

Mapping the incidentalome: estimating incidental findings generated through clinical pharmacogenomics testing

Matthew J. Westbrook, MTS¹, M. Frances Wright, BA¹, Sara L. Van Driest, MD, PhD^{2,3}, Tracy L. McGregor, MD, MSCI²⁻⁴, Joshua C. Denny, MD, MS⁵, Rebecca L. Zuvich, PhD⁴, Ellen Wright Clayton, MD, JD¹⁻³ and Kyle B. Brothers, MD^{1,2,6}

Purpose: Greater clinical validity and economic feasibility are driving the more widespread use of multiplex genetic technologies in routine clinical care, especially for applications in pharmacogenomics. Empirical data on the numbers and types of incidental findings generated through such testing are needed to develop policies and practices related to their clinical use. Of particular importance are disparities in findings relevant to different ancestry groups.

Methods: The Pharmacogenomic Resource for Enhanced Decisions in Care and Treatment Resource, or PREDICT, is an institutional program to implement prospective clinical genotyping of 34 pharmacogenomic-related genes to guide drug selection and dosing. We curated 5,566 journal articles to quantify and characterize the incidental, nonpharmacogenomic genotype–phenotype associations that could be generated through this clinical genotyping project.

Results: We identified 372 putative incidental genotype–phenotype associations that might be revealed in patients undergoing clinical genotyping for pharmacogenomic purposes. Of these, 287 associations were supported by at least one study demonstrating an odds ratio ≥ 2.0 or ≤ 0.5 . Numbers of potentially relevant findings varied widely by ancestry group.

Conclusion: Rigorous clinical policies for the clinical management of incidental findings are needed because the sheer number of significant findings could prove overwhelming to health-care institutions, providers, and patients.

Genet Med 2013;15(5):325–331

Key Words: disease susceptibility; genomics; incidental findings; incidentalome; pharmacogenomics

INTRODUCTION

Multiplex technologies, which can detect an array of desired laboratory results pertinent to a particular clinical issue, often yield ancillary or incidental findings not related to the original motivation for testing. Kohane et al.¹ coined the term incidentalome in 2006 to refer to the potentially voluminous collection of ancillary findings that can be generated through multiplex genetic testing technologies. They predicted that this set of ancillary findings could pose a threat to the implementation of genomic medicine because the number of findings generated, particularly through whole-genome sequencing, could raise a number of challenges. These authors and others have raised concerns that because of the size of the incidentalome, follow-up testing to characterize incidentally identified risks could become very expensive,^{1,2} especially given the inevitability of false-positive results.¹

Even more significant than the challenge of cost, perhaps, is the challenge of scale. The incidental findings generated by a single-gene genetic test are limited in number and can be handled relatively effectively and efficiently by a clinical geneticist, a genetic counselor, or another knowledgeable clinician. But the number of results generated by a high-throughput assay such as a single-nucleotide polymorphism (SNP) chip covering a range of genes, or even a whole-genome sequence, could become

unwieldy for providers to evaluate and validate. The methods for addressing incidental findings on this larger scale is therefore an issue that must be addressed in discussions focused on identifying when genetic results are ready for clinical application. This is especially important given that the time already required for clinicians to provide routine preventive care is substantial.³

Policies designed to address the challenge of scale will need to be informed by the number and clinical relevance of incidental findings potentially generated using specific multiplex genetic technologies. All other things being equal, stronger associations are more likely to be clinically relevant, and most experts agree that findings returned clinically should be clinically “actionable.” The clinical relevance of a particular association for a specific patient additionally depends on the patient’s gender, age, medical history, and health behaviors. Finally, ancestral origin influences the strength of genetic associations, and therefore the clinical relevance of an association could vary among ancestral groups. For all these reasons, it is important to examine whether the number of associations varies among groups of patients.

In this article, we provide an empirical assessment of the incidentalome, taking a clinical pharmacogenomics project as a test case. The Vanderbilt University Medical Center recently designed and built a gene-agnostic and pharmaceutical-agnostic

¹Center for Biomedical Ethics and Society, Vanderbilt University, Nashville, Tennessee, USA; ²Department of Pediatrics, Vanderbilt University, Nashville, Tennessee, USA; ³Monroe Carell Jr. Children’s Hospital at Vanderbilt, Nashville, Tennessee, USA; ⁴Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, USA; ⁵Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, USA; ⁶Department of Pediatrics, University of Louisville School of Medicine, Louisville, Kentucky, USA. Correspondence: Kyle B. Brothers (kyle.brothers@louisville.edu)

infrastructure to support the integration of pharmacogenomic variants into routine clinical care.⁴ This project, named the Pharmacogenomic Resource for Enhanced Decisions in Care and Treatment (PREDICT), currently uses Illumina's VeraCode ADME Core Panel (Illumina, San Diego, CA) as its genotyping platform. This SNP-based platform interrogates 184 SNPs in 34 genes that were selected on the basis of their importance in pharmacogenomics. Thus, they represent especially well-studied targets. In addition to their importance in pharmacogenomics, many have implications for other components of clinical care. Because this panel covers only a small fraction of human genes, and thus only a small slice of the incidentalome, it provides a manageable case study to examine the incidentalome-related challenges that will arise as multiplex genetic testing technologies become more widespread.

In this article, we report on the first attempt to “map” the subset of the incidentalome that could be generated by a multigene SNP test intended for pharmacogenomic use, using the VeraCode ADME Core Panel as a case study. We conducted a systematic review of published articles to quantify and characterize the total number of ancillary findings associated with the genes included in this panel to provide the data necessary to inform practices on reporting incidental findings. These practices should ensure that the aims of efficiency and efficacy envisioned for personalized medicine can be attained, in addition to managing the effect of an incidentalome that could be overwhelming or distracting for patients and providers.

MATERIALS AND METHODS

Initial article database

We performed a comprehensive literature review of all articles available in PubMed as of 22 June 2011 referencing at least one of the 34 genes included in the VeraCode ADME Core Panel (Table 1). We used the Genopedia tool, an online database of genomics-related articles designed and maintained by the Human Genome Epidemiology Network, to generate our initial panel of articles.⁵

Inclusion and exclusion criteria

A genotype–phenotype association was included in our database if the gene being studied was one of the 34 genes included in the VeraCode ADME Core Panel and the phenotype was a medical condition or characteristic with clinical significance. Excluded phenotypes were those not usually evaluated in the

clinical setting (e.g., DNA damage, chromosomal aberrations) and complications of diseases or therapies (e.g., transplant rejection or survival after treatment). Pharmacogenomic associations and associations with nonpharmacological treatment outcomes were also excluded, because these constitute the primary purpose for the genetic test in the PREDICT program, not “incidental” findings. Genome-wide association studies and articles for which no abstract or text in English was available were also excluded. An association was considered statistically significant if the 95% confidence interval for at least one reported odds ratio did not cross 1.0 or another valid statistic indicated significance at the 95% confidence level.

Database curation

Our initial database of published articles included reports on potential genotype–phenotype associations, as well as other studies referencing our genes of interests. We first excluded all articles that were not focused on identifying an association between at least one gene of interest and a clinically relevant phenotype. Next, we determined whether the study produced a statistically significant (positive) finding or a statistically nonsignificant (negative) finding for each gene of interest referenced.

We applied our inclusion and exclusion criteria in two stages. First, we used a computerized algorithm to identify articles that were excluded because they did not report on relevant genotype–phenotype associations (Supplementary Figure S1a online). Second, we hand-curated remaining articles to record positive and negative genotype–phenotype associations and to identify additional articles that should be excluded (Supplementary Figure S1b online).

Computerized curation

Computerized curation proceeded in two stages. First, we used MedEx, a tool originally designed to extract medication information from full-text clinical narratives, to identify articles with titles that refer explicitly to medications.⁶ These articles were classified as pharmacogenomic in focus and were excluded from this study. Second, additional search criteria were used to identify articles meeting other exclusion criteria. For example, articles with titles that contained the text “DNA damage” were excluded because this indicated a focus on a biomarker not usually evaluated in the clinical setting (Supplementary Figure S1 online).

The computerized algorithm was refined to minimize the number of articles that would require hand curation and also to minimize the number of articles incorrectly excluded from review. Random samples of 100 included and excluded articles were reviewed to determine whether articles were mislabeled as qualified or disqualified by the computerized algorithm. The algorithm was then revised iteratively until 100% specificity was attained.

Hand curation

Five authors then hand-curated remaining articles. Articles were assigned to reviewers at random, and one author (K.B.B.) reviewed a subset of each reviewer's inclusions and exclusions to ensure accuracy and consistency. Articles reporting incomplete

Table 1 Genes included in Illumina's VeraCode ADME core panel

<i>ABCB1</i>	<i>CYP2C19</i>	<i>DPYD</i>	<i>SLC22A1</i>	<i>TPMT</i>
<i>ABCC2</i>	<i>CYP2C8</i>	<i>GSTM1</i>	<i>SLC22A2</i>	<i>UGT1A1</i>
<i>ABCG2</i>	<i>CYP2C9</i>	<i>GSTP1</i>	<i>SLC22A6</i>	<i>UGT2B15</i>
<i>CYP1A1</i>	<i>CYP2D6</i>	<i>GSTT1</i>	<i>SLCO1B1</i>	<i>UGT2B17</i>
<i>CYP1A2</i>	<i>CYP2E1</i>	<i>NAT1</i>	<i>SLCO1B3</i>	<i>UGT2B7</i>
<i>CYP2A6</i>	<i>CYP3A4</i>	<i>NAT2</i>	<i>SLCO2B1</i>	<i>VKORC1</i>
<i>CYP2B6</i>	<i>CYP3A5</i>	<i>SLC15A2</i>	<i>SULT1A1</i>	

or ambiguous results were subjected to a secondary review. If an article reported on the association between a gene and more than one phenotype, each genotype–phenotype association was recorded and analyzed separately. Records were managed through the online research database tool REDCap.⁷

For putative genotype–phenotype associations that qualified for inclusion, we recorded the phenotype tested and the population studied. Phenotypes were grouped according to clinical and pathophysiological relationships. For example, angina, acute coronary syndrome, and myocardial infarction were grouped as one phenotype, and aerodigestive tract cancers of the head and neck were grouped as another (**Supplementary Figure S2** online). Populations were grouped at the level of the nation-state (e.g., Brazil, China, Finland) or region (e.g., eastern Europe), except for cases in which the article explicitly stated that research participants came from diverse ethnic or national origins. To facilitate an analysis focused on the largest US ancestral origin groups, findings for European Americans were combined with those from western and central Europe, and findings for African Americans were combined with those from western and sub-Saharan Africa (**Supplementary Table S1** online).

For statistically significant findings, we recorded the odds ratio and 95% confidence interval from the strongest association reported, in terms of magnitude of the odds ratio. We also recorded whether the study examined other factors that could influence the clinical significance of the association including interactions with health behaviors, environmental exposures, occupational exposures, and other gene markers.

Analysis of clinical relevance

To account for some of the factors that influence the clinical relevance of incidental genotype–phenotype associations, we constructed tables of genotype–phenotype associations relevant to two hypothetical patients living in the United States. We did not define or apply criteria for clinical actionability, but instead focused on validity, for which criteria are less stringent. Specifically, we defined an association as “strong” if at least one publication reported an odds ratio for that association to be ≥ 2.0 or ≤ 0.5 .^{8–10} Associations were considered “replicated” if more than one publication reported a positive finding and no publications reported a negative finding. Finally, a finding was considered “clinically relevant” for a hypothetical patient if the association had been demonstrated to have a strong correlation in at least one study conducted with participants from that patient’s ancestral group. Our two case studies involved tallying the number of findings meeting this criterion for a healthy female of European ancestry and for an otherwise identical patient of African ancestry.

RESULTS

Excluded articles

Altogether, we reviewed 5,566 unique articles. A small number of articles (94) could not be evaluated because the abstract was not available through PubMed and the full text of the article could not be obtained in English.

In total, 3,850 articles were excluded: 2,391 by the computerized filter and 1,459 through hand curation. Examples of excluded articles are systematic reviews, articles reporting the frequency of genetic variants in populations, and articles reporting novel gene mutations. Among the excluded articles, 2,277 were found to report only pharmacogenomic findings and 166 were found to report only associations with complications of diseases or treatments.

Genotype–phenotype associations

After exclusions, 1,715 studies were found to have tested associations between at least one gene of interest and a qualifying phenotype. These studies included single-gene studies, small candidate-gene studies, and large pathway-based studies, with an average of 2.0 (SD 1.4) genes of interest examined per article. At least one qualified genotype–phenotype association was found in 26 of the 34 genes included on the VeraCode ADME Core Panel.

Altogether, we examined 806 putative genotype–phenotype associations, of which 434 had been tested but not supported by a statistically significant finding. A total of 91 putative associations were supported by only one positive study with no published attempts at replication, and 14 had been replicated in at least two studies with no published negative findings. There was mixed evidence on most putative genotype–phenotype associations; 267 associations were found to be statistically significant in at least one study and statistically nonsignificant in at least one study (**Table 2**).

On average, each gene carried statistically significant associations with 10.9 phenotypes, strong associations with 8.4 phenotypes, and strong associations that had been replicated with 0.4 phenotypes (**Table 3**). The median number of articles examining each genotype–phenotype association was 2. The most studied genotype–phenotype association was a possible association between *GSTM1* and lung cancer. This association was examined in 83 articles, 28% of which reported statistically significant findings.

Ancestral origin

A total of 158 associations were reported in studies relevant to European Americans, whereas 14 associations were reported

Table 2 Putative genotype–phenotype associations

Evidence category	Genotype–phenotype associations	
	All	Strong correlation ^a
At least one positive finding	372	287
Replicated (≥ 2 positive findings, 0 negative findings)	14	12
No replication attempts (1 positive finding, 0 negative findings)	91	66
Mixed evidence (≥ 1 positive finding, ≥ 1 negative finding)	267	209

^aAt least one positive finding demonstrated odds ratio ≥ 2.0 or ≤ 0.5 .

Table 3 Statistically significant genotype–phenotype associations by gene

Gene	Associated phenotypes		
	All	Strong correlation ^a	Strong correlation, replicated ^b
<i>ABCB1</i>	12	9	1
<i>ABCC2</i>	3	2	0
<i>ABCG2</i>	3	2	0
<i>CYP1A1</i>	40	31	2
<i>CYP1A2</i>	16	10	1
<i>CYP2A6</i>	6	6	1
<i>CYP2B6</i>	2	2	0
<i>CYP2C19</i>	8	6	0
<i>CYP2C8</i>	1	0	0
<i>CYP2C9</i>	7	6	0
<i>CYP2D6</i>	9	7	0
<i>CYP2E1</i>	24	18	0
<i>CYP3A4</i>	7	7	0
<i>CYP3A5</i>	7	4	0
<i>DPYD</i>	0	0	0
<i>GSTM1</i>	60	49	3
<i>GSTP1</i>	42	31	2
<i>GSTT1</i>	55	45	0
<i>NAT1</i>	7	3	1
<i>NAT2</i>	34	31	0
<i>SLC15A2</i>	0	0	0
<i>SLC22A1</i>	0	0	0
<i>SLC22A2</i>	0	0	0
<i>SLC22A6</i>	0	0	0
<i>SLCO1B1</i>	3	1	0
<i>SLCO1B3</i>	0	0	0
<i>SLCO2B1</i>	0	0	0
<i>SULT1A1</i>	11	9	1
<i>TPMT</i>	0	0	0
<i>UGT1A1</i>	7	4	0
<i>UGT2B15</i>	1	1	0
<i>UGT2B17</i>	3	2	0
<i>UGT2B7</i>	1	1	0
<i>VKORC1</i>	3	0	0

^aAt least one study indicated that at least one gene variant conferred risk with odds ratio ≥ 2.0 or ≤ 0.5 . ^bFindings replicated in more than one publication, at least one of which demonstrated a correlation at the odds ratio ≥ 2.0 or ≤ 0.5 level, with no negative findings.

in studies relevant to African Americans. We identified only one study conducted in a US population explicitly described as white Hispanic. Seventy-two associations were identified in participants living in the United States from multiple ancestry groups (Supplementary Table S1 online).

Clinical relevance

In all, 287 genotype–phenotype associations were supported by at least one study demonstrating a strong correlation

(Table 2). Of these associations, 103 were supported by more than one such study. The subset of these strong genotype–phenotype associations that could be identified in a healthy female patient of European ancestry is shown in Table 4. Genotyping on the VeraCode ADME Core Panel could identify 100 clinically relevant genotype–phenotype associations in a female patient of European ancestry, of which 39 have been replicated. By comparison, the same genotyping could identify nine clinically relevant genotype–phenotype associations in a female patient of African ancestry, of which only one has been replicated (Table 5).

DISCUSSION

Only two previous studies have sought to characterize the incidental findings that could be generated through multiplex genetic tests. In 2002, Hirschhorn *et al.*¹¹ reported most of the gene–disease associations that had been identified up to that point, excluding genes of known monogenic disorders and associations with human leukocyte antigen markers and blood group antigens. Multiplex genetic tests were not being used clinically at that time, so the results from the study were not framed in terms of incidental findings. However, had genome-wide SNP chips been in clinical use at that time, polymorphisms in 268 genes would have generated incidental findings across 133 common diseases and traits. Only 166 genotype–phenotype associations had been examined in at least three studies, and only six of those were reproduced in 75% or more of the relevant studies. At that time, the portion of the incidentalome that had been characterized was quite small.

In 2007, Henrikson *et al.*¹² reviewed abstracts of 555 articles for associations between genetic variants relevant to pharmacogenomics and at least one condition. They found that among 42 pharmacogenomic variants, only 22 (52.4%) had been found to be associated with a disease in more than one study, and only 7 (16.7%) had been associated with multiple conditions in two or more studies. We studied a group of 34 genes that overlapped significantly with those studied by Henrikson *et al.* We found that 20 (58.8%) had been associated with at least one disease in more than one study and 14 (41.2%) with multiple conditions in two or more studies.

This study confirmed that replicated ancillary findings are generated through pharmacogenetic tests, but it was not comprehensive enough to quantify the number of results that may be generated. In addition, the relevant science has advanced in the past 5 years. Our study provides an updated and more comprehensive account of the number of genotype–phenotype associations generated through pharmacogenomics testing.

The incidentalome at the system level

Our study indicates at least two ways the number of potential genotype–phenotype associations will be important to institutions implementing clinical genotyping. First, the process of identifying the complete set of relevant incidental findings was time intensive. Even with informatics tools, classifying

Table 4 Selected strong (odds ratio ≥ 2.0 or ≤ 0.5) genotype–phenotype associations relevant to a female patient of European ancestry

Gene	Selected ^a phenotypes with strong (odds ratio ≥ 2.0 or ≤ 0.5) association
<i>ABCB1</i>	Breast cancer, colorectal cancer, inflammatory bowel disease
<i>ABCG2</i>	Gout
<i>CYP1A1</i>	Breast cancer, colorectal cancer, coronary artery disease, endometrial cancer, head and neck cancer, leukemia, lung cancer, ovarian cancer, psoriasis, type 2 diabetes
<i>CYP1A2</i>	Colorectal cancer, endometrial cancer, hypertension, ovarian cancer
<i>CYP2C9</i>	Mood disorder
<i>CYP2D6</i>	Colorectal cancer, lung cancer, Parkinson disease, scleroderma
<i>CYP2E1</i>	Colorectal cancer, gastric cancer, head and neck cancer, lung cancer, scleroderma
<i>CYP3A4</i>	Lung cancer
<i>CYP3A5</i>	Hypertension
<i>GSTM1</i>	Alcoholic pancreatitis, asthma, basal cell or squamous cell carcinoma, breast cancer, esophageal cancer, head and neck cancer, hypertension, lung cancer, ovarian cancer, renal cancer, rheumatoid arthritis, urinary tract cancer
<i>GSTP1</i>	Breast cancer, colorectal cancer, gastric cancer, Hodgkin lymphoma, lung cancer, Parkinson disease, urinary tract cancer
<i>GSTT1</i>	Asthma, breast cancer, colorectal cancer, coronary artery disease, esophageal cancer, gastric cancer, head and neck cancer, leukemia, lung cancer, melanoma, non-Hodgkin lymphoma, renal cancer, urinary tract cancer
<i>NAT1</i>	Asthma, lung cancer
<i>NAT2</i>	Breast cancer, cervical cancer, endometriosis, gastric cancer, head and neck cancer, lung cancer, ovarian cancer, urinary tract cancer
<i>SULT1A1</i>	Breast cancer, colorectal cancer, endometrial cancer, gastric cancer, head and neck cancer, urinary tract cancer
<i>UGT1A1</i>	Endometrial cancer

^aRelatively minor or rare phenotypes omitted.

Table 5 Strong (odds ratio ≥ 2.0 or ≤ 0.5) genotype–phenotype associations relevant to a female patient of African ancestry

Gene	Phenotypes with strong (odds ratio ≥ 2.0 or ≤ 0.5) association
<i>ABCG2</i>	Coronary artery disease
<i>CYP1A1</i>	Breast cancer
<i>GSTM1</i>	Asthma, liver cancer, lung cancer
<i>GSTP1</i>	Asthma, esophageal cancer, lung cancer
<i>GSTT1</i>	Breast cancer

genotype–phenotype associations required a significant amount of time, care, and specialized knowledge. We estimate that, even with the use of a computerized algorithm that excluded ~40% of articles, our hand curation of articles required 800 person-hours. Our review did not even consider sample size and power, quality of study design, or the possibility of translating results across different genotyping technologies, nor did we directly evaluate clinical actionability. Given that these more detailed evaluations would take even more time, effort, and expertise, our experience highlights the importance of efforts such as the Human Genome Epidemiology network and the Evaluation of Genomic Applications in Practice and Prevention initiative to combine efforts across institutions to evaluate the quality of data on genotype–phenotype associations and their readiness for use in clinical practice.^{13–15}

Second, our findings make it clear that institutions seeking to translate incidental genotype–phenotype associations into clinical care will need to develop robust informatics systems for delivering this information to providers and patients. In our

panel of 34 relatively well-studied genes, the mean number of strong associations (odds ratio ≥ 2.0 or ≤ 0.5) generated for each gene was 8.4 phenotypes. Given the pleiotropy of these genes, reporting genotype information alone to providers will not be sufficient. Patients and providers will need more sophisticated reports organizing and synthesizing data on the relevant disease risks and the quality of the data supporting such assessments.

The incidentalome at the patient level

Even if patients and providers are provided with interpreted reports on genotype–phenotype associations, there will still be a need for contextualization and prioritization at the patient level. Although the size of the incidentalome is large, the number of clinically significant genotype–phenotype associations identified in each patient will vary. Some patients will face an overwhelming number of incidental findings. For example, patients with certain variants in *GSTM1* could carry significantly elevated risks for more than 40 different phenotypes. Other patients may carry no high-risk variants in the 34 genes probed in such a panel.

One solution to this challenge would be to prioritize results. Health-care institutions, following the lead of direct-to-consumer genomic testing companies, may choose to make all results available to patients through tools such as a Web portal. This approach is consistent with the commonly reported (but not unanimous) patient preference to have access to “everything” there is to know from their genetic testing.^{16–18} But for patients with large numbers of incidental findings, clinic visits will need to focus on only the 5 or 10 most significant results, or the results the provider judges to be most important given the patient’s current medical situation. As long as patients are able to access “everything” through

another mechanism, clinical efforts that focus on only the most important results may be well received by patients.

Such an approach may raise concerns with health-care providers who fear liability for failure to address all potential results.^{19–21} This study demonstrates that clinical approaches that treat all genotype–phenotype associations as laboratory “results” in need of clinical attention will be unworkable. Multiplex genetic technologies have the potential in some patients to generate too many incidental findings for their providers to address them all meaningfully.²² Standards of care related to addressing lab findings will need to be reframed to limit the responsibility of providers to only those results that meet appropriate standards of relevance, utility, and quality. The necessity, and even wisdom, in such an approach is supported by efforts in other fields to prioritize clinical time and attention.²³

Prioritizing results

Berg *et al.*²⁴ proposed a “binning system” by which genetic variants can be “triaged” in the clinical, diagnostic setting according to specific reporting criteria. These authors identify three “bins” into which genotype–phenotype associations may be categorized. Bin 1 would contain clinically valid results that also carry clinical utility according to the current literature. Bin 2 would contain clinically valid results that are not considered to be actionable. This bin is further stratified into bins 2A, 2B, and 2C. Bin 2A would hold results that are unlikely to cause distress (such as risks for common diseases) to patients, whereas bins 2B and 2C would hold results that patients are more likely to find distressing (such as risks for Alzheimer disease or Huntington disease). Bin 3 would contain results with unknown clinical implications and would thus hold the majority of incidental findings. The authors argue that such a binning system, when used appropriately, would lead to relatively few results falling into the “clinical utility” bin (bin 1) and would thus allow patients and providers to focus on those results that are most likely to be useful.

Our study supports the assessment that such an approach to prioritizing results will be needed if incidental findings are to be incorporated into clinical care. However, it also highlights that the “devil is in the details.” The valid results generated by the VeraCode ADME core panel could range from 12 associations meeting very strict criteria (replicated findings with at least one study showing a “strong” correlation and no negative findings) to 105 associations meeting less stringent criteria (findings demonstrated in at least positive study and no negative findings). Criteria to identify which of those are also clinically useful should be balanced to ensure that the number of results in bin 1 remains manageable.

Our work also indicates that national and international collaborations such as the Human Genome Epidemiology network and the Evaluation of Genomic Applications in Practice and Prevention initiative will be important in addressing the daunting task of identifying the proper bin for specific findings. Currently available online databases, although useful for a range of applications, do not provide the information needed for clinical applications. For example, the Genopedia interface

for the Human Genome Epidemiology Navigator (used to generate the initial data set for our study) greatly overestimates the relevant genotype–phenotype associations. It catalogs the gene names and MeSH terms that are referenced together in publications,²⁵ but it has not been curated to identify articles that demonstrate clinically relevant associations. On the other hand, the National Human Genome Research Institute catalog of genome-wide association studies underestimates the number of valid genotype–phenotype associations, because it catalogs only studies using genome-wide methodologies.^{26,27} For example, of the 287 strong genotype–phenotype associations identified in this study, only six were found in the National Human Genome Research Institute genome-wide association catalog (data not shown).

Disparities in the incidentalome

This study also demonstrates that the racial and ethnic disparities in genomic science observed in genome-wide association studies are recapitulated among case–control studies.^{28–30} We identified 45 phenotypes in which risk could be assessed at a clinically relevant level among European-American women, but only six phenotypes that could be assessed among African-American women. If we accept that genotype–phenotype associations need to be replicated within an ancestral group before they are implemented in medical care for members of that group, then it is clear that a great deal more scientific work will be required before the benefits made possible through genome-based personalized medicine can be provided equitably across racial and ethnic groups.

Limitations

The primary limitation of this study is that we did not assess quality of study design, adequacy of sample size, or power of each study. We also did not differentiate between different variants within genes; different variants within genes may carry different risks. In addition, we did not assess the clinical actionability of identified genotype–phenotype associations. Because of these factors, our estimate of the number of genotype–phenotype associations relevant to the medical care of patients is likely an overestimate. However, the associations that could be eliminated using more strict criteria are likely to be replaced over time by novel associations and new confirmatory findings, both of which are being reported at an increasing rate.

The complex relationships among race, ethnicity, and genetic ancestry also posed significant challenges. The vast majority of studies we reviewed treated place of residence or self-identified race/ethnicity as an analog for genetic ancestry. Although current data support the generalization that genetic ancestry and self- or observer-reported race/ethnicity are correlated,^{31–34} the appropriate methods for operationalizing genetic ancestry in the clinical application of genetic test findings remain unresolved. In particular, we have speculated that findings generated in western or central Europe will be relevant to the health of European Americans, whereas findings generated in west and sub-Saharan Africa will be relevant to African Americans. This

is an assumption that will require more careful analysis before clinical application of findings, perhaps on a study-by-study basis.

Conclusion

As this quantitative literature review has shown, the sheer number of potential incidental findings generated through whole-genome sequencing is likely to pose an information-management challenge, both for informatics systems and for health-care providers and patients. Managing and categorizing all the genotype–phenotype associations generated through clinical genotyping is likely to overwhelm the resources of individual institutions. Collaborations through national or international networks will be required. Likewise, the amount of time a health-care provider would need to address the number of findings generated through such testing for some patients is likely to exceed the practical limitations of most clinic settings. That such a large number of findings can be generated through a relatively small panel of 34 relatively well-studied genes implies that the “incidentalome” generated through whole-genome sequencing will raise even more significant challenges. The development of relatively stringent policies for “binning” results is an important prerequisite for the effective use of incidental genomic findings in clinical care.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at <http://www.nature.com/gim>

ACKNOWLEDGMENTS

The authors acknowledge the contributions of Dan Roden, Jill Pulley, Jonathan Haines, Dana Crawford, Jonathan Schildcrout, Hua Xu, Jennifer Madison, Erica Bowton, Melissa Basford, Kendrick Newton, and Denise Lillard to this work. These studies were funded by the Vanderbilt Genome-Electronic Records Project, the National Institutes of Health (NIH)/National Human Genome Research Institute grant 1U01HG006378-01, and the Vanderbilt Institute for Clinical and Translational Research, National Center for Advancing Translational Sciences/NIH grant UL1TR000011.

DISCLOSURE

The authors declare no conflict of interest.

REFERENCES

- Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. *JAMA* 2006;296:212–215.
- McGuire AL, Burke W. An unwelcome side effect of direct-to-consumer personal genome testing: raiding the medical commons. *JAMA* 2008;300:2669–2671.
- Yarnall KS, Pollak KI, Østbye T, Krause KM, Michener JL. Primary care: is there enough time for prevention? *Am J Public Health* 2003;93:635–641.
- Pulley JM, Denny JC, Peterson JF, et al. Operational implementation of prospective genotyping for personalized medicine: the design of the Vanderbilt PREDICT project. *Clin Pharmacol Ther* 2012;92:87–95.
- UK Biobank Ethics and Governance Framework. 2007. <http://www.ukbiobank.ac.uk/docs/EGF20082.pdf>. Accessed 2 June 2008.
- Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. *J Am Med Inform Assoc* 2010;17:19–24.
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–381.
- Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001;29:306–309.
- Burke W, Psaty BM. Personalized medicine in the era of genomics. *JAMA* 2007;298:1682–1684.
- Wright CF, MacArthur DG. Direct-to-consumer genetic testing. In: Best DH, Swensen JJ, (eds). *Molecular Genetics and Personalized Medicine*. Springer: New York, 2012:215–236.
- Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002;4:45–61.
- Henrikson NB, Burke W, Veenstra DL. Ancillary risk information and pharmacogenetic tests: social and policy implications. *Pharmacogenomics J* 2008;8:85–89.
- Human Genome Epidemiology Network (HuGENet). 2012. <http://www.cdc.gov/genomics/hugenet/default.htm>. Accessed 21 May 2012.
- Evaluation of Genomic Applications in Practice and Prevention (EGAPP) 2012. 2012. <http://www.egappreviews.org/>. Accessed 21 May 2012.
- Khoury MJ. Human genome epidemiology: translating advances in human genetics into population-based data for medicine and public health. *Genet Med* 1999;1:71–73.
- Halverson CM, Ross LF. Attitudes of African-American parents about biobank participation and return of results for themselves and their children. *J Med Ethics* 2012;38:561–566.
- Murphy J, Scott J, Kaufman D, Geller G, LeRoy L, Hudson K. Public expectations for return of results from large-cohort genetic research. *Am J Bioeth* 2008;8:36–43.
- Ruiz-Canela M, Valle-Mansilla JI, Sulmasy DP. What research participants want to know about genetic research results: the impact of “genetic exceptionalism”. *J Empir Res Hum Res Ethics* 2011;6:39–46.
- Bosch JRt, Grody WW. Keeping up with the next generation: massively parallel sequencing in clinical diagnostics. *J Mol Diagn* 2008;10:484.
- Biesecker LG, Shianna KV, Mullikin JC. Exome sequencing: the expert view. *Genome Biol* 2011;12:128.
- Marchant GE, Campos-Outcalt DE, Lindor RA. Physician liability: the next big thing for personalized medicine? *Personalized Medicine* 2011;8:457–467.
- Ormond KE, Wheeler MT, Hudgins L, et al. Challenges in the clinical application of whole-genome sequencing. *Lancet* 2010;375:1749–1751.
- Maciosek MV, Coffield AB, Edwards NM, Flottemesch TJ, Goodman MJ, Solberg LI. Priorities among effective clinical preventive services: results of a systematic review and analysis. *Am J Prev Med* 2006;31:52–61.
- Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. *Genet Med* 2011;13:499–504.
- HuGE Navigator—Genopedia—Search. 2012. <http://www.hugenavigator.net/HuGENavigator/startPagePedia.do>. Accessed 22 June 2011.
- Hindorf LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009;106:9362–9367.
- Hindorf LA, MacArthur J, Wise A, et al. A catalog of published genome-wide association studies. <http://www.genome.gov/gwastudies>. Accessed 11 September 2012.
- Collins FS, Manolio TA. Merging and emerging cohorts: necessary but not sufficient. *Nature* 2007;445:259.
- Fullerton SM. The input-output problem: whose DNA do we study, and why does it matter? In: Burke W (ed). *Achieving Justice in Genomic Translation: Rethinking the Pathway to Benefit*. Oxford University Press: Oxford; New York, 2011.
- Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. *Trends Genet* 2009;25:489–494.
- Dumitrescu L, Ritchie MD, Brown-Gentry K, et al. Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. *Genet Med* 2010;12:648–650.
- Burchard EG, Ziv E, Coyle N, et al. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med* 2003;348:1170–1175.
- Sinha M, Larkin EK, Elston RC, Redline S. Self-reported race and genetic admixture. *N Engl J Med* 2006;354:421–422.
- Tang H, Quertermous T, Rodriguez B, et al. Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am J Hum Genet* 2005;76:268–275.