin)[34], Lauren Nathan (University of California Los Angeles)[34], Judith Ockene (Fallon Clinic, University of Massachusetts)[34], Mary Jo O'Sullivan (University of Miami)[34], Lawrence S. Phillips (Emory University)[34], Ross L. Prentice (Fred Hutchinson Cancer Research Center)[34], Alexander P. Reiner (Fred Hutchinson Cancer Research Center, University of Washington)[1, 2, 3, 5, 9, 11, 12, 13, 14, 15, 20, 25, 34], John Robbins (University of California at Davis)[34], Jennifer G. Robinson (University of Iowa)[9, 11, 18, 34], Jacques E. Rossouw (National Heart, Lung, and Blood Institute, National Heart, Lung, and Blood Institute WHI Project Office)[5, 14, 17, 20, 34], Haleh Sangi-Haghpeykar (Baylor College of Medicine)[34], Gloria E. Sarto (University of Wisconsin)[34], Sally Shumaker (Wake Forest University)[34], Michael S. Simon (Wayne State University)[34], Marcia L. Stefanick (Stanford University School of Medicine)[34], Evan Stein (Medical Research Labs)[34], Hua Tang (Stanford University)[2, 34], Kira C. Taylor (University of Louisville)[1, 3, 13, 15, 20, 34], Cynthia A. Thomson (University of Arizona)[34], Timothy A. Thornton (University of Washington)[13, 15, 18, 34], Linda Van Horn (Northwestern University)[34], Mara Vitolins (Wake Forest University)[34], Jean Wactawski-Wende (University of Buffalo)[34], Robert

Wallace (University of Iowa)[2, 34], Sylvia Wassertheil-Smoller (Boston University)[18, 34], Donglin Zeng (University of North Carolina Chapel Hill)[9, 34]
*deceased

NHLBI GO ESP Project Team
Deborah Applebaum-Bowden (National Heart, Lung, and Blood Institute)[4, 7, 12, 17], Michael Feolo (National Center for Biotechnology Information)[12], Weiniu Gan (National Heart, Lung, and Blood Institute)[7, 8, 16, 17], Dina N. Paltoo (National Heart, Lung, and Blood Institute)[4, 6, 11, 17], Jacques E. Rossouw (National Heart, Lung, and Blood Institute, National Heart, Lung, and Blood Institute WHI Project Office)[5, 14, 17, 20, 34], Phyliss Sholinsky (National Heart, Lung, and Blood Institute)[4, 12, 17], Anne Sturcke (National Center for Biotechnology Information)[12]

ESP Groups
[1]Anthropometry Project Team, [2]Blood Count/Hematology Project Team, [3]Blood Pressure Project Team, [4]Data Flow Working Group, [5]Early MI Project Team, [6]ELSI Working Group, [7]Executive Committee, [8]Family Study Project Team, [9]Lipids Project Team, [10]Lung Project Team, [11]Personal Genomics Project Team, [12]Phenotype and Harmonization Working Group, [13]Population Genetics and Statistical Analysis Working Group, [14]Publications and Presentations Working Group, [15]Quantitative Analysis Ad Hoc Task Group, [16]Sequencing and Genotyping Working Group, [17]Steering Committee, [18]Stroke Project Team, [19]Structural Variation Working Group, [20]Subclinical/Quantitative Project Team

ESP Cohorts
[21]Acute Lung Injury (ALI), [22]Atherosclerosis Risk in Communities (ARIC), [23]Cardiovascular Health Study (CHS), [24]Chronic Obstructive Pulmonary Disease (COPDGene), [25]Coronary Artery Risk Development in Young Adults (CARDIA), [26]Cystic Fibrosis (CF), [27]Early Pseudomonas Infection Control (EPIC), [28]Framingham Heart Study (FHS), [29]Jackson Heart Study (JHS), [30]Lung Health Study (LHS), [31]Multi-Ethnic Study of Atherosclerosis (MESA), [32]Pulmonary Arterial Hypertension (PAH), [33]Severe Asthma Research Program (SARP), [34]Women's Health Initiative (WHI)