



Published in final edited form as:

Genet Med. 2010 June ; 12(6): 355–363.

CLIA-tested Genetic Variants on Commercial SNP Arrays: Potential for Incidental Findings in Genome-Wide Association Studies

Andrew D. Johnson, PhD^{1,2,*}, Anupama Bhimavarapu, BS^{1,2,3,*}, Emelia J. Benjamin, MD MSc^{1,4}, Caroline Fox, MD MS^{1,2,3}, Daniel Levy, MD^{1,2,3,4}, Gail P. Jarvik, MD PhD⁵, and Christopher J. O'Donnell, MD MPH^{1,2,3,§}

¹The National Heart, Lung and Blood Institute's (NHLBI) Framingham Heart Study; Framingham, MA, 01702; USA

²Division of Intramural Research and Center for Population Studies, NHLBI, Bethesda, MD, 20892; USA

³Harvard Medical School, Boston, MA, 02116; USA

⁴Boston University School of Medicine and Public Health, Boston, MA, 02118; USA

⁵Division of Medical Genetics, University of Washington Medical Center, Seattle, WA, 98195; USA

Abstract

Purpose—Increases in throughput and affordability of genotyping products have led to large sample sizes in genetic studies, increasing the likelihood that incidental genetic findings may occur. We set out to survey potential notifiable variants on arrays used in genome-wide association studies and in direct-to-consumer genetic services.

Methods—We used multiple bioinformatics strategies to identify and map variants tested for genetic disorders in ≥ 2 CLIA-approved labs (based on the GeneTests database). We subsequently surveyed 18 commercial SNP arrays and HapMap for these variants.

Results—Of 1,362 genes tested according to GeneTests, we identified 298 specific targeted mutations measured in ≥ 2 labs, encompassing 56 disorders. Only 88 of 298 mutations could be identified as known SNPs in genomic databases. We found 18 of 88 SNPs present in HapMap or on commercial SNP arrays. Homozygotes for rare alleles of some variants were identified in the Framingham Heart Study, an active GWAS cohort ($n=8,410$).

Conclusions—Variants in genes including *APOE*, *F5*, *HFE*, *CYP21A2*, *MEFV*, *SPINK1*, *BTD*, *GALT*, and *G6PD* were found on SNP arrays or in the HapMap. Some of these variants may warrant further review to determine their likelihood to trigger incidental findings in the course of GWAS or DTC testing.

[§]Corresponding Author: Christopher J. O'Donnell, MD MPH, The Framingham Heart Study, 73 Mt. Wayte Ave., Suite #2, Framingham, MA 01702, Fax: 508-626-1262, Telephone: 508-935-3435, odonnellc@nhlbi.nih.gov.

*These authors contributed equally.

Disclosures

None

None of the authors have a conflict of interest to disclose.

Keywords

SNP; GWAS; incidental finding; return of results; genetic

INTRODUCTION

Recently, there have been remarkable advances in our understanding of the genetic determinants of common, complex human diseases. The Human Genome Project provided a public reference human genome sequence¹, and the HapMap project² created a genome-wide database of genetic variation of some four million single nucleotide polymorphisms (SNPs). With the identification of so many SNPs and information regarding correlations among these SNPs, genome-wide association studies (GWAS) using high-density arrays of 100K – 1000K SNPs became feasible. GWAS have already led to the discovery of thousands of genetic variants contributing to variability in a range of common diseases including diabetes, cancer, and cardiovascular disease³.

Along with the rapid advance in GWAS and knowledge of genetic variation, an important logistical and ethical issue that has received increasing attention is the potential benefits and risks of identifying “incidental” genetic variants that may compel participant notification⁴. The specific criteria to be used for participant notification of a finding in biomedical research is currently under debate by different interested parties (researchers, participants, practitioners, lawyers, health insurers, those with financial interest in genetic tests), involving issues relating to informed consent, ethics, law and clinical practice⁵. For the purposes of this study we adopted one set of criteria recently published for defining a notifiable genetic variant in population cohort research studies⁶. These criteria from Bookman et al.⁶ indicate a notifiable genetic variant is one with a) established analytic validity, b) for which genetic testing can strongly predict a deleterious clinical outcome with reasonable certainty, and c) for which efficacious early medical intervention exists to reduce the risk of disease or its complications, or which may impact reproductive decisions⁶.

Do participants want to receive genetic results? In the administration of informed consent for genetic research studies some research participants are asked whether they wish to receive the results of their genetic tests. One survey study of potential biobanking participants who did not commit to participation indicated a majority of people would elect to be notified of any significant discoveries regarding their allelic status if it would provide information and future health risk and treatment is available⁷. However, the issue has not been widely studied in consented populations and the extent of information sought in the practice of informed consent is heterogeneous, as are the contexts in which informed consent is applied⁸. With the passage in May 2008 of the Genetic Information Nondiscrimination Act (GINA), which limits health insurers’ and employers’ abilities to use genetic information in a discriminatory manner, some, but not all, potential damaging uses of genetic tests reported back to research participants have been limited. GINA could therefore increase the incentive for individuals to participate in genetic research studies, though the impact of GINA on genetic research remains unknown⁹. In addition, in the past two years, a number of companies have begun to offer genetic testing for SNP variants identified by GWAS and/or genomewide SNP scans direct-to-consumer (DTC) (e.g., 23andMe, DeCode, Variagenics, Knome), suggesting to potential buyers that this information could be clinically useful to them. Accordingly, participants in genetic research may develop greater interest in personal genetic information as DTC information increasingly enters the public consciousness¹⁰.

Despite the growing discussion of many facets around implementing participant notification procedures for actionable genetic results, there is little information regarding the number and types of “notifiable” genetic variants on commercial SNP arrays currently used in GWAS or the potential implications of the use of SNP arrays harboring these variants. We used available genome resources to systematically identify the overlap of SNPs tested in GWAS with genetic tests conducted in CLIA-certified labs. We further provide estimates of the prevalence of notifiable tests from one “real world” study of genomewide association, the Framingham Heart Study (FHS), with 9,274 genotyped and consented participants.

MATERIALS AND METHODS

Abstraction of Disease Variants Tested in CLIA-certified Labs

From a master list of 1,362 genes catalogued by GeneTests, we characterized the types of genetic tests reported as being tested in CLIA-labs (Figure 1). We focused on targeted mutation analyses reported in ≥ 2 CLIA-certified laboratories ($n=89$ disorders, $n=88$ genes). “Targeted mutation analysis” was defined as “testing for either 1) a nucleotide repeat expansion, or 2) ≥ 1 specific mutation, and excluding deletion/duplication analysis or family-specific mutation analysis”. We chose to focus on variants tested in ≥ 2 CLIA-certified laboratories because we felt this would narrow the range of variants to ones where there was a higher level of community consensus on the utility of testing a specific variant. From a list of diseases and genes tested by targeted mutation analysis, we established, where possible, the specific genetic variants tested (Supplemental Table 1). Many of these CLIA-tested variants have been shown to be causally related to disease in prior genotype-phenotype studies. While some DTC tests are conducted in CLIA-certified labs they are not included in the GeneTests CLIA labs dataset since their primary purpose is not to detect highly penetrant clinical variants.

Bioinformatics Search for Unique Identifiers for Variants Tested in CLIA-certified labs

Two reviewers (AB, ADJ) conducted an exhaustive search to ascertain whether the identified targeted mutations possessed reference SNP identifiers (rsIDs). To find rsIDs we employed GeneTests, OMIM, HGMD, dbSNP (Build 129), ExPASy, primary literature (used in some cases to identify and map probe sequences for variants), and the BLAST-Like Alignment Tool (BLAT), human genome reference sequence (Build 36) and SNP tracks all via the UCSC genome browser. For some variants, a detailed search was required to locate the appropriate position in the human genome. Some instances required accounting for varied positions in alternate protein isoforms, resolving multiple names for the same variant among the CLIA-certified laboratories, and confirming and aligning sequences from the variant region based on published variant and probe sequences.

Locating Variants on Commercial SNP Arrays

Once a set of variants with verified SNP rsIDs was confirmed (listed in Supplemental Table 1), we determined if the SNPs themselves or perfect proxy SNPs ($r^2=1.0$) are present on commercially-available Affymetrix or Illumina SNP arrays using SNAP (SNP Annotation and Proxy search tool, version 1.3)¹¹. This tool provides rapid retrieval of HapMap linkage disequilibrium (LD) information, best LD proxies and genotyping array membership for user-defined query SNP lists, and handles a number of related informatics issues relating to SNP queries which otherwise could result in false negative queries¹². Eighteen arrays (listed in Table 1) were scanned for the presence of CLIA-tested variants or proxies in the HapMap CEU population.

Determination of SNP Prevalence in Five Populations

We determined the prevalence of these SNPs and proxy SNPs in the FHS Original cohort¹³, Offspring cohort¹⁴ and Third Generation cohort¹⁵ of the FHS, and in four HapMap populations²: the CEPH (Utah residents with ancestry from northern and western Europe) (CEU), Han Chinese in Beijing, China (CHB), Japanese in Tokyo, Japan (JPT), and Yoruba in Ibadan, Nigeria (YRI). Genotyping was completed in 9,274 Framingham Heart Study participants as part of the SHARe (SNP Health Association Resource) project, using Affymetrix 500K mapping arrays (250K *Nsp I* and 250K *Sty I* arrays) and Affymetrix 50K supplemental Human Gene Focused arrays. Genotyping resulted in 503,551 SNPs with a call rate >95% and Hardy-Weinberg equilibrium p-value >10⁻⁶. Imputation of 2.5 million autosomal SNPs in HapMap was conducted using MACH¹⁶. To examine allele frequencies, we restricted the FHS DNA samples to those with genome-wide call rates of ≥ 97%, samples with genome-wide heterozygosity rates ≤ 5 SD from the norm, and those without unresolved pedigree errors. The final population examined for genotype frequencies here included 8,410 individuals from the FHS study (Original cohort n=962, Offspring cohort n=3,576, Third Generation cohort n=3,872). The Framingham Heart Study protocol is approved by the Institutional Review Board of the Boston University Medical Center, and all participants in the SHARe project provided written informed consent to participate in genetic research.

Localization of SNPs from Published GWAS on Commercial SNP Arrays

From the NHGRI catalog of published GWAS assaying at least 100,000 SNPs, we derived a table of top genetic associations from GWAS. The NHGRI catalog included all SNP-trait associations with p-values < 9.5×10⁻⁶ in the original reports. Information regarding the risk allele frequency in controls from the original report for each SNP was taken from the published table online. We queried SNAP for the relative representation of these GWAS SNPs and their proxies on commercially-available SNP arrays. Information regarding the allele frequencies of SNPs was also obtained from the FHS cohorts and HapMap populations.

RESULTS

A schematic overview of results derived from GeneTests is shown in Figure 2. There were 217 genetic diseases for which genetic tests are performed by ≥2 CLIA-certified labs; a test by targeted mutation analysis is available in ≥2 labs for 89 disorders. Of these, 7 were mitochondrial diseases we did not consider further since SNP arrays generally do not include mitochondrial markers. Additionally, for 26 of the genetic diseases, the particular variants being tested were not specified on the GeneTests site or on the websites of the specific laboratories conducting testing.

For the remaining 56 disorders, 298 specific targeted mutations were reported as tested. A complete list and description of these variants and disorders is in Supplemental Table 1. Of 298 tested variants we identified rsIDs using available genome databases for 88 SNPs (29.5%). Seventy of the 88 SNPs were not found in HapMap. Characteristics of the remaining 18 SNPs present in HapMap are given in Table 1, including gene, SNP description, genotype counts in FHS when available, and minor allele frequencies reported in HapMap populations and calculated, or estimated from imputed results, in FHS. Twelve of these SNPs, representing 9 genes and 9 corresponding diseases, were found to be located on commercial SNP arrays. When we sought evidence for proxy SNPs in HapMap for the 18 variants, we identified six “perfect proxy” SNPs ($r^2=1.0$); physical distances between the CLIA-tested SNPs and the HapMap proxies ranged from 4.1 – 51.4 kb. Four out of these 6 proxies were themselves found on arrays, however the inclusion of these proxies did not

substantially increase the burden of potentially notifiable variants on commercial arrays since they tagged common disease variants for which evidence for clinical impact was modest (e.g., *MTHFR* E429A, *HFE* H63D, *GALT* L218L). Each commercial array contained on average 4.1 of the SNPs or their proxies (range 0 to 11, of a maximum possible 24 SNPs). Of the 18 SNPs found in HapMap, six variants were not on any of the commercially available arrays. One of these SNPs, rs2070075, had a proxy SNP rs12553321 ($r^2=1.0$) that was on two arrays (Illumina Hap1Million, single and dual formats) but did not have strong evidence for a substantial disease burden (*GALT* L218L).

We determined the occurrence and prevalence of potentially notifiable SNPs in a large genome-wide scan performed in the Framingham Heart Study (FHS). Of the 12 SNPs that were directly located on at least one array, 8 SNPs or their proxies were genotyped on arrays used in FHS GWAS (the 100K study on the Offspring cohort: Affymetrix 50K *Xba* I, 50K *Hind* III; the 550K SHARe study including all 3 FHS cohorts). One SNP, rs1801133 in *MTHFR*, was only present on a lower density array used in the FHS 100K GWAS studies, and thus, genotype data was only available in a subset of the Offspring cohort ($n=1,325$). These 8 SNPs corresponded with tests performed in ≥ 2 US CLIA-certified laboratories for HFE-associated hereditary hemochromatosis, biotinidase deficiency, familial Mediterranean fever, hereditary pancreatitis, and *MTHFR* deficiency (see Table 1). In the FHS, the frequency of the minor alleles ranged from 0.00018 to 0.36. These allele frequencies were generally similar to those reported in HapMap.

Representation of GWAS SNPs cross Arrays

From the NHGRI catalog of GWAS findings we derived a list of variants reported to be significantly associated with common diseases, disease traits and other human measurements ($n=325$). We focused on the occurrence and prevalence in FHS of SNPs reported as significantly associated only with common diseases or disease traits in at least one GWAS report ($n=164$, Supplemental Table 2). For common, chronic disease conditions such as adult onset diabetes mellitus, Crohn's or other inflammatory bowel diseases, or breast or colorectal cancer, 50% of the SNPs reported in GWAS ($n=82$) were found directly on arrays included in the FHS 550K study. The average allele frequency for all SNPs ($n=164$) either genotyped or imputed in the FHS cohort was 27.9%, with only two SNPs, rs10498345 and rs16901979, having a MAF $\leq 5\%$. The average MAF for the full set of GWAS SNPs ($n=164$) in the HapMap CEU was 27.6%, similar to that in Framingham. Average MAFs for GWAS SNPs in other HapMap populations were: CHB (28.5%), JPT (27.9%) and YRI (32.2%). Most of the SNPs associated with a common, chronic disease were either directly represented on FHS arrays (550K, 50%; 550K+100K, 62%) or had a highly correlated proxy SNP ($r^2 \geq 0.8$) on the FHS arrays (550K, 86%; 550K+100K, 88%). A similar pattern was observed for other commercial arrays, highlighting the overlap among them and their focus on detecting common variants. By contrast, CLIA-tested variants were less uniformly represented on commercial arrays (12 of 298 variants (4.0%) found across *all* arrays), and were less common among participants (average genotyped or imputed MAF of the 12 variants in FHS was 9.1%).

DISCUSSION

GWAS have been conducted in hundreds of thousands of research participants and are currently underway in many other populations, with increasingly large sample sizes being employed and combined to maximize statistical power. We sought to identify and characterize the potentially "notifiable" genetic variants residing on commercially available SNP arrays used for GWAS. Since the current guidelines for the use of genetic tests from research findings that we adopted indicate these tests should be conducted in CLIA-certified facilities, we considered known genes tested by these laboratories⁶. From our review of all

types of genetic tests currently measured for 1,362 genes, it appears that only 11.7% of the genes are tested for targeted mutation analysis in ≥ 2 CLIA-certified laboratories. Using a variety of bioinformatics methods, we confirmed only 12 SNPs tested in ≥ 2 CLIA-certified laboratories for 9 diseases/genes are currently found on ≥ 1 SNP array used for GWAS. Because GWAS studies increasingly impute millions of SNPs from genotyped SNPs using HapMap information¹¹, it might be possible that many additional potentially notifiable SNPs could be found among the imputed SNPs. However, we confirmed only 4 additional SNPs found on ≥ 1 commercially available SNP scan that are “perfect proxies” for the 12 potentially notifiable SNPs based on the HapMap. Further, we found 6 additional HapMap SNPs tested for 6 diseases/genes in ≥ 2 CLIA-certified laboratories that are not currently found on any SNP arrays. Thus, our findings suggest that SNP arrays used for GWAS, and the HapMap itself, generally harbor very few potentially notifiable genetic variants.

Potential Notifiable SNPs

For a genetic result to be considered notifiable to research participants from community-based populations, a number of criteria have been suggested, including evidence that the notifiable disease has important health implications, that penetrance is relatively high though not necessarily complete ($RR > 2.0$), the risk for disease is strong, the magnitude of risk conferred by the genetic variant is significant, and there are proven therapeutic or preventive interventions available, or there are significant reproductive implications⁶. When we assessed the state of the literature for each of the genes (and the specific associated variants available on the commercially-available SNP arrays) to determine whether they met the criteria and might qualify for notification, we found that few, if any, potentially notifiable variants that reside on arrays used in GWAS meet these criteria. Three disorders – congenital adrenal hyperplasia due to 21-hydroxylase deficiency (*CYP21A2*), biotinidase insufficiency (*BTD*) and galactosemia (*GALT*) – primarily display onset in newborns or infants and are identified with high sensitivity by newborn screening (National Newborn Screening and Genetics Resource Center). A fourth condition, methylenetetrahydrofolate reductase (*MTHFR*) deficiency, is associated with developmental delays in physical and cognitive functions, as well as mental retardation and various psychiatric disturbances¹⁷. However, this condition is incompletely penetrant and routine newborn screening is not conducted.

Similarly, it is unclear whether most of these variants are strongly associated with adult diseases for which treatments are available and that might qualify them for genetic notification. For the two separate amino acid substitutions detectable for *MTHFR*, Ala222Val (677C>T) and Glu429Ala (1298A>C), the evidence for association with cardiovascular disease remains uncertain and the indication for notification is weak. Glucose-6-phosphate dehydrogenase deficiency (Val68Met in *G6PD*, not measured in FHS), Familial Mediterranean fever (FMF, Pro369Ser, Arg408Gln in *MEFV*, for both FHS MAF=0.8%), the nonclassic form of congenital adrenal hyperplasia (Val237Glu, FHS MAF=0.02%), and hereditary pancreatitis (Asn34Ser in *SPINK1*, FHS MAF=1.0%) may manifest postnatally and are uncommon. In the FHS sample we observed a total of 6 homozygotes for the minor alleles of variants in these genes, for which the disease-related gene variants are incompletely penetrant. Knowledge regarding genetic status relating to these conditions might be beneficial, specifically for avoiding triggers for the severe episodic illnesses in hereditary pancreatitis and G6PD deficiency, colchicine prophylaxis in FMF, and steroid replacement in congenital adrenal hyperplasia (CAH); however, there is insufficient evidence regarding the penetrance of these conditions in general community-based populations, and whether there are significant benefits from genetic notification.

We examined evidence for return of results for each of these four conditions. For glucose-6-phosphate dehydrogenase deficiency, a *G6PD* mutation, Val68Met, is present on genotyping

arrays. However, this mutation has only been found to cause a disease phenotype in combination with another mutation, Asn126Asp, which is not present on any of the arrays^{18,19}, making interpretation of the status for Val68Met alone unclear. The *MEFV* mutations present on arrays and related to the autosomal recessive FMF, for which 3 homozygotes have been potentially identified in FHS. Since onset of FMF typically occurs in adulthood, often via a difficult diagnosis after multiple exploratory surgeries, there may be a case for notification. However, the mutations identified here (Pro369Ser, Arg408Gln) are not among the most commonly observed disease variants, their role in FMF has not been clearly elucidated, and some studies suggest incomplete penetrance of these and other *MEFV* alleles^{20–23}. Thus, a careful clinical review and consideration of further mutational screening would likely be necessary in cases where notification regarding *MEFV* mutation homozygosity was considered.

Individuals with the nonclassic form of CAH present postnatally and exhibit moderate enzyme deficiency and in some cases signs of hyperandrogenism. They may be heterozygous for one or more mutations or deletions (compound heterozygotes), including the Val237Glu mutation that is among a cluster of tested mutations in exon 6, and which is present on the array used in the FHS GWAS. In FHS, there were 3 potential heterozygotes for the Val237Glu mutation, which appears to be a functional null²⁴, raising the question of whether such participants might benefit from notification as it could explain and possibly lead to treatment of symptoms if there is also a second undetected mutation, which in combination, might lead to undiagnosed nonclassical CAH. Finally, for hereditary pancreatitis, while both heterozygosity and, particularly, homozygosity for the *SPINK1* mutation (Asn34Ser) is clearly and strongly associated with symptomatic disease²⁵, the pancreatitis shows highly incomplete penetrance, likely due to required environmental triggers, such as infection and there may be little impact on treatment²⁶, although avoidance of alcohol and some drugs may be recommended and earlier recognition of first pancreatitis episode could be a benefit.

Genetic testing for two other common genetic mutations, *F5* (Arg506Gln) and *HFE* (Cys282Tyr), is sometimes conducted in adults with clinical manifestations of venous thromboembolism or hemochromatosis, respectively. Carriers of the Factor V Leiden variant (Arg506Gln) with a history of venous thromboembolism are at increased risk for a second thromboembolic event²⁷. However, there is no evidence of a clear benefit from screening asymptomatic persons for variants in the *F5* gene²⁸. Furthermore, from available clinical trial evidence, there is little evidence that genetic testing predicts responsiveness or aids in decisions regarding use of more intensive anticoagulation in persons with recurrent venous thromboembolism²⁹, or that long-term prophylactic anticoagulation is beneficial for asymptomatic, heterozygous individuals. However, knowledge of Factor V Leiden carrier status might result in modifying exposure to thromboembolic risk factors, such as smoking or prophylactic aspirin use for sedentary periods (e.g., airplane flights). Factor V Leiden may contribute to pregnancy loss, however, testing for this variant is a usual part of the evaluation for recurrent pregnancy loss.

HFE homozygotes have an increased prevalence of liver enzyme abnormalities with increased hepatic iron stores³⁰. There are no randomized trials assessing long term outcomes of phlebotomy in *HFE* Cys282Tyr homozygotes. Since homozygous individuals may or may not have biochemical expression of iron overload, and many will not develop disease and end-organ damage, the use of phlebotomy is reserved for homozygotes with abnormal iron levels. In our current study, we identified 31 potential FHS homozygotes for the rare allele of Cys282Tyr in *HFE* (FHS MAF=5.8%). While available clinical trial data are limited in asymptomatic homozygotes for *HFE* variants, it is possible that manifestations of iron overload can be delayed or averted by interventions to reduce iron intake or overload³¹.

Best practice guidelines exist which do recommend a predictive referral of Cys282Tyr homozygotes for hemochromatosis examination in a clinical setting, but they do not suggest general population screening due to issues of incomplete penetrance³². Even though *HFE* is incompletely penetrant and a test has not been conducted in a CLIA lab, the measurement of iron levels is routine and safe, and if a clinical imbalance was detected a safe and effective treatment exists with routine phlebotomy. Thus, further consideration may be warranted of the potential benefits versus risks for reporting variants in the *HFE* gene to research participants depending on the specific context of the research study in question. In some contexts important clinical information may be available to supplement knowledge of the genotype in deciding whether to return results (e.g., for *F5* or *HFE*, if a medical history of DVT, or high iron levels, is known, respectively), though such situations invoke the boundary between research and clinical practice, how they are defined and what is expected in each.

While not an explicit goal of this study, we briefly consider the next steps, ethical obligations and caveats in the potential incidental notification process, a subject that has been discussed extensively elsewhere^{5,6,33,34}. If there is a clear medical benefit that outweighs the risks of notifying participants who have consented to such notification, then there seems to be an ethical obligation to inform. We did not immediately deem any of the variants considered here to clearly meet notifiability criteria. However, we decided to bring these findings to our independent Ethics Advisory Board to solicit input and a recommendation was made for further consultation with outside experts. If variants are deemed notifiable a number of additional practical steps might be considered, including: 1) consulting the consent form for the study to assess whether individuals clearly indicated an opt-in or opt-out preference to be informed, in recognition of individuals who would not want this information, 2) reviewing available genotype quality assessments (e.g., cluster plots), if any, for the variants in question to determine potential false positives, and considering additional validation genotyping in a CLIA-certified lab, 3) if clinical information is available and consent provided, conducting a clinical review for evidence of related, possibly undiagnosed disorders if this is deemed appropriate to the situation at hand, and 4) verifying that the participants are alive and can be contacted. Identification of resources including a specialized care provider, such as a genetic counselor and/or medical geneticist, and development of educational materials may also be warranted.

Implication of SNP Associations Identified from GWAS

To our knowledge our study represents the first study that seeks to quantify the overlap between SNPs tested in GWAS and genetic tests administered in CLIA-certified labs. Because many CLIA-lab tests target rare, Mendelian conditions caused by low frequency genetic variants, it is not surprising that the overlap is low between most CLIA-lab tests and GWAS SNPs. GWAS studies are providing strong evidence for associations between individual SNPs and chronic diseases and quantitative disease traits. Because there is substantial overlap (Supplemental Table 2) among commercial arrays for “significant” GWAS SNPs, most GWAS that use high density SNP arrays are likely to identify or reliably impute genotypes for most GWAS-reported risk alleles. A number of direct-to-consumer companies have formed to market tests based upon GWAS findings^{8,35}. However, for nearly all GWAS associations to date on which DTC reporting is largely based, the magnitude of risk is modest (relative risk <1.5) -- excepting age-related macular degeneration and *CFH* polymorphism -- and evidence from clinical trials for interventions that alter risk based on SNPs is lacking. Thus, at present there appears to be insufficient evidence to classify SNPs from most GWAS as “notifiable” using evidence-based criteria for such reporting⁶. Continual re-evaluation of the rapidly evolving evidence base is warranted regarding the threshold for reporting SNPs as further GWAS are conducted. As

consumers purchase DTC tests and look for explanations of the results participants may also increasingly query researchers about their own individual genetics results^{10,35}.

Limitations

Our survey and the resulting tables are only as comprehensive as the information provided by the laboratories contributing test information to GeneTests. While many laboratories specified which genetic variants they were testing for a given disease, some laboratories did not do so; as a result some variants could not be identified and checked against SNP arrays. We also focused on US CLIA-certified labs. Other databases exist (e.g., EuroGentest) but the participating labs generally conform to standards other than CLIA. We did not specifically consider combinations of signals at sex-linked markers to identify ploidy-related sex chromosome disorders like Turner or Klinefelter syndromes. Detecting such ploidy events requires an additional level of analysis on the part of researchers beyond simple SNP calling. Many ploidy disorders are diagnosed in childhood, so their incidental detection in GWAS is most relevant to young cohorts, however, some may persist undiagnosed into adulthood³⁶ and thus could trigger incidental findings in older populations.

Many of the disease-causing SNPs are not identified and tagged in human genetic databases (e.g., dbSNP, OMIM) as these databases were originally designed to generally capture variation with a MAF of $\geq 1\%$ in general populations. Moreover, the SNP identifiers in the OMIM and UCSC databases did not always correlate with those in dbSNP, or with each other. This is largely due to redundant alias SNPids¹². We found only 29% of CLIA-tested SNPs have an identifier in genomics databases. The ways in which some CLIA-tested variants are identified, tested and named can create confusion in identifying a precise genomic position for these variants. For example, a number of CLIA-tested variants are known by multiple names referring to restriction digests used, protein positions in precursor or mature proteins, or varying positions with respect to intron borders¹². With increasing sample numbers, denser arrays and deeper genome sequencing, a more systematic informatics effort would be beneficial to ensure that both common and rare disease-associated variants are clearly identified in public databases. Without more systematic cataloging of rare disease variants there are clear informatics barriers to the identification of notifiable genetic findings in genomics research. The current study focuses on SNPs to the exclusion of other types of polymorphisms such as insertions, deletions, inversions and CNVs that GWAS arrays generally do not directly capture³⁷.

Conclusions

Our results indicate a small fraction of disease-causing variants currently tested by CLIA-certified labs are in HapMap and/or commercially-available SNP arrays. Of this limited number of potentially-notifiable variants, few seem to fit criteria for notifiability. For several, newborn disease screening is widely practiced in most US states. For other variants, the penetrance with known adult diseases is low and/or there is insufficient evidence for therapeutic or preventive interventions that would avert disease risk in adults. Nevertheless, for some variants, such as the Cys282Tyr mutation in the *HFE* gene associated with hereditary hemochromatosis, further consideration may be warranted regarding the benefits versus risks of participant notification. Although DTC SNP tests are increasing, there is insufficient evidence regarding virtually all common SNPs implicated in GWAS to suggest these variants should be considered notifiable at present. Given rapid advances in knowledge regarding SNPs on GWAS arrays and regarding the potential benefits and risks of reporting genetic variants, further research and continual updating of our knowledge regarding genetic return of results will be needed to allow informed and appropriate reporting of genetic variants that are detected in the course of research studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. We thank Shih-Jen Hwang for assistance with FHS genotyping results.

Funding Sources

This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195). ADJ was supported by an NHLBI IRTA fellowship award. AB was supported by the NIH Summer Internship Program in Biomedical Research. GPJ was supported by a State of Washington Life Sciences Discovery Fund.

Reference List

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921. [PubMed: 11237011]
2. International HapMap Consortium. A second generation human haplotype map of 3.1 million SNPs. *Nature* 2007;449:851–861. [PubMed: 17943122]
3. Johnson AD, O'Donnell CJ. An open access database of genome-wide association results. *BMC Med Genet* 2009;10:6. [PubMed: 19161620]
4. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. *J Amer Med Assoc* 2006;296:212–215.
5. Wolf SM, Lawrenz FP, Nelson CA, et al. Managing incidental findings in human subjects research: analysis and recommendations. *J Law Med Eth* 2008;36:219–248.
6. Bookman EB, Langehorne AA, Eckfeldt JH, et al. Reporting genetic results in research studies: summary and recommendations of an NHLBI working group. *Am J Med Genet A* 2006;140:1033–1040. [PubMed: 16575896]
7. Hoeyer K, Olofsson BO, Mjorndal T, Lynoe N. Informed consent and biobanks: A population-based study of attitudes towards tissue donation for genetic research. *Scand J Public Health* 2004;32:224–229. [PubMed: 15204184]
8. Hoeyer K. Donors perceptions of consent to and feedback from biobank research: time to acknowledge diversity? *Public Health Genomics*. 2009 Nov 26; epublished.
9. Hudson KL, Holohan MK, Collins FS. Keeping pace with the times--the Genetic Information Nondiscrimination Act of 2008. *N Engl J Med* 2008;358:2661–2663. [PubMed: 18565857]
10. Evans JP, Green RC. Direct to consumer genetic testing: Avoiding a culture war. *Gen in Med* 2009;11:568–569.
11. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 2008;24:2938–2939. [PubMed: 18974171]
12. Johnson AD. Single-nucleotide polymorphism bioinformatics: A comprehensive review of resources. *Circ Cardio Gen* 2009;2:530–536.
13. Dawber TR, Kannel WB, Lyell LP. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci* 1963;107:539–556. [PubMed: 14025561]
14. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* 1979;110:281–290. [PubMed: 474565]
15. Splansky GL, Corey D, Yang Q, et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* 2007;165:1328–1335. [PubMed: 17372189]

16. Li Y, Abecasis G. MACH 1.0: Rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet* 2006;79:2290.
17. Cohen AF, Sedel F, Papo T. Cystathionine betasynthase and MTHFR deficiencies in adults. *Rev Neurol (Paris)* 2007;163:904–910. [PubMed: 18033026]
18. Hirono A, Kawate K, Honda A, Fujii H, Miwa S. A single mutation 202G>A in the human glucose-6-phosphate dehydrogenase gene (G6PD) can cause acute hemolysis by itself. *Blood* 2002;99:1498. [PubMed: 11852882]
19. Town M, Bautista JM, Mason PJ, Luzzatto L. Both mutations in G6PD A- are necessary to produce the G6PD deficient phenotype. *Hum Mol Genet* 1992;1:171–174. [PubMed: 1303173]
20. Asktentjevich I, Torosyan Y, Samuels J, et al. Mutation and haplotype studies in familial Mediterranean fever reveal new ancestral relationships and evidence for a high carrier frequency with reduced penetrance in the Ashkenazi Jewish population. *Am J Hum Genet* 1999;64:949–962. [PubMed: 10090880]
21. Cazeneuve C, Sarkisian T, Pecheux C, et al. MEFV-gene analysis in Armenian patients with familial Mediterranean fever: diagnostic value and unfavorable renal prognosis of the M694V homozygous genotype - genetic and therapeutic implications. *Am J Hum Genet* 1999;65:88–97. [PubMed: 10364520]
22. Touitou I. The spectrum of familial Mediterranean fever (FMF) mutations. *Eur J Hum Genet* 2001;9:473–483. [PubMed: 11464238]
23. Sarkisian T, Ajrapetian H, Beglarian A, et al. Familial Mediterranean fever in Armenian population. *Georgian Med News* 2008;156:105–111. [PubMed: 18403822]
24. Robins T, Barbaro M, Lajic S, Wedell A. Not all amino acid substitutions of the common cluster E6 mutation in CYP21 cause congenital adrenal hyperplasia. *J Clin Endocrinol Metab* 2005;90:2148–2153. [PubMed: 15623806]
25. O'Reilly DA, Witt H, Rahman SH, et al. The SPINK1 N34S variant is associated with acute pancreatitis. *Eur J Gastroenterol Hepatol* 2008;20:726–731. [PubMed: 18617776]
26. Rosendahl J, Bodeker H, Mossner J, Teich N. Hereditary chronic pancreatitis. *Orphanet J Rare Dis* 2007;2:1–10. [PubMed: 17204147]
27. Ho WK, Hankey GJ, Quinlan DJ, Eikelboorn JW. Risk of recurrent venous thromboembolism in patients with common thrombophilia: a systematic review. *Arch Intern Med* 2006;166:729–736. [PubMed: 16606808]
28. Press RD, Bauer KA, Kujovich JL, Heit JA. Clinical utility of factor V leiden (R506Q) testing for the diagnosis and management of thromboembolic disorders. *Arch Pathol Lab Med* 2002;126:1304–1318. [PubMed: 12421138]
29. Ridker PM, Goldhaber SZ, Danielson E, et al. Long-term, low-intensity warfarin therapy for the prevention of recurrent venous thromboembolism. *N Engl J Med* 2003;348:1425–1434. [PubMed: 12601075]
30. Adams PC, Reboussin DM, Barton JC, et al. Hemochromatosis and iron-overload screening in a racially diverse population. *N Engl J Med* 2005;352:1769–1778. [PubMed: 15858186]
31. Niederau C, Fischer R, Purschel A, et al. Long-term survival in patients with hereditary hemochromatosis. *Gastroenterology* 1996;110:1107–1119. [PubMed: 8613000]
32. King C, Barton DE. Best practice guidelines for the molecular genetic diagnosis of Type I (HFE-related) hereditary hemochromatosis. *BMC Med Gen* 2006;7:81.
33. Cho MK. Understanding incidental findings in the context of genetics and genomics. *J Law Med Ethics* 2008;36:280–285. [PubMed: 18547195]
34. Van Ness B. Genomic research and incidental findings. *J Law Med Ethics* 2008;36:292–297. [PubMed: 18547197]
35. Janssens AC, Gwinn M, Bradley LA, Oostra BA, van Duijn CM, Khoury MJ. A critical appraisal of the scientific basis of commercial genomic profiles used to assess health risks and personalize health interventions. *Am J Hum Genet* 2008;82:593–599. [PubMed: 18319070]
36. Bojesen A, Juul S, Gravholt CH. Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study. *J Clin Endocrinol Metab* 2003;88:622–626. [PubMed: 12574191]
37. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* 2008;118:1590–1605. [PubMed: 18451988]

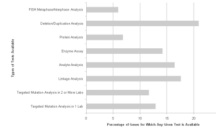


Figure 1. Breakdown of genetic testing types initially considered in this report from the Gene Tests database. Abbreviations: FISH – fluorescent *in situ* hybridization.

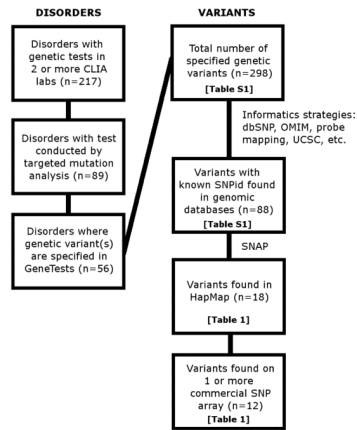


Figure 2. Flowchart depicting identification process for potential notifiable variants represented on commercially available genotyping arrays. Abbreviations: CLIA – Clinical Laboratory Improvements Act, SNP – single nucleotide polymorphism, SNAP – SNP Annotation and Proxy Search tool.

arrays according to HapMap.
 genotypes in FHS or HapMap, or

		Commercial SNP genotyping arrays										FHS allele information										HapMap allele information									
		Affymetrix										FHS allele information										HapMap allele information									
Hap650	Hap650	Hap1M	Hap1M (quad)	Hap1M (single)	Hap1M (dual)	CARE	iSelect	50K Map Xba I	50K Map Hind III	50K Gene Focused	250K Map Nsp I	250K Map Sty I	5.0	6.0	AA	Ab	bb	FHS MAF	CEU MAF	CHB MAF	JPT MAF	YRI MAF									
				x	x		x			x					8373	3	0	0.0002	0	0	0	0	0								
																		0.0046	0.008	0	0	0									
																		0.0035	0.008	0	0	0									
							x			x					8213	130	2	0.008	0.009	0.049	0.048	0.009									
							x			x					8246	128	3	0.0079	0.008	0.024	0.036	0.009									
										x					8160	163	1	0.010	0	0	0.012	0									
										x					7683	676	17	0.042	0.026	0	0	0									
							x											0.162	0.182	0.125	0.042	0									
x																		0.160	0.142	0.044	0.056	0.183									
x										x					6004	2002	185	0.145	0.142	0.044	0.022	0									
																		0.160	0.142	0.044	0.023	0.183									
							x											0.058	0.042	0	0	0									
																		0.103	0.133	0.033	0	0									
							x											-	0	0	0	0.111									
																		0.359	0.242	0.511	0.364	0.108									
																		0.292	0.358	0.20	0.178	0.102									
																		-	0.625	0.80	0.822	0.542									
																		-	0	0	0	0									
																		-	0	0	0	0									
																		-	0	0	0.012	0.009									

Commercial SNP genotyping arrays										FHS allele information				HapMap allele information					
550	Hap650	Hap610 (quad)	Hap1Million (single)	Hap1Million (dual)	CARE iSelect	Affymetrix						AA	Ab	bb	FHS MAF	CEU MAF	CHB MAF	JPT MAF	YRI MAF
						50K Map Xba I	50K Map Hind III	50K Gene Focused	250K Map Nsp I	250K Map Sty I	5.0								
															n/a	0	0.011	0.017	
														0.043	0.042	0.011	0	0	
				x										0.043	0.042	0.011	0	0	
														-	0	0	0	0.114	